Thirty-day Hospital Readmission Prediction Model Based on Common Data Model with Weather and Air Quality Data

Borim Ryu
  Seoul National University

Sooyoung Yoo (✉ yoosoo0@snubh.org)
  Seoul National University Bundang Hospital

Seok Kim
  Seoul National University Bundang Hospital

Jinwook Choi
  Seoul National University

Abstract

Many epidemiological studies have established an association between environmental exposure and clinical outcome for hospital admissions. However, few studies have explored the impact of environmental factors, such as ambient air pollution and meteorological factors, on hospital readmissions using predictive analysis. In this study, we aimed to develop a model to predict unplanned hospital readmissions within 30 days of discharge based on the common data model considering weather and air quality factors. Moreover, we validated the proposed model externally. We developed and compared the following machine learning methods: decision tree, random forest, AdaBoost, and gradient boosting machine–based models. We performed 10-fold cross-validation for internal validation, and external validation was performed by applying the model to unseen data. The performance of the prediction model was evaluated using the area under the receiver operating characteristic curve. PM10, rainfall, and maximum temperature were the weather and air quality variables that most impacted the model. Among the four machine learning models, the AdaBoost-based model demonstrated the best performance and was the most accurate in predicting the readmission of patients with musculoskeletal diseases. External validation demonstrated that the model based on weather and air quality factors is transportable.

Introduction

The relationship among climate, human health, and diseases has been established through multiple studies in the past[1], and various risk factors for hospital admission have been studied based on demographic, environmental, and clinical factors[2–9]. Temperature variation due to global warming has been linked to hospital admission rates in a few studies[10–13]. Humidity is another important health factor. Pollutants, including carbon monoxide and fine air particulates, have been associated with increased admissions for multiple conditions[13–15]. Further understanding of the association between the ambient climate condition and hospital admission will help healthcare stakeholders understand the seriousness of the effect of weather change and implement healthcare-resource and patient-care plans..

The common data model (CDM) is a healthcare data model based on standard terminology. An example is the CDM developed by the Observational Medical Outcomes Partnership (OMOP) and maintained by the Observational Health Data Sciences and Informatics (OHDSI)[16–18]. A system developed by converting data into a CDM is easily applicable through the distribution of the source code of the program without the need to install the software on a specific institution's system[19]. CDM is a data model based on common standard terms. Therefore, it guarantees standardized content from the data model and exhibits high extensibility.

In the present study, we developed and validated four prediction models for hospital readmission within 30 days of discharge using the OMOP CDM as well as weather and air quality factors. In addition, the model performance was externally validated to examine the model's extensibility. To the best of our knowledge, the present study is the first to create a patient-level prediction model for hospital readmission

within 30 days using OMOP CDM and ambient weather data. A predictive model that combines weather and environmental data with a patient's residence information is expected to enhance clinical decision making at the individual patient level.

Results

Out of the 61,922 index hospitalizations from the Seoul National University Hospital (SNUH) data included in our cohort research, 5,794 resulted in a 30-day readmission through emergency-room visits (Table 1). The mean age of the readmitted individuals was 75.2 years, and more than half of the readmitted patients were males. The average length of stay was 2.5 days for the readmitted group and 0.2 days for the non-readmitted group.

Table 1
Basic characteristics of study data for each visit type.

Characteristics		Derived cohorts		P value
		Readmitted (N = 5,794)	Non-readmitted (N = 56,128)	
Age, y, mean (SD)		75.2 (6.8)	74.7 (6.7)	0.000
Gender	Male, n (%)	54.8	49.7	
	Female, n (%)	45.2	50.3	
Age during hospital visit	60s	23.8	26.0	
	70s	49.7	50.2	
	80s	23.7	21.4	
	90s	2.7	2.4	
Season during admission	Spring	25.4	24.1	0.049
	Summer	25.5	26.8	
	Fall	24.4	24.8	
	Winter	24.6	24.3	
Average length of stay, mean (SD)		2.5 (4.3)	0.2 (0.4)	
Charlson Comorbidity Index, mean		1.11	0.52	

Table 2

Number of visits in each disease group and outcome incidence rate in our research cohorts.

Disease groups	Train/test population (internal)		Valid population (external)	
	Target size (N)	% incidence	Target size (N)	% incidence
Diseases of the circulatory system (I00-I99)	9357	14.0	87063	10.3
Mental and behavioral disorders (F00-F99)	3174	16.3	7228	17.5
Diseases of the musculoskeletal system and connective tissue (M00-M99)	13564	11.8	41015	11.7
Diseases of the respiratory system (J00-J99)	10310	15.7	87604	15.1

Table 3
Comparison of disease-specific performance in each model based on the area under the receiver operating characteristic curve.

Disease groups	Prediction models	Clinical covariates		Clinical covariates and W-scores	
		Internal	External	Internal	External
Diseases of the circulatory system	DT	0.653	0.664	0.674	0.679
	RF	0.693	0.688	0.686	0.681
	ADA	0.698	0.672	0.708	0.670
	GBM	**0.726***	**0.704***	**0.717***	**0.696***
Mental and behavioral disorders	DT	0.612	0.706	0.691	0.737
	RF	0.703	0.743	**0.692***	**0.686***
	ADA	0.716	0.747	0.654	0.728
	GBM	**0.747***	**0.751***	0.676	0.727
Diseases of the musculoskeletal system and connective tissue	DT	0.680	0.856	**0.690****	**0.889****
	RF	0.719	0.909	0.734	0.882
	ADA	**0.726****	**0.917****	**0.739***	**0.915***
	GBM	**0.751***	**0.883***	0.725	0.900
Diseases of the respiratory system	DT	0.634	0.651	0.607	0.622
	RF	0.653	0.658	0.643	0.638
	ADA	0.663	0.639	0.667	0.655
	GBM	**0.672***	**0.669***	**0.675***	**0.667***

* The best performances for each disease

** Major improvements in external validation

Table 4
Weather and air quality predictors in W-score.

Disease groups	covariateName	covariateValue	CovariateMean	
			CovariateMean WithOutcome	CovariateMean WithNoOutcome
Diseases of the circulatory system	PM10	0.0016	12.59	13.13
	Rainfall	0.0011	2.22	2.32
	Humidity	0.0005	0.29	0.28
	Min Temperature	0.0006	0.65	0.59
	Max Temperature	0.0005	0.85	0.83
Mental and behavioral disorders	PM10	0.0016	12.36	13.24
	Rainfall	0.0012	1.95	2.30
	Humidity	0.0014	0.41	0.31
	Min Temperature	0.0008	0.67	0.58
	Max Temperature	0.0004	0.71	0.81
Diseases of the musculoskeletal system and connective tissue	PM10	0.0015	12.85	13.18
	Rainfall	0.0012	2.24	2.32
	Humidity	0.0008	0.31	0.30
	Min Temperature	0.0007	0.63	0.55
	Max Temperature	0.0007	0.86	0.84
Diseases of the respiratory system	PM10	0.0038	12.86	13.01
	Rainfall	0.0032	2.30	2.29
	Humidity	0.0005	0.31	0.29
	Min Temperature	0.0012	0.60	0.54
	Max Temperature	0.0036	1.01	0.97

Table2 presents the number of patient visits and readmission incidence rate in different disease groups. The internal and external validation results of the proposed readmission prediction models are presented in Table3, where we can observe the differences in model performance among different diseases. The external validation results indicate that the proposed models show significantly improved performance for the musculoskeletal disease group.Supplemetary Table S3-S6 shows top 20 predictors of each model in this study. According to Table4, PM10, rainfall, and maximum temperature were the weather and air quality variables that most impacted the model among the disease groups.

The receiver operating characteristic curves in Fig. 1 reflect the predictive model performances for the (a) internal and (b) external validation of the Adaboost (ADA) model based on clinical covariates and the (c) internal and (d) external validation of the decision tree (DT) model based on clinical covariates and W-score in patients with diseases of the musculoskeletal system and connective tissue, respectively. The clinical covariate and W-score model exhibited the greatest AUC for both the internal and external validations in the musculoskeletal disease group.

Discussion

We developed a 30-day unplanned hospital readmission prediction model based on OMOP-CDM transformed patient medical records and meteorological public data. We also obtained the weather and air quality records for the patients' residence localities. Furthermore, we established a W-score for individual visits based on the Korean weather warning issuance criteria. Moreover, we developed models capable of predicting readmission at the time of patient discharge using weather forecast data at the clinical site.

Many epidemiological studies have established an association between environmental factors and hospital readmissions[11,12,20,21]. However, few studies have examined the impact of environmental factors, such as ambient air pollution or climate, on hospital readmissions using predictive analysis and the result of the health outcome on the prediction model among hospitals.

We developed a model to predict hospital readmission at the time of discharge based on patient-level clinical diagnosis and drug prescription data before discharge as well as the weather and air quality records for the patient's residence locality. The variables used in the proposed model were designed based on diagnosis and drug information to make the model extensible, considering the standard term mapping issues that may arise in the process of converting electronic health record (EHR) data to OMOP CDM. This is because the diagnosis and drug information are based on terms commonly used by many hospitals. In addition, the Korea Meteorological Administration (KMA) provides short-term weather forecast data for the following three to ten days. Therefore, the proposed forecasting model may be applied to actual clinical settings in the future using the short-term forecast data.

The performance of the proposed model for the respiratory disease cohort was lower than expected. Moreover, the performance of the proposed model for the musculoskeletal disease cohort demonstrated

good scalability. The result of the proposed model in this study was presumed to be due to the occurrence of readmission for acute events that require post-operative management, rather than hospitalization due to the occurrence of chronic diseases in tertiary hospitals. Many patients who needed trauma management after surgery were not hospitalized for a sufficient period. The results of a disease-specific predictive model can be observed in further studies based on our research.

We could not externally validate the proposed model across multiple organizations. However, the proposed model can be easily reintegrated when migrating to a different EHR, either as an embedded frame in the EHR or as a standalone CDM application. Furthermore, the proposed model can perform better using a sophisticated weather data function approach. Our research provides a basis for future applications of the proposed model to clinical settings to manage visiting patients based on clinical and weather data.

In summary, providing a clinical basis for a patient's future risk of readmission at the time of discharge will assist hospitals in developing a patient care plan in advance. We developed a model for predicting hospital readmission based on environmental factors. External verification of the model demonstrated that a high-accuracy model can be developed based on weather and air quality factors. Improving the accuracy of the readmission prediction model will help in establishing patient care plans and making clinical decisions at the time of discharge.

Methods

Study population and clinical data description

Our retrospective cohort study was conducted using OMOP-CDM-converted EHR data between January 1, 2017 and December 31, 2018 from SNUH and the Seoul National University Bundang Hospital (SNUBH) in the Seoul metropolitan area, South Korea. These hospitals have converted the EHR data over a 15-year period into the OMOP CDM.

We considered consecutive hospitalizations among adults over 65 years who were discharged alive and underwent at least one hospitalization or emergency-room visit during our study period. We focused on patients living in the Seoul metropolitan area, including the Gyeonggi Province in South Korea, to create prediction models that consider weather and environmental variables during the study period.

In addition, we categorized patients into subgroups based on weather- and environment-related diseases studied previously[22–28]. Patients diagnosed with mental and behavioral disorders (F00-F99), circulatory system diseases (I00-I99), respiratory diseases (J00-J99), and musculoskeletal system and connective tissue–related diseases (M00-M99) at discharge were included together in subgroups based on the International Classification of Diseases, 10th Revision.

The primary outcome of this study was 30-day unplanned hospital readmission. We referred to the Hospital-Wide All-Cause Unplanned Readmission (HWR) measure from Centers for Medicare & Medicaid

Services (CMS)[29]. According to the HWR measure, CMS classified the planned readmissions into planned disease or treatment groups, including chemotherapy, organ transplant, and rehabilitation. All admissions other than the scheduled admissions were considered to be unscheduled visits.

Figure 2 illustrates the study cohort design derived using SNUH data, which are mainly used as the training dataset in our research. Figure 3 shows the overall study process in this research.

This study was performed in accordance with the relevant guidelines and regulations of the SNUH Institutional Review Board (IRB) and was approved by the SNUH IRB. As it is an observational study and the data source was de-identified, this study was approved based on waivers of informed consent or exemptions by the SNUH IRB (IRB No: B-1504-296-302). This study was also performed in accordance with the relevant guidelines and regulations of the SNUBH Institutional Review Board (IRB) and was approved by the SNUBH IRB. As it is an observational study and the data source was de-identified, this study was approved based on waivers of informed consent or exemptions by the SNUBH IRB (IRB No: X-1908-559-901).

Clinical features, such as the gender of the patient, age of the subject on the index date, diagnosis conditions, drug exposures for patient medications, and the Charlson comorbidity index (Romano adaptation), were obtained using all conditions prior to the end of the readmission interval.

Diagnosis and drug prescription were used as clinical variables for individual patients. Moreover, each variable was extracted from the standardized CONDITION_ERA and DRUG_ERA in the CDM table as a higher concept of individual diagnosis and drugs. In OMOP CDM, a CONDITON_ERA data table is defined as the duration in which the patient is assumed to have a given condition[30]. The CONDITION_ERA table provided a chronological period of diagnosis. DRUG_ERA is defined as the duration in which the patient is assumed to be exposed to a particular active drug ingredient. The DRUG_ERA table provided successive periods of individual drug prescriptions combined following certain rules to produce continuous eras.

Weather and air quality data

Weather and air quality data were derived from KMA's weather data open portal (https://data.kma.go.kr) and the official website of the Korean Ministry Of Environment (MOE) (https://www.airkorea.or.kr/eng)[31,32].

Records of daily mean temperature (ºC), daily mean relative humidity (RH) percentage (%), and daily rainfall (mm) during the study period were obtained from the KMA website. The daily mean concentration of ambient particulate matter (PM in $\mu g/m^3$), sulfur dioxide (SO_2 in $\mu g/m^3$), nitrogen dioxide (NO_2 in $\mu g/m^3$), and ozone (O_3 in $\mu g/m^3$) from all general monitoring stations were collected from the Air Korea website for the study period. The daily median was averaged across the data for any missing record on a particular day. KMA and Air Korea data needed to be preprocessed into postal zip codes owing to the varying levels of location information granularity.

W-score: Weather and air quality scores for individual visits

We calculated a patient-level W-score based on weather and air quality data for each patient visit based on the patient's residence locality. The score was derived using the KMA's standards for special weather reports[33]. A special weather report refers to a forecast that calls attention to or warns against a serious disaster that is expected to occur because of a weather phenomenon. An "advisory" is issued if a disaster is expected because of a specific weather phenomenon, and a "warning" is issued if significant damage is expected. KMA issues weather reports on strong winds, wind waves, heavy rains, heavy snow, dry weather, storm tidal waves, earthquakes, cold waves, typhoons, yellow dust, and heat waves (Supplemetary Table S1 and S2).

Data such as the daily average particulate matter (PM10), maximum temperature, minimum temperature, relative humidity, and precipitation were used. Moreover, W-scores of individual patient visits were calculated using weather conditions, such as fine dust warning, heat wave, cold wave, dryness, and heavy rain, respectively. The meteorological warning issuance criteria of the KMA were used for calculating W-scores for each element. We obtained the W-score by calculating the sum of the weather element–specific forecast values for 7 days from the discharge date so that the weather forecast data from KMA can be utilized at the time of the patient's discharge.

Model development

The prediction model for re-admission within 30 days was developed to reflect variables such as clinical diagnosis and drug prescription prior to the patient's discharge date as well as to predict the occurrence of re-admission of the patient by considering the W-score for the weather forecast at the patient's residence location after the discharge date (Fig. 4).

We developed tree-based machine learning models, namely, DT, random forest (RF), ADA, and gradient boosting machine (GBM)–based classifiers, based on the weather and air quality feature set using the patient-level prediction R package developed by OHDSI. Models were trained and tested on SNUH data. Ten-fold cross-validation was primarily used for internal validation. Moreover, the models were externally validated using the SNUBH dataset. Each model performance was evaluated using the area under the receiver operating characteristic curve.

Declarations

Acknowledgements

Author Contributions

B.R. analyzed the data and drafted the manuscript as the first author, S.K. helped prepare and evaluate the data, S.Y. helped analyze the data and managed the overall study, and J.C. supervised the overall study. S.Y. and J.C contributed equally to the research as co-corresponding authors. All authors reviewed the manuscript.

Conflicts of Interest

None of the authors have any conflicts of interest to declare.

References

1.   Mcmichael, A. J. *et al. Climate change and human health RISKS AND RESPONSES Editors*. (2003).

2.   Wang, C.-L. *et al.* Factors associated with emergency department visit within 30 days after discharge. (2016). doi:10.1186/s12913-016-1439-x

3.   Silverstein, M. D., Qin, H., Mercer, S. Q., Fong, J. & Haydar, Z. Risk factors for 30-day hospital readmission in patients≥ 65 years of age. in *Baylor University Medical Center Proceedings* **21**, 363–372 (Taylor & Francis, 2008).

4.   Hong, J., Choi, K., Lee, J. & Lee, E. A study on the factors related to the readmission and ambulatory visit in an university hospital: using patient care information DB. *J. Korean Soc. Med. Informatics* **6**, 23–33 (2000).

5.   Boland, M. R., Parhi, P., Gentine, P. & Tatonetti, N. P. Climate Classification is an Important Factor in Assessing Quality-of-Care Across Hospitals. *Sci. Rep.* **7**, 3–8 (2017).

6.   Shebeshi, D. S., Dolja-Gore, X. & Byles, J. Unplanned readmission within 28 days of hospital discharge in a longitudinal population-based cohort of older australian women. *Int. J. Environ. Res. Public Health* **17**, 3136 (2020).

7.   Kansagara, D. *et al.* Risk prediction models for hospital readmission: A systematic review. *JAMA - Journal of the American Medical Association* **306**, 1688–1698 (2011).

8.   Low, L. L. *et al.* Predicting 30-Day Readmissions: Performance of the LACE Index Compared with a Regression Model among General Medicine Patients in Singapore. *Biomed Res. Int.* **2015**, (2015).

9.   Van Walraven, C. *et al.* Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *Cmaj* **182**, 551–557 (2010).

10.   Bishop-Williams, K. E. *et al.* Understanding weather and hospital admissions patterns to inform climate change adaptation strategies in the healthcare sector in uganda. *Int. J. Environ. Res. Public Health* **15**, 2402 (2018).

11.   Lam, H. C. Y., Chan, J. C. N., Luk, A. O. Y., Chan, E. Y. Y. & Goggins, W. B. Short-term association between ambient temperature and acute myocardial infarction hospitalizations for diabetes mellitus patients: A time series study. *PLoS Med.* **15**, 1–18 (2018).

12.   Lim, Y. H. *et al.* Ambient temperature and hospital admissions for acute kidney injury: A time-series analysis. *Sci. Total Environ.* **616–617**, 1134–1138 (2018).

13.   Dominici, F. *et al.* Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Assoc.* **295**, 1127–1134 (2006).

14.   Ab Manan, N., Noor Aizuddin, A. & Hod, R. Effect of Air Pollution and Hospital Admission: A Systematic Review. *Ann. Glob. Heal.* **84**, 670 (2018).

15.   Su Oh, J. *et al. Ambient Particulate Matter and Emergency Department Visit for Chronic Obstructive Pulmonary Disease. 28 ⬚ ⬚ 1 ⬚* **28**, (2017).

16.   OHDSI – Observational Health Data Sciences and Informatics. Available at: https://www.ohdsi.org/. (Accessed: 29th December 2020)

17.   Hripcsak, G. *et al.* Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. in *Studies in Health Technology and Informatics* **216**, 574–578 (IOS Press, 2015).

18.   Marc Overhage, J., Ryan, P. B., Reich, C. G., Hartzema, A. G. & Stang, P. E. Validation of a common data model for active safety surveillance research. *J. Am. Med. Informatics Assoc.* **19**, 54–60 (2012).

19.   Reps, J. M., Schuemie, M. J., Suchard, M. A., Ryan, P. B. & Rijnbeek, P. R. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *J. Am. Med. Informatics Assoc.* **25**, 969–975 (2018).

20.   Blecker, S., Kwon, J. Y., Herrin, J., Grady, J. N. & Horwitz, L. I. Seasonal Variation in Readmission Risk for Patients Hospitalized with Cardiopulmonary Conditions. *Journal of General Internal Medicine* **33**, 599–601 (2018).

21.   Slama, A. *et al.* Impact of air pollution on hospital admissions with a focus on respiratory diseases: a time-series multi-city analysis. *Environ. Sci. Pollut. Res.* **26**, 16998–17009 (2019).

22.   Blecker, S., Kwon, J. Y., Herrin, J., Grady, J. N. & Horwitz, L. I. Seasonal Variation in Readmission Risk for Patients Hospitalized with Cardiopulmonary Conditions. *Journal of General Internal Medicine* **33**, 599–601 (2018).

23.   Ross, J. S. *et al.* Statistical models and patient predictors of readmission for heart failure: A systematic review. *Archives of Internal Medicine* **168**, 1371–1386 (2008).

24.   Gould, D. *et al.* Patient-related risk factors for unplanned 30-day readmission following total knee arthroplasty: A protocol for a systematic review and meta-analysis. *Syst. Rev.* **8**, 1–8 (2019).

25.   Wang, H., Wang, L., Sun, Z., Jiang, S. & Li, W. Unplanned hospital readmission after surgical treatment for thoracic spinal stenosis: incidence and causative factors. *BMC Musculoskelet. Disord.* **22**, 1–8 (2021).

26.   Han, X. *et al.* Factors associated with 30-day and 1-year readmission among psychiatric inpatients in Beijing China: A retrospective, medical record-based analysis. *BMC Psychiatry* **20**, 1–12 (2020).

27.   Biese, K. *et al.* Predictors of 30-Day Return Following an Emergency Department Visit for Older Adults. *N. C. Med. J.* **80**, 12–18 (2019).

28.   Weinreich, M. *et al.* Predicting the risk of readmission in pneumonia a systematic review of model performance. *Annals of the American Thoracic Society* **13**, 1607–1614 (2016).

29.   for Medicare, C. & Services, M. *2016 Measure Information About the 30-Day All-Cause Hospital Readmission Measure*.

30.   OMOP CDM v6.0. Available at: https://ohdsi.github.io/CommonDataModel/cdm60.html#OMOP_CDM_v60. (Accessed: 21st May 2021)

31.   Open MET Data Portal. Available at: https://data.kma.go.kr/resources/html/en/aowdp.html. (Accessed: 21st May 2021)

32.   AirKorea. Available at: https://www.airkorea.or.kr/eng. (Accessed: 21st May 2021)

33.   Weather Forecast > Weather Forecast. Available at: https://www.kma.go.kr/eng/biz/forecast_01.jsp. (Accessed: 25th May 2021)
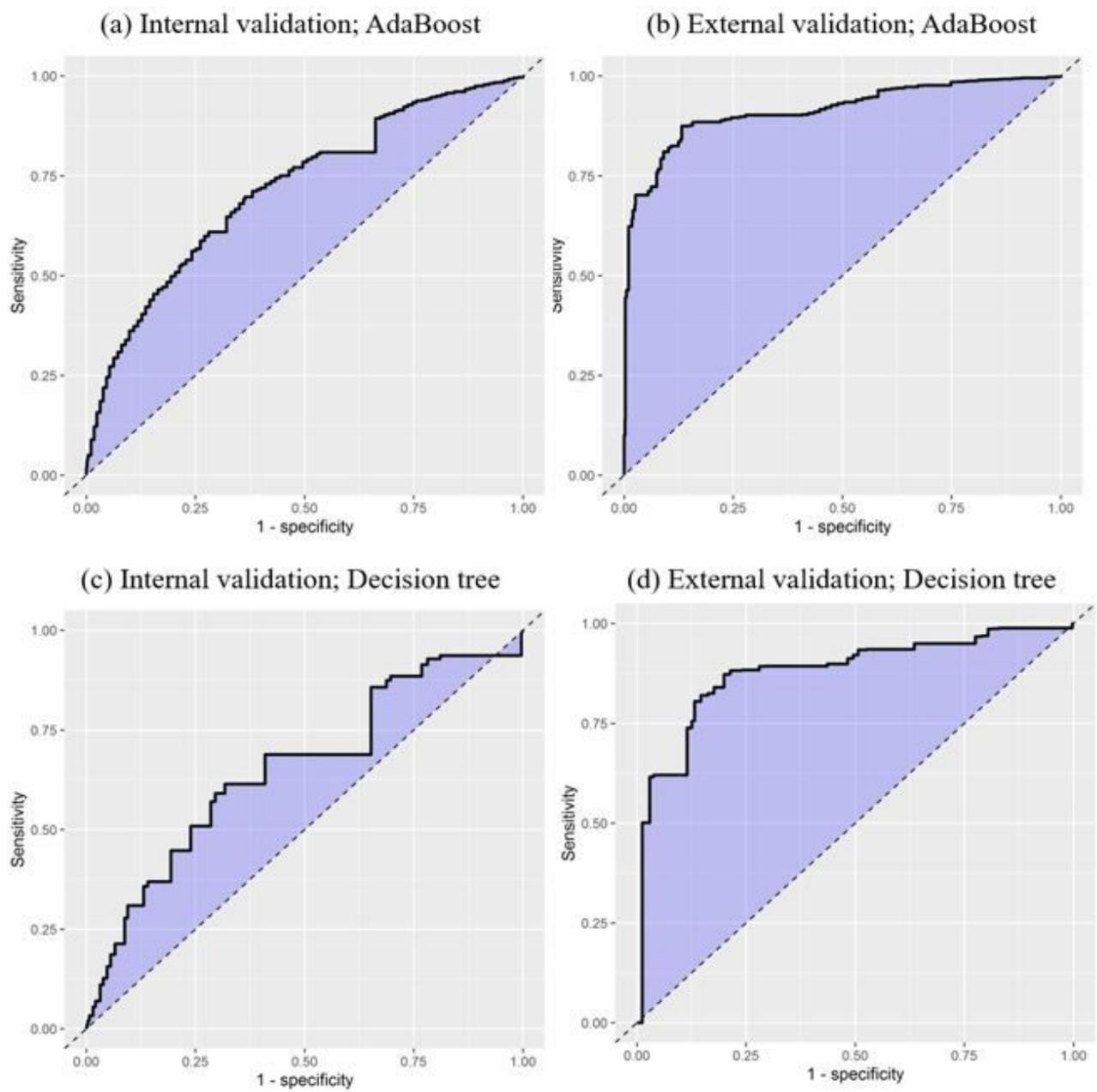
Figures

Figure 1

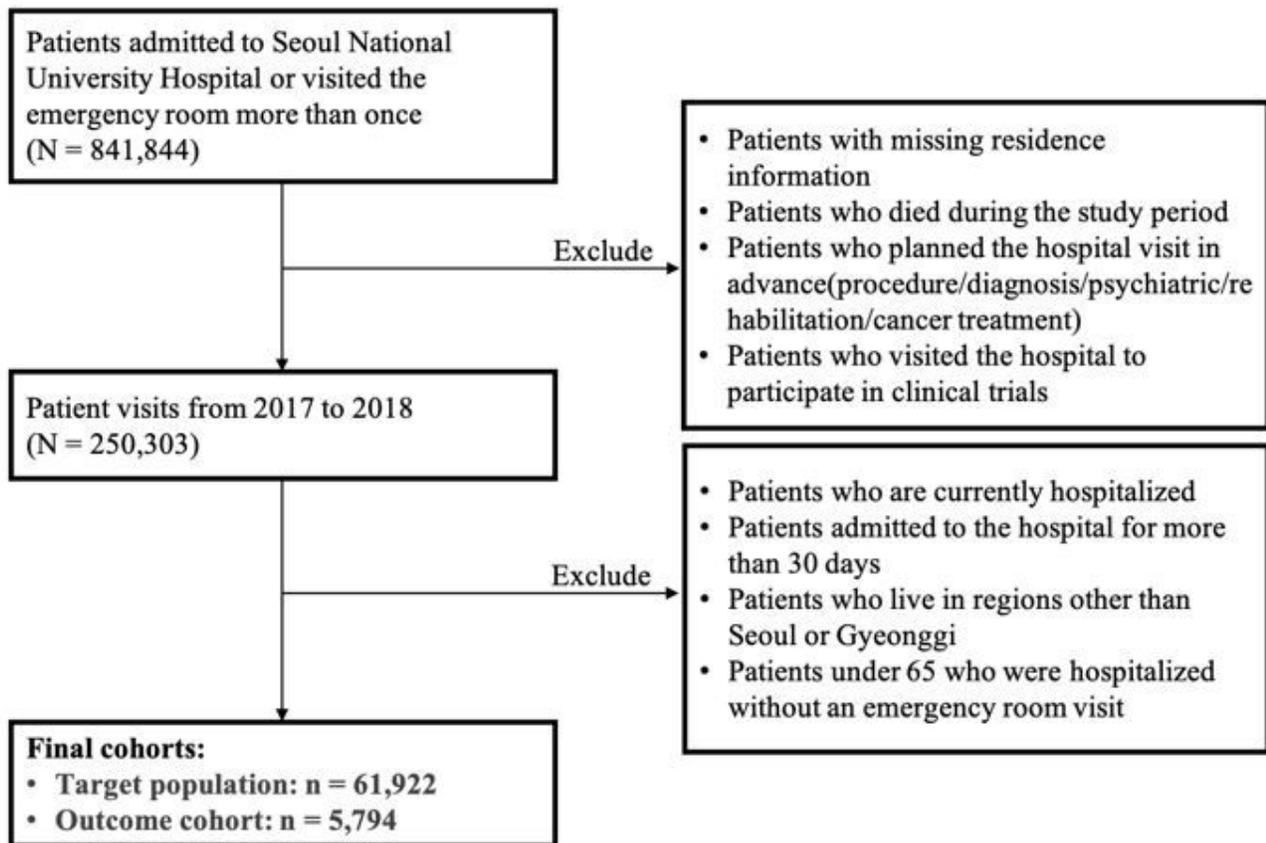ROC curves for the validation of the Adaboost and decision tree models.
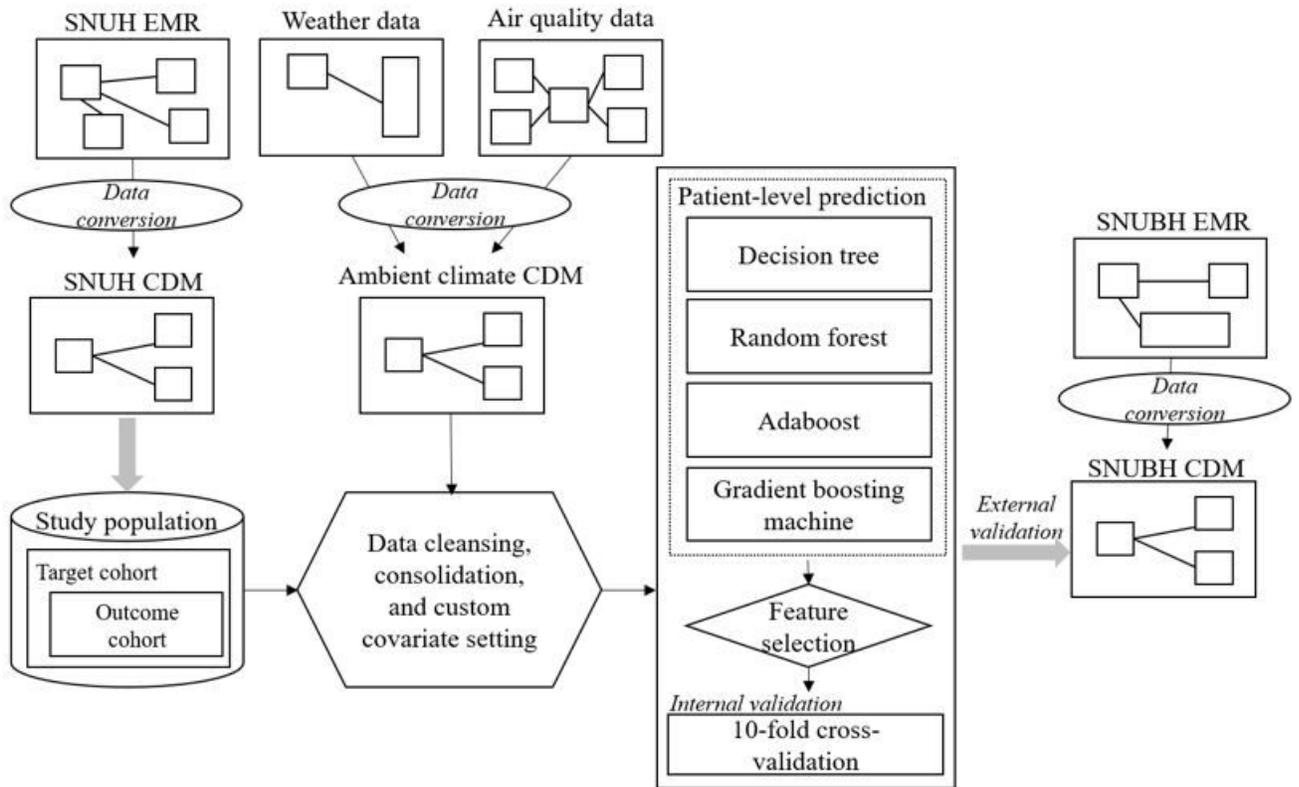
Figure 2

Study cohort design.

Figure 3

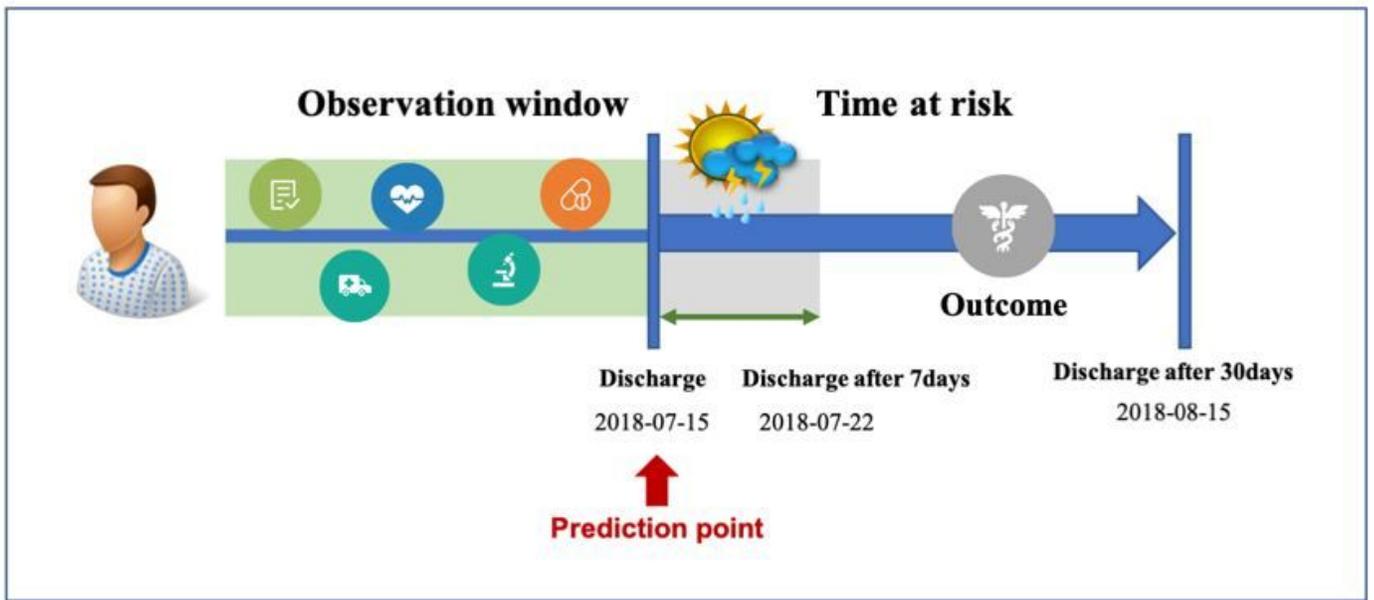Overall methodology of this study.

Figure 4

Prediction window.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesS1S2.docx
- SupplementaryTablesS3S6.xlsx
- SupplementaryTablesS1S2.docx
- SupplementaryTablesS3S6.xlsx