



HAL
open science

Machine learning with environmental predictors to forecast hospital visits and admissions: a systematic review

Levi Monteiro Martins, Elsa Coz, Delphine Maucort-Boulch, Mohand-Saïd Hacid

► **To cite this version:**

Levi Monteiro Martins, Elsa Coz, Delphine Maucort-Boulch, Mohand-Saïd Hacid. Machine learning with environmental predictors to forecast hospital visits and admissions: a systematic review. *Environmental Systems Research*, 2025, 14 (1), pp.12. <10.1186/s40068-025-00401-x>. <hal-05098915>

HAL Id: hal-05098915

<https://hal.science/hal-05098915v1>

Submitted on 5 Jun 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-ND 4.0 - Attribution - No Derivative Works - International License

REVIEW

Open Access



Machine learning with environmental predictors to forecast hospital visits and admissions: a systematic review

Levi Monteiro Martins^{1,2,3,4*} , Elsa Coz^{1,2,4} , Delphine Maucort-Boulch^{1,2,4}  and Mohand-Saïd Hacid^{1,3}

Abstract

Many studies have demonstrated a correlation between environmental monitoring data and healthcare service demand, highlighting its contribution to the global problem of emergency department crowding. To address this problem, forecasting models are essential for resource allocation and general management to improve patient outcomes. Machine Learning (ML), especially Deep Learning (DL), offers promise for forecasting patient volume. In this work, we present a systematic review of the use of ML to predict the health impacts of environmental exposures in the context of hospital visits and admissions. Standardized tools for conducting a systematic review were used, including the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS), and Prediction model Risk Of Bias ASsessment Tool (PROBAST). The search included studies from 2012 to 2025. We focus on answering how ML has been applied and what the major environmental predictors are used. As a result, 36 studies were retained from PubMed, Embase, and IEEE Xplore databases. We found that many studies exhibited a high risk of bias due to poor handling of missing values, inadequate outcome definitions, biased participant selection, and a low number of events per variable. Additionally, we found that the most used air pollutants and meteorological variables were PM_{2.5}, PM₁₀, NO₂, SO₂, CO, O₃, and temperature. Furthermore, the most common models were Random Forest and feed-forward neural networks. In addition, land use, remote sensing, demographic, and socioeconomic data offer promising avenues for improving model performance.

Keywords Air pollution, Deep learning, Hospital visits, Hospital admissions, Environmental predictors, Machine learning, Meteorological conditions

Introduction

Air pollution is one of the most considerable environmental risks to health. It is associated with many diseases and symptoms that affect health and quality of life regardless of location, socioeconomic status, and age. According to the World Health Organization (2021), in 2019, 99% of the global population was exposed to air pollution levels that increased their risk for diseases such as heart disease, stroke, chronic obstructive pulmonary disease, cancer, pneumonia, etc.

An important factor that affects air pollution concentration is meteorology. Due to the weather, pollutants released into the atmosphere can undergo different

*Correspondence:

Levi Monteiro Martins
levi.monteiro-martins@univ-lyon1.fr

¹ Université Claude Bernard Lyon 1, Villeurbanne cedex 69622, France

² Biostatistiques Santé, CNRS UMR 5558 Laboratoire de Biométrie et Biologie Évolutive, Villeurbanne cedex 69622, France

³ Base de Données, CNRS UMR 5205 Laboratoire d'Informatique en Image et Systèmes d'Information, Villeurbanne cedex 69622, France

⁴ Service de Biostatistique et Bioinformatique, Hospices Civils de Lyon, Lyon cedex 03 69424, France



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

chemical reactions, and they can be dispersed or concentrated in the air depending on the speed and wind direction (Ministère de la Transition écologique et de la Cohésion des territoires 2022). Exposure to these environmental factors can increase hospital visits and admissions.

This rise in hospital visits and admissions can lead to the worldwide problem of Emergency Department (ED) crowding, i.e., when the need for ED services exceeds the available resources (Schneider et al. 2003; Pearce et al. 2023). This problem is associated with various adverse outcomes, including service care delays, which can increase in-hospital mortality, low patient satisfaction, difficulties in clinician decision-making, equity in care, etc. (Chen et al. 2020).

Anticipating the volume of hospital visits and admissions can help reduce the burden on the healthcare system. Many studies using artificial intelligence (AI) in medical settings, such as bioinformatics, medical imaging, pervasive sensing, etc., are published yearly. Many of them highlight the ability of ML to abstract large, complex, and unstructured data, offering a robust way of analyzing heterogeneous data (Ravi et al. 2017). AI, such as machine learning (ML) and Deep Learning (DL), can help address this problem.

Previous studies have been presented in the setting of patient volume forecasting. Wargon et al. (2009) systematically reviewed nine studies. They aimed to present forecasting techniques to help ED physicians anticipate medical and nursing needs based on expected hospital visit volume to the ED. As a result, most of the models used were linear regression and traditional autoregressive time series models. Despite the broad search strategy adopted, no ML approaches were found, likely due to the low popularity of ML in earlier years (Mendo et al. 2021). In addition, two recent studies, Silva et al. (2023) and Jiang et al. (2022), systematically reviewed seven and thirty-five studies, respectively. Both studies investigated the factors affecting hospital visit forecasting, such as the predictors and the type of models used.

While the studies mentioned above provide important insights into understanding the use of different methods to forecast hospital visits, it is crucial to acknowledge certain limitations in the existing studies. One notable gap is the lack of research examining the usage of environmental predictors. To bridge this gap, we aim to conduct a systematic review to critically appraise studies that use ML and DL in the volumetric prediction of hospital visits and admissions, using well-established environmental predictors. We aimed to answer the following questions:

1. How has ML been applied to predict the number of hospital visits and admissions?
2. What are the major environmental predictors of hospital visits and admissions?

A narrative analysis of the included studies was conducted to explore the findings and guide decision-making.

Methodology

This systematic review was conducted in accordance with the PRISMA statement (Page et al. 2021) and the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) (Moons et al. 2014). The Risk of bias (RoB) and the concerns regarding the applicability were assessed based on the Prediction model Risk Of Bias ASsessment Tool (PROBAST) (Wolff et al. 2019). The review protocol is registered on the International Prospective Register of Systematic Reviews (PROSPERO) under the registration number: [CRD42023390400](https://www.crd42023390400).

Search strategy

We selected PubMed as the primary database since it is the preferred choice among researchers in health sciences and clinicians for conducting literature reviews (Lam and McDiarmid 2016; Vassar et al. 2017). To ensure comprehensive coverage, we also included Embase, as it has been demonstrated that Embase and PubMed complement each other in literature searches (Frandsen et al. 2021; Woods and Trewheellar 1998). Additionally, we included IEEE Xplore, which has been cited significantly more frequently than other Science, Technology, Engineering, and Mathematics (STEM) journal databases—such as the ACM Digital library, Engineering Village, and MathSciNet—between 2000 and 2019, indicating that computer scientists are frequently using this resource in research (Tomaszewski 2021). Combining these three databases ensures a multidisciplinary and comprehensive literature search, capturing both health, clinical, and technological perspectives relevant to our research questions.

In terms of the inclusion period, we opted for January 1st, 2012, to March 03rd, 2025, as it aligns with significant trends in the field. The increase in ML publications has been particularly pronounced since 2012, with a nearly sevenfold rise since 2015 (Nestor Maslej et al. 2024). By focusing on studies published during this period, we ensure that our review incorporates the most recent and relevant studies, reflecting the rapid advancements and evolving trends in ML applications. In addition, the search was based on all fields (e.g., titles, abstracts, metadata, and full-text). The search strategy for this systematic review was developed using medical

subject headings (MeSH) terms and text keywords related to hospital visits and admissions using ML and DL together with air pollution and weather environmental predictors. The search queries were developed in PubMed and adapted for the other databases. The queries are presented in Table 4 of Appendix A. Additional filters, such as “Journals and Conferences” were used for the search in the IEEE Xplore database. After retrieving relevant articles, we excluded duplicated and ineligible articles (Conference Abstracts, Tables of Contents, Conference Programs, Conference Guides, Organizing Committees, Conference Speakers, Indexes, and Editorial Boards) marked by automation tools developed in Python. Then, based on the inclusion criteria, we performed a manual screening. The screening consisted of two steps: (i) Title and abstracts, and (ii) Full-text screening of articles obtained if identified to meet criteria. Two reviewers, LMM and EC, independently performed this screening process. LMM screened 100% of the articles, and EC screened 20% of randomly selected articles. In cases where LMM and EC had conflicting screening decisions, the authors DMB and MSH were designated for conflict resolution.

Selection criteria

We included full-text English-language studies meeting the following criteria:

- Focus on hospital visits or admissions prediction;
- Use environmental predictors, including air pollution, meteorological factors, or both;
- Use ML models.

Reviews and surveys were excluded.

Summary measures

Knowing the number of outcome events relative to the number of independent variables analyzed in a model, also known as the Events Per Variable ratio (EPV), and the sample size is essential when mitigating the risk of over-fitting during the study design phase (Peduzzi et al. 1996). Thus, whenever the EPV was missing, we computed it by utilizing the time period covered by the data used in the study and the outcome granularity (e.g., daily, weekly, yearly) to estimate the total number of observations. Then, we divided this value by the total number of candidate predictors, yielding the estimated EPV (e-EPV).

Data extraction

From selected articles, we used CHARMS, a checklist designed to offer a systematic method for extracting data, ensuring a consistent capture of all pertinent information.

This checklist comprises 11 key domains: source of data, participants, outcome predicted, candidate predictors, sample size, missing data, model development, model performance, model evaluation, results, interpretation, and discussion (Moons et al. 2014). Additionally, the RoB was assessed using PROBAST (Wolff et al. 2019). The key domains adopted by PROBAST are aligned with the domains included in CHARMS. It contains four key domains: participants, predictors, outcome, and analysis. Data extraction and RoB were performed by a single reviewer (LMM) and supervised by the reviewer EC.

Results

The Fig. 1 shows the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flowchart summarizing the selection process. Following the database search, we retrieved 334 records. After automatic filtering, the remaining 301 records underwent title and abstract screening. One hundred ninety-eight records were excluded. The remaining 103 records underwent full-text screening. Sixty-seven articles were excluded since they did not match the inclusion criteria. No conflicts occurred during the screening process, making further consultation unnecessary. The remaining 36 articles were included in our analysis.

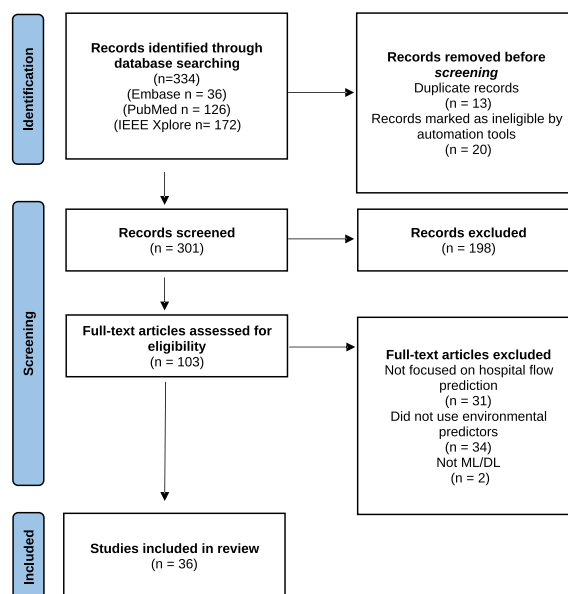


Fig. 1 The study selection process was performed via three steps: identification, screening, and inclusion. In the identification step, we searched three different databases. We removed duplicates automatically and removed records marked as ineligible by automation tools. Next, we screened the studies found using a two-stage process and removed studies using pre-defined selection criteria. The flowchart is adapted from the PRISMA statement

CHARMS extraction results

In this section, we show the results of data extraction using CHARMS (Moons et al. 2014).

Data sources, participants and outcomes

This review included 36 studies that used data from multiple existing registries (e.g., administrative or routine care hospital databases). Fig 2 shows the high heterogeneity of the data types used and how they were combined among different studies. These studies combined several types of data sources that we categorized into eleven

categories: air pollution, remote sensing, i.e., data collected without making physical contact with the object (e.g., satellite data), weather, calendar (e.g., day of the week, holidays, etc.), time series features (e.g., kurtosis, skewness, etc.), public health trends (e.g. influenza, acute diarrheea, respiratory infection, etc.), land use (e.g., energy, roads, built surface, etc.), hospital visits/beds, website searches, demographic-socioeconomic, and others not described by authors.

Regarding the study population, patient selection criteria included single or multiple reasons for hospital visits



Fig. 2 Distribution of predictor types used in each study. These types include air pollution, hospital visits and number of beds, weather, calendar, time series features, etc

and admissions. Most studies focused on patients diagnosed with respiratory and circulatory system diseases. In some studies, patients were selected based on diagnoses generated in the ED (e.g., ICD-10 diagnosis code), whereas in other studies, patients were selected based on the department to which they were admitted. The total count of studies for each subspecialty is presented in Fig. 3. We grouped them into four main categories: all subspecialties, diseases of the respiratory system, circulatory system, and other subspecialties.

Regarding geography, data from eighteen locations were found across studies; see Fig. 4. Most studies were located in China (28%), followed by the USA (14%).

The studies used datasets from 2001 to 2021. Table 1 presents the study period for each study. Most studies (28%) used data that began in 2015, and 19% had the end period in 2019. In most studies, the predicted outcome was the count of admissions or visits. However, seven studies (Khatri and Tamil 2017; Hovland et al. 2013; Qiu et al. 2020; Chen et al. 2019; Meng et al. 2015; Ram et al.

2015; Medina Quero et al. 2018) rather considered an ordinal outcome representing the hospital demand (e.g., normal, high, extreme). Additionally, the granularity of aggregation of the data (e.g., hourly, daily, etc.) and the forecast horizon (FH), i.e., the length of time forecasts are made, varied between studies (Table 1).

Ten studies forecasted ≤ 7 steps ahead, i.e., seven steps after the last input used for training (Sciannameo et al. 2022; Hu et al. 2020; Meng et al. 2015; Medina Quero et al. 2018; Moustris et al. 2012; de Souza Tadano et al. 2016; Mossop et al. 2023; Zhong et al. 2024; Kurucz et al. 2024; Susmann et al. 2025). Additionally, twelve studies forecasted ≥ 8 and ≤ 99 steps ahead (Usmani et al. 2021; Duwalage et al. 2020; Tuominen et al. 2022; Meng et al. 2015; Ram et al. 2015; Lin et al. 2021; Cengiz and Terzi 2012; Chou-Chen and Barboza 2024; Guo et al. 2024; Zhong et al. 2024; Mutlu et al. 2024; Kurucz et al. 2024). Furthermore, fourteen studies forecasted ≥ 100 and ≤ 365 steps ahead (Jing et al. 2020; Stonko et al. 2018; Zhang et al. 2018; Duwalage et al. 2020; Qiu et al.

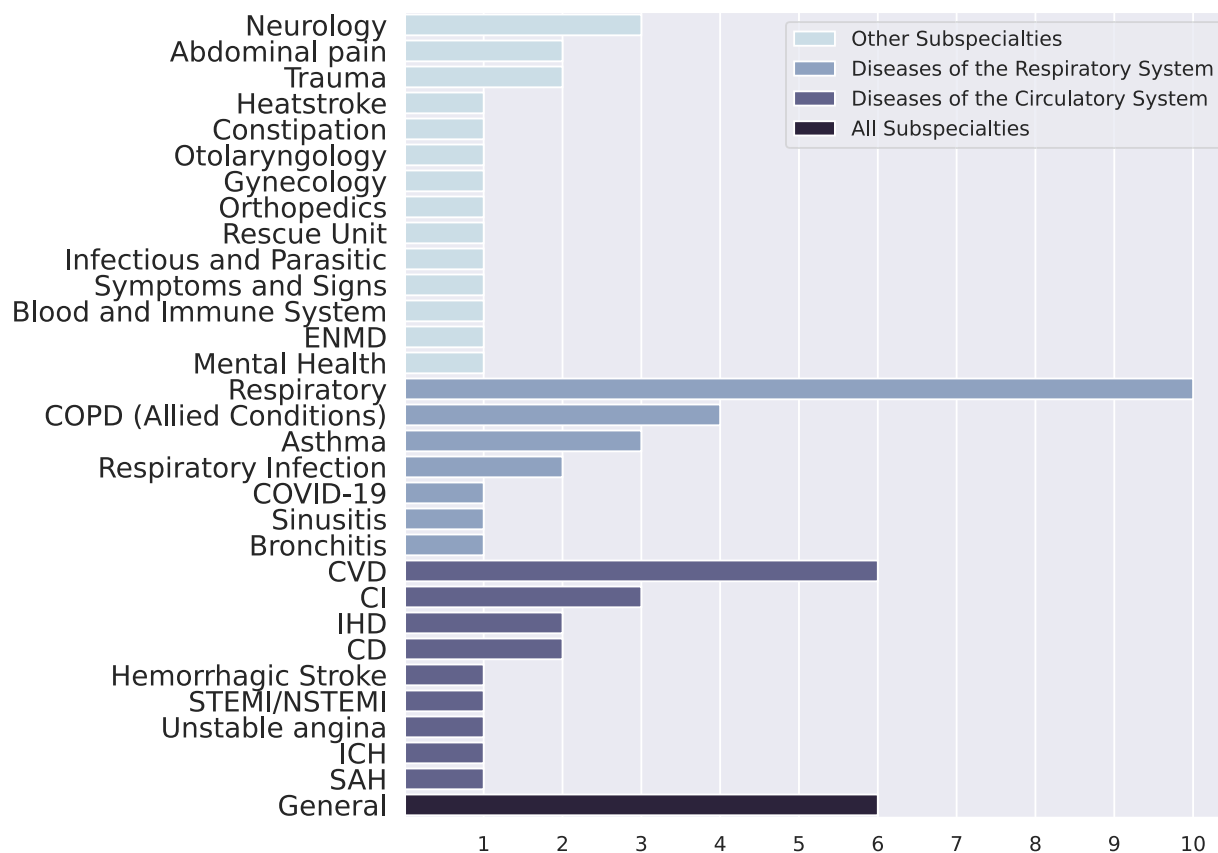


Fig. 3 Total count of studies for each medical subspecialty. Cardiovascular Disease (CVD), Chronic Obstructive Pulmonary Disease (COPD), Cerebrovascular Disease (CD), Cerebral Infarction (CI), Endocrine, nutritional, and metabolic disorder (ENMD), Ischemic Heart Disease (IHD), Intracerebral Hemorrhage (ICH), Subarachnoid Hemorrhage (SAH), non-ST-elevated Myocardial Infarction and ST-Elevated Myocardial Infarction (STEMI/NSTEMI)

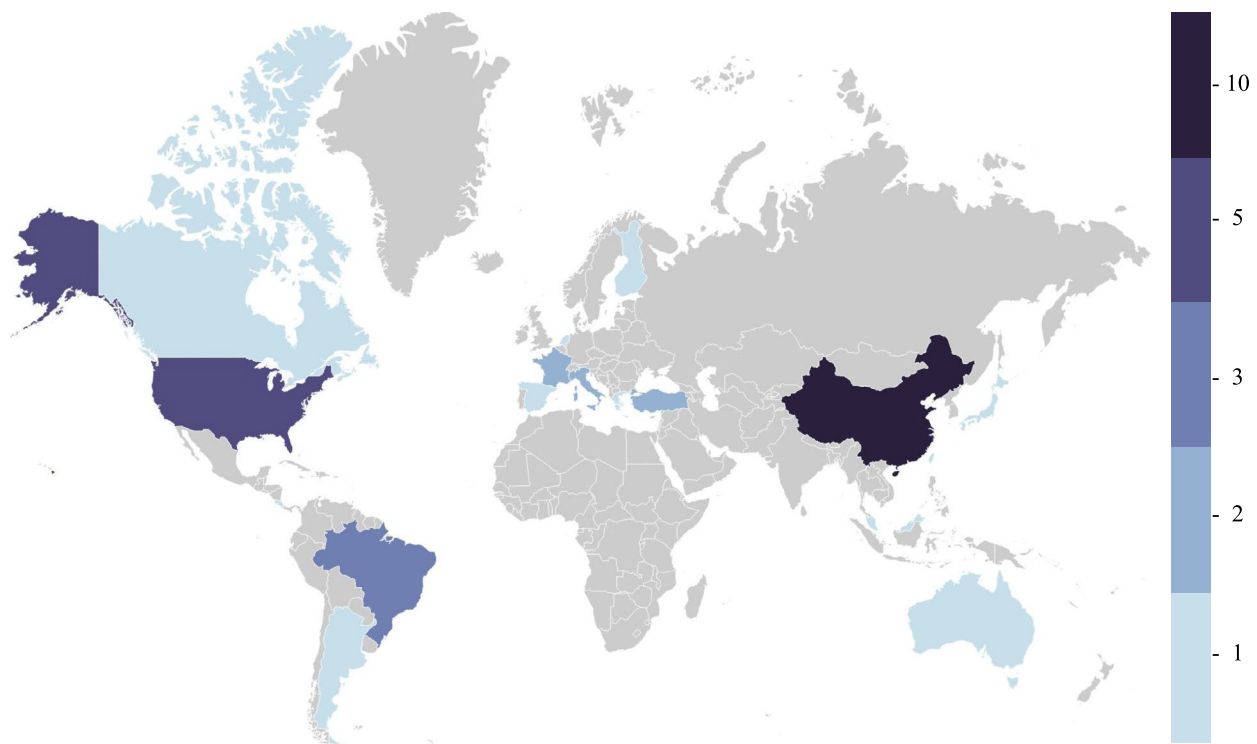


Fig. 4 Geographic distribution of the studies in this review. One article was found for each location: Argentina, Australia, Canada, Costa Rica, Finland, Greece, Japan, Malaysia, Singapore, Spain, Taiwan, and the Netherlands. Ten articles were located in Continental China, five in the USA, three in Brazil, and two in France, Italy, and Turkey

2020; Chen et al. 2019; Tavera Busso et al. 2021; Lin et al. 2021; Lu and Qiu 2023; Gafni-Pappas and Khan 2023; Song et al. 2022; Zhong et al. 2024; Yang et al. 2025; de Souza Tadano et al. 2024). Lastly, three studies forecasted ≥ 366 steps ahead (Hovland et al. 2013; Ogata et al. 2021; Cappelli et al. 2024). We could not determine the FH adopted in the studies (Khatri and Tamil 2017; Siqueira et al. 2023). Only two studies focused on hourly forecasting, seven on weekly forecasting, two on monthly forecasting, and twenty-eight on daily forecasting.

Candidate predictors and sample size

The number of candidate predictors, including feature-engineered (i.e., predictors generated through different transformations applied to the raw data) and lagged predictors (i.e., predictors derived from past values of a time series), ranged from 1 to 920. Studies employed candidate predictors within different ranges: 1–9 (Mossop et al. 2023; Stonko et al. 2018; Zhang et al. 2018; Huang et al. 2020; Siqueira et al. 2023; Meng et al. 2015; Usmani et al. 2021; Duwalage et al. 2020; Moustiris et al. 2012; Lin et al. 2021; Khatri and Tamil 2017; Hovland et al. 2013; Ram et al. 2015; Cengiz and Terzi 2012; de Souza Tadano et al. 2024; Mutlu et al. 2024; Cappelli et al. 2024), 12–33 (de Souza Tadano et al.

2016; Jing et al. 2020; Qiu et al. 2020; Gafni-Pappas and Khan 2023; Ogata et al. 2021; Medina Quero et al. 2018; Tuominen et al. 2022; Song et al. 2022; Chou-Chen and Barboza 2024; Zhong et al. 2024; Yang et al. 2025; Susmann et al. 2025), and 75–266 (Sciannameo et al. 2022; Lu and Qiu 2023; Chen et al. 2019; Tavera Busso et al. 2021; Guo et al. 2024; Kurucz et al. 2024). In addition, the study with the highest number of candidate predictors was the study presented by Hu et al. (2020), with 920 candidate predictors. These predictors were generated by extracting various time series features from eleven distinct time series, including mean, standard deviation, and kurtosis.

The most used air pollution candidate predictors were PM10, NO₂, SO₂, O₃, CO, and PM_{2.5}, mainly consisting of daily average values from one or multiple monitoring stations. We show the air pollutants and atmospheric predictors used in Fig. 5. Regarding the atmospheric candidate predictors, most studies used variables generated from temperature, including daily maximum, daily minimum, and daily mean values from one or more monitoring stations. Additionally, other candidate predictors, such as thermal amplitude, visibility, snow depth, significant tornado parameter, and daily planetary boundary layer, were observed in at least one study as part of the

Table 1 Characteristics of the included studies

Author (year)	Subspecialty	Location	Study population	Period	Granularity	FH	Lag	e-EPV	Model
Sciannameo et al. (2022)	COVID-19	Italy	3 110	2020–21	Daily	0–7	0 14 28	46	CNN-LSTM
Hu et al. (2020)	CVD IHD CD	China	NR	2015–17	Daily	7	1–28	1	Stacking Ensemble LR SVR XGBoost RF GBDT
Usmani et al. (2021)	CVD Respiratory diseases	Malaysia	NR	2006–16 2009–16 2010–16	Monthly	0–40 0–29 0–24	0	19 13 11	LSTM ELSTM ANN VAR
Jing et al. (2020)	Department of Trauma Otolaryngology Gynecology Abdominal Pain Orthopedics Neurology Comprehensive Department Rescue Unit	China	486 137	2015–18	Daily	1–210	0	81	GRU SARIMA
Stonko et al. (2018)	Trauma	USA	10 612	2013–16	Daily	1–164	0	274	ANN
Zhang et al. (2018)	General	China	NR	2012–13	Daily	1–184	0	183	ETS Hybrid ETS-LR Hybrid ETS-ANN
Khatri and Tamil (2017)	Asthma COPD Other respiratory conditions	USA	NR	2011–13	Daily	NR	3–7	200	MLP
Hovland et al. (2013)	COPD	USA	453 082	2001–06 2009–11	Daily	1–2 921	0	324	GLM MARS RF
Duwalage et al. (2020)	General	Australia	1 525 301	2009–15	Daily	14 90 365	0	365	GAM GLM MLR SARIMA RF
Tuominen et al. (2022)	General	Finland	386 579	2015–19	Daily	0–28	0–1	9	ARIMA ARIMAX Naïve Snaïve RF LMS RLS
Ogata et al. (2021)	Heatstroke	Japan	NR	2015–18	12-hours	1–730	0	110 1770	GLM GAM RF XGBoost
Qiu et al. (2020)	CVD	China	NR	2015–17	Daily	1–219	0–6	73	Logistic Regression SVM ANN RF XGBoost LightGBM

Table 1 (continued)

Author (year)	Subspecialty	Location	Study population	Period	Granularity	FH	Lag	e-EPV	Model
Chen et al. (2019)	Hemorrhagic Stroke	China	1 932	2015–17	Daily	1–111	1–14	139	Logistic Regression RF SVMLinear KNN XGBTree XGBLinear
Meng et al. (2015)	Neurology department: CI ICH SAH	China	2 180	2012–14	Daily	1–70 1–5	0	8 122	MLP
Tavera Busso et al. (2021)	Respiratory diseases	Argentina	1 612 049	2005–11	Daily	1–256	1–7	67	MLR GLM GAM DRF XRT GBM
Ram et al. (2015)	Asthma	USA	NR	2013	Daily	1–9	0	10	DT NB SVM ANN ANN (Adaptive Boosting) Stacking Ensemble (ANN, DT)
Huang et al. (2020)	COPD	China	20 000	2015–18	Weekly	1–16	0	39	MLWLR (Kernels: Gaussian, Exponential, and Laplacian) with K-means clustering LR RG LWLR DT
Medina Quero et al. (2018)	COPD	Spain	NR	2015–16	Daily	0 1	0	29	LSTM
Moustris et al. (2012)	Asthma	Greece	3 602	2001–04	Weekly	1	0	30	TLRN
de Souza Tadano et al. (2016)	Respiratory diseases	Brazil	NR	2007–09	Daily	1	0–7	219	ELM ESN MLP
Lin et al. (2021)	CVD	Taiwan	10 306	2009–13	Daily Weekly	1–365 1–52	0	37 260	LSTM
Lu and Qiu (2023)	CD Stroke	China	102 708	2015–18	Daily	1–365	1–6	69	RF GBDT RG ANN LSTM Stacking Ensemble (w/elastic net)
Mossop et al. (2023)	Mental Health	Canada	NR	2014–21	Weekly	1 1–4	0	2872 574	ARIMA SARIMA CNN LSTM BiLSTM Hybrid CNN-LSTM Hybrid CNN-BiLSTM Hybrid SCNN-BiLSTM

Table 1 (continued)

Author (year)	Subspecialty	Location	Study population	Period	Granularity	FH	Lag	e-EPV	Model
Siqueira et al. (2023)	Respiratory diseases	Brazil	175 949	2014–16 2017–19	Daily	NR	0-7	219	GLM ELM ESN Hybrid ELM-ESN
Gafni-Pappas and Khan (2023)	General	USA	NR	2017–19	Daily	1–365	0	55	ARIMA ETS FP RF XGBoost Hybrid FP-XGBoost
Song et al. (2022)	Respiratory diseases	China	NR	2016–19	Daily	1–365	3–7	162	ANN
Cengiz and Terzi (2012)	Respiratory diseases	Turkey	NR	2007–09	Weekly	1–26	0	52	GLM GAM ANN
Chou-Chen and Barboza (2024)	Respiratory diseases	Costa Rica	NR	2001–19	Weekly	1–52	0 1 2 10	40	RF Xgboost Facebook Prophet Ensemble
Guo et al. (2024)	Infectious and parasitic disease Respiratory infections	Singapore	NR	2009–18	Weekly	1–8	1–4	3	Naive AR RF GBM Hybrid LSTM-CNN Ensemble
Zhong et al. (2024)	CVD Respiratory diseases Neurological diseases Symptoms and signs Blood and immune system diseases Endocrine, nutritional, and metabolic disorders	France	58 691	2015–19	Hourly Daily Monthly	3 hours 6 hours 12 hours 1-365 days 1-12 months	0 1 2 3	166 5.45 3984	ARIMA GBM RNN LSTM
Yang et al. (2025)	CI	China	42 849	2016–20	Daily	1–313	1 2 3	104	RF LSTM
de Souza Tadano et al. (2024)	CVD Respiratory diseases	Brazil	NR	2011–13 2015–16	Daily	1–100	0 1	98	ELM ESN MLP RBF
Mutlu et al. (2024)	Asthma Bronchitis Sinusitis Respiratory infections	Turkey	19 674	2017–19	Daily	1–13	0 1	22	MLP ARIMA GLM
Kurucz et al. (2024)	IHD UA NSTEMI STEMI	Netherlands	214 953	2010–17	Daily	1 3 7 30	1–14	21	MLR RF

Table 1 (continued)

Author (year)	Subspecialty	Location	Study population	Period	Granularity	FH	Lag	e-EPV	Model
Susmann et al. (2025)	General	France	NR	2014–18	Daily	1	1 7	101	Naive BJML ARIMA ARIMAX QR QGAM GRF DRF GBM Ensemble
Cappelli et al. (2024)	CVD Respiratory diseases	Italy	NR	2013–21	Daily	1–657	0	470	RF

Artificial Neural Network (ANN), Autoregressive model (AR), Autoregressive Integrated Moving Average (ARIMA), Autoregressive integrated moving average with exogenous features (ARIMAX), Daily Minimal Bed Requirement (BJML; *Besoin Journalier Minimal en Lits*), Bidirectional Long Short-Term Memory (BiLSTM), Convolutional Neural Network (CNN), Convolutional Long-Short Term Memory (CNN-LSTM), Distributed Random Forest (DRF), Decision Trees (DT), estimated Events-per-variable (e-EPV), Extreme Learning Machines (ELM), Error-Trend-Seasonal (ETS), Echo State Networks (ESN), Generalized Additive Model (GAM), Gradient Boosting Machine (GBM), Gradient Boosting Decision Tree (GBDT), Generalized Linear Model (GLM), Gated Recurrent Unit (GRU), Generalized Random Forests (GRF), k-nearest neighbors algorithm (KNN), Linear Regression (Linear Regression), Long Short-Term Memory (LSTM), enhanced LSTM (ELSTM), Least Mean Squares (LMS), Locally Weighted Linear Regression (LWLR), Multiple Locally Weighted Linear Regression (MLWLR), Multivariate Adaptive Regression Splines (MARS), Multilayer Perceptron (MLP), Multiple Linear Regression (MLR), Not Reported (NR), Naive Bayes (NB), Recursive Least Squares (RLS), Random Forest (RF), Radial basis function network (RBF), Ridge Regression (RG), Recurrent neural networks (RNN), Quantile regression (QR), Quantile Generalized Additive Model (QGAM), Seasonal Autoregressive Integrated Moving Average (SARIMA), Seasonal Naive (Snaive), Support Vector Machine (SVM), Support Vector Regression (SVR), Stacked-dilated-causal BiLSTM (SCNN-BiLSTM), Time Lagged Recurrent Networks (TLRN), eXtreme Gradient Boosting (XGBoost), Extremely Randomized Trees (XRT), Vector Autoregression (VAR)

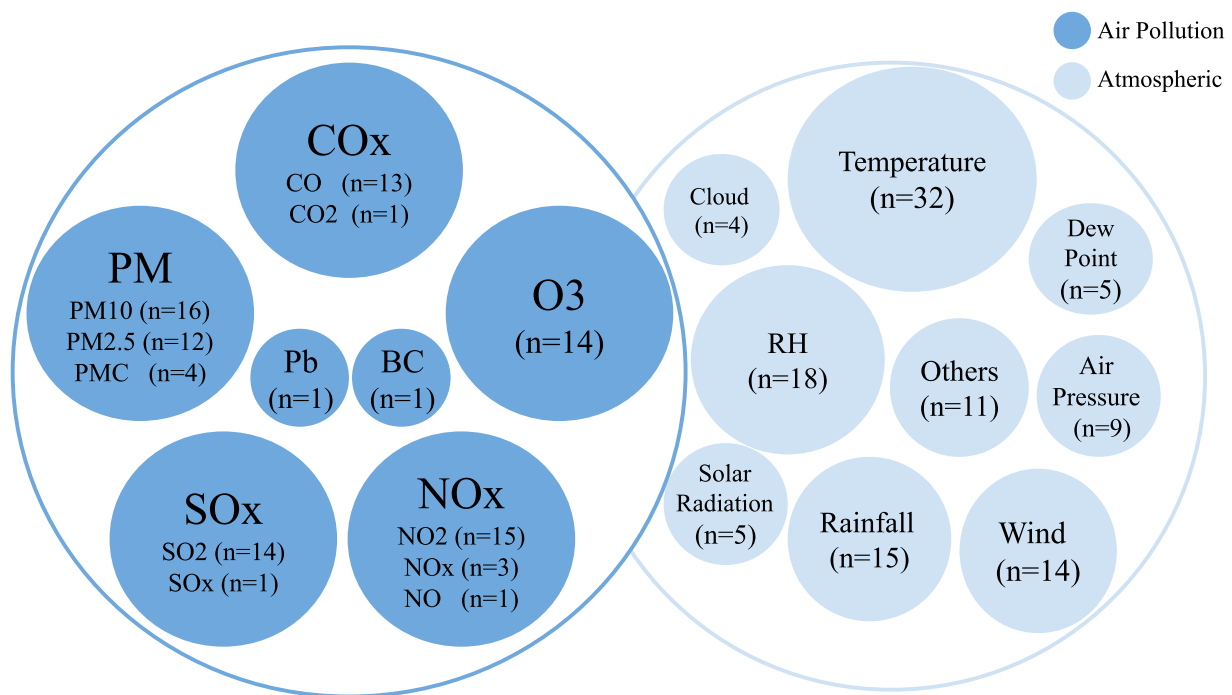


Fig. 5 Total count of the use of air pollution and atmospheric candidate predictors of studies

predictor set. Moreover, relative humidity, rainfall, wind, solar radiation, air pressure, dew point, and cloud were used as atmospheric candidate predictors. The Table 5 of Appendix B shows all the predictors used.

Nineteen studies used lagged predictors. The lags used ranged from 0 to 28 days across studies. Most studies included immediate lags and lags up to the seventh day (Table 1).

When it comes to data transformation, six studies used z-score normalization (Sciannameo et al. 2022; Hu et al. 2020; Ram et al. 2015; Siqueira et al. 2023; de Souza Tadano et al. 2024; Kurucz et al. 2024), four used logarithmic transformation (Zhang et al. 2018; Hovland et al. 2013; Huang et al. 2020; Chou-Chen and Barboza 2024), and four used min-max normalization (Hu et al. 2020; Lu and Qiu 2023; Yang et al. 2025; Mutlu et al. 2024). These transformations can approximate the data distribution to a normal distribution, and in some cases, they can reduce the variability of data (Feng et al. 2014). In one particular study (Ram et al. 2015), Twitter data was transformed by calculating the ratio of asthma-related tweets to the total number of tweets in the general Twitter stream collected from the same geographic area. This transformation was achieved using the term frequency-inverse document frequency algorithm (TF-IDF). One study (Song et al. 2022) used average sliding windows to reduce the noise in the data, one (Cappelli et al. 2024) used seasonal-trend decomposition using locally estimated scatterplot smoothing (STL), and one used quadratic (Cengiz and Terzi 2012) transformation. The others did not mention any data transformation.

The datasets used varied in time span and study population size (Fig. 6). The time span used for learning varied from 50 to 6939 days, which corresponds to 1–231 months and 0–19 years. Most studies used ≤ 1461 days of data (42%) and >2271 days (25%). Most studies did not report the study population size (50%). Moreover, considering the time in days, the estimated patient per day ratio is mostly ≤ 51 (11%), and ≤ 145 among studies (11%). Furthermore, concerning the sample size, most studies had e-EPV >20 and < 100 . Ten studies had an e-EPV ≤ 20 , six had e-EPV values ≥ 100 and ≤ 200 , and nine had e-EPV >200 .

Missing data

In most studies (61%) the missing values information was not reported (Sciannameo et al. 2022; Usmani et al. 2021; Jing et al. 2020; Zhang et al. 2018; Ogata et al. 2021; Chen et al. 2019; Meng et al. 2015; Tavera Busso et al. 2021; Ram et al. 2015; Huang et al. 2020; Medina Quero et al. 2018; Moustris et al. 2012; de Souza Tadano et al. 2016; Lin et al. 2021; Mossop et al. 2023; Siqueira et al. 2023; Cengiz and Terzi 2012; Chou-Chen and Barboza 2024; Guo et al. 2024; Zhong et al. 2024; Yang et al. 2025; Mutlu et al. 2024). Regarding data imputation, four studies (Hu et al. 2020; Qiu et al. 2020; Lu and Qiu 2023; Susmann et al. 2025) applied linear interpolation, and three studies (Stonko et al. 2018; Khatri and Tamil 2017; Susmann et al. 2025) removed records with missing values. One study (Hovland et al. 2013) used Holt-Winters smoothing forward and backward. One study (Duwalage et al. 2020) used KNN, and one study (Tuominen et al. 2022) imputed missing values for a predictor using a constant zero. Concerning the information on missing value proportions, only nine studies (Stonko et al. 2018; Khatri and Tamil 2017; Hovland et al. 2013; Duwalage et al. 2020; Qiu et al. 2020; Gafni-Pappas and Khan 2023; Kurucz et al. 2024; Susmann et al. 2025; Cappelli et al. 2024) mentioned the predictors with missing values and their respective percentages of missing values. In addition, one study (Hu et al. 2020) mentioned the overall percentage of missing values.

Model development

In Fig. 7, we present all models found in this systematic review. We classify the models into eight categories: regression, autoregression, ML, gradient boosting, feedforward neural networks (FNN), recurrent neural networks (RNN), hybrid, and others, and discuss their advantages and limitations in Section 3.2. ML and regression models were the categories most adopted across studies. Among the most frequently used models,

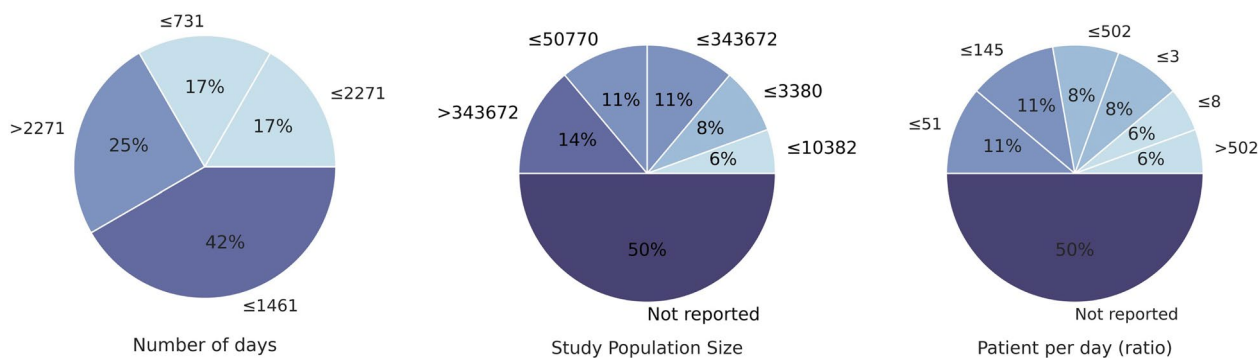


Fig. 6 Number of days, study population size, and patient per day ratio

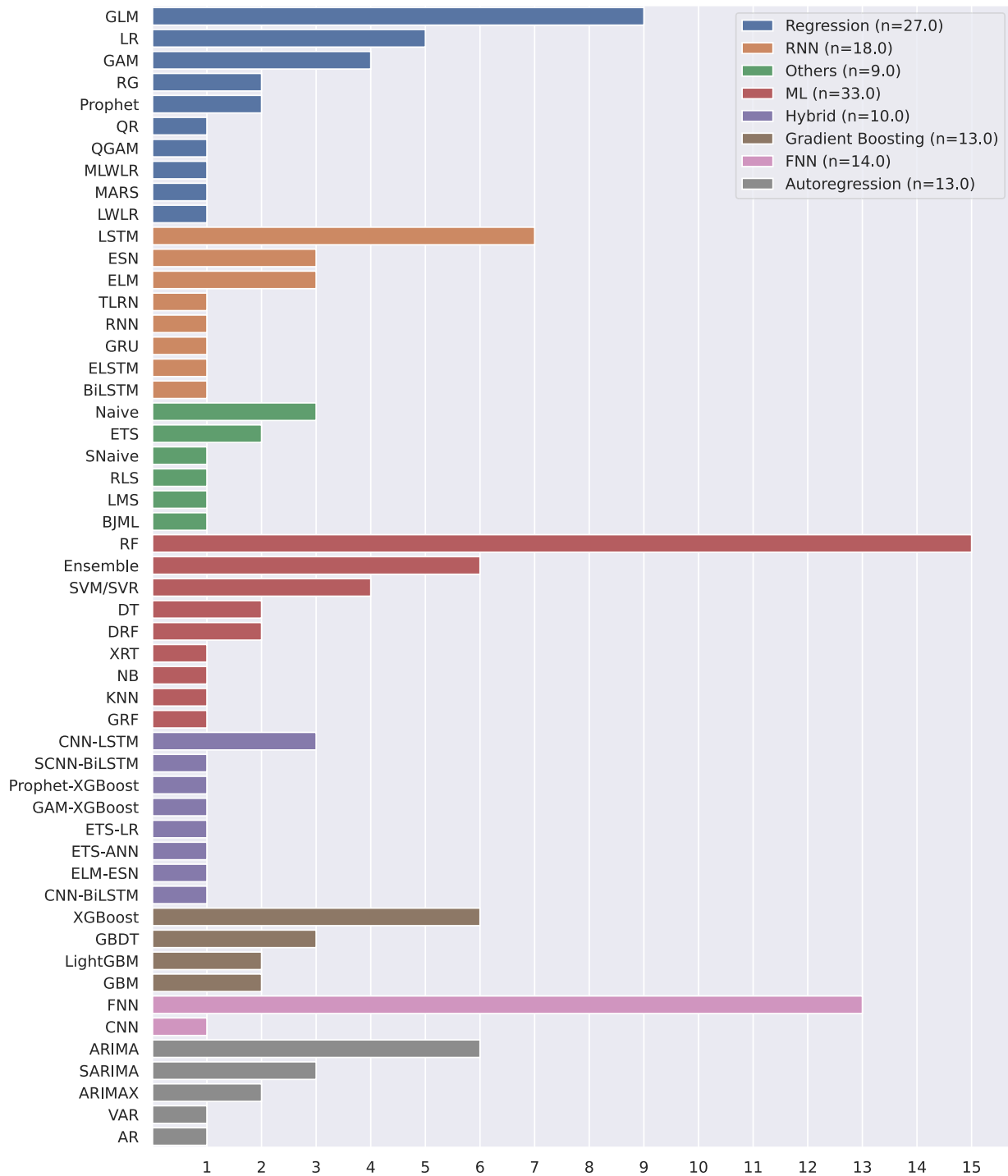


Fig. 7 Total count of model types used across studies

random forest (RF) was found in 15 studies, and FNN was used in 13 studies. In addition to FNN, we found a convolutional neural network (CNN) in one study. Furthermore, we found various derivations of DL such

as RNN, including echo state networks (ESN), gated recurrent unit (GRU), time-lagged recurrent networks (TLRN), long short-term memory (LSTM), enhanced LSTM (ELSTM), and bidirectional LSTM (BiLSTM).

Hybrid approaches such as CNN-LSTM, CNN-BiLSTM, and stacked CNN-BiLSTM (SCNN-BiLSTM) were also found.

The most commonly used models for each category were general linear model (GLM) for regression, LSTM for RNN, RF for ML, CNN-LSTM for a hybrid approach, XGBoost for gradient-boosting, conventional feed-forward neural networks for FNN, ARIMA, and SARIMA for autoregressive models. Naive and exponential smoothing (ETS) for other types of models.

A list of the models used in each study is presented in Table 1. In addition, Fig. 8 presents the total number of articles included in this systematic review grouped by the year of publication and the total number of models used grouped by year and model type. A peak of seven studies was observed in 2024, including different model types such as regression, autoregression, ML, gradient boosting, FNN, and RNN. In addition, we noticed the emergence of hybrid models in 2018, gradient boosting

in 2019, and autoregressive models in 2020. Other models, such as ML, FNN, and RNN, were used intermittently from 2012 to 2025. No studies were found for the year 2014.

Predictor selection, including lag range selection, varied across studies and was applied to select the most relevant features, reduce multicollinearity, and improve computational efficiency. Since most studies used more than one model, authors opted for using one or more predictor selection approaches. Five studies used Pearson correlation (Tavera Busso et al. 2021; Ram et al. 2015; Huang et al. 2020; Medina Quero et al. 2018; Cappelli et al. 2024), two used Chi-squared test (Hovland et al. 2013; Tavera Busso et al. 2021), five used least absolute shrinkage and selection operator (LASSO) (Chen et al. 2019; Tavera Busso et al. 2021; Gafni-Pappas and Khan 2023; Guo et al. 2024; Kurucz et al. 2024), one used Spearman’s correlation (Tavera Busso et al. 2021). In addition to the Chi-squared test, Hovland

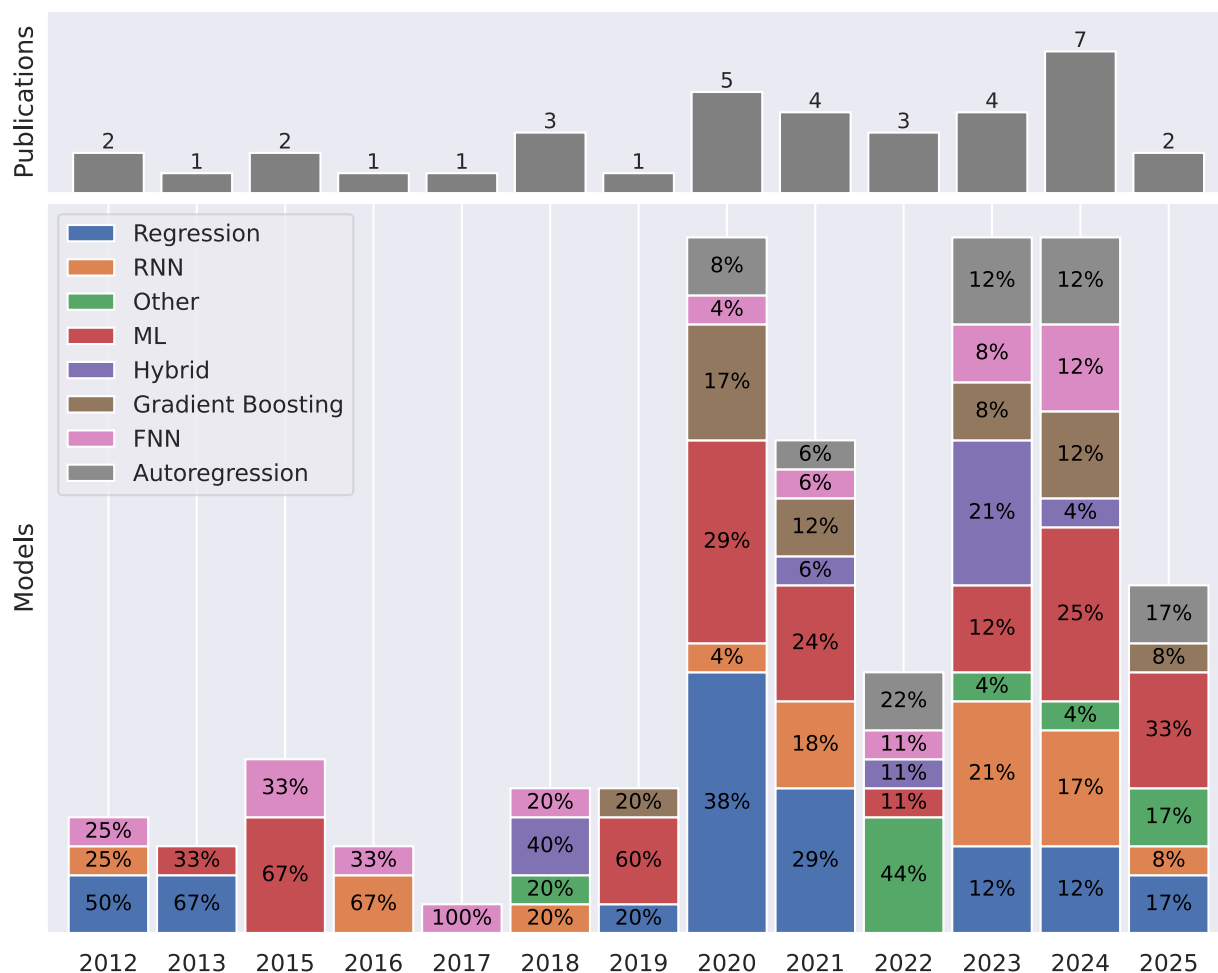


Fig. 8 Publications by year, and the total of model types used grouped by year and model type

et al. (2013) used GLM Maximum likelihood tests to identify significant factors ($p < 0.05$) and RF feature importance based on the Gini criterion. Moreover, Song et al. (2022) used the results of DL models to determine the correlation between multiple combinations between two predictors, and Qiu et al. (2020) used a generalized additive model (GAM) with generalized cross-validation to select the lag range.

The study presented by Hu et al. (2020) used multivariate adaptive regression splines (MARS) internal test to identify the important predictors and early stopping method for sequential forward floating selection (SFFS). Tavera Busso et al. (2021), when using GAM, the final model was obtained after excluding predictors with a p -value > 0.1 . Tuominen et al. (2022) used simulated annealing (SA) and floating search (FS). In the study presented by Ogata et al. (2021), optimal predictors were selected through recursive feature elimination, i.e., by removing the weakest features until a certain number of features was reached. Ram et al. (2015) used backward feature selection. Six studies (Sciannameo et al. 2022; Jing et al. 2020; Zhang et al. 2018; Lin et al. 2021; Mossop et al. 2023; de Souza Tadano et al. 2024) conducted multiple experiments to find the best combination of predictors, optimal lag or time steps, and the best model structure (e.g., the best number of layers in an artificial neural network). Other studies did not report predictor selection methods.

Model performance, evaluation, and results

The evaluation of models relied on several metrics. Studies usually used more than one evaluation metric. In Fig. 9, we present how these metrics have been used across studies. For regression tasks, the RMSE and MAPE were the most used metrics. For classification tasks, the predominant metrics used were accuracy, the area under the curve (AUC) or the area under the receiver operating characteristic curve (ROC), sensitivity, and precision.

All the studies found in this review focused on developing prediction models with internal validation, i.e., through some form of re-sampling technique such as cross-validation, bootstrapping, etc. Seventeen studies (Usmani et al. 2021; Zhang et al. 2018; Khatri and Tamil 2017; Hovland et al. 2013; Duwalage et al. 2020; Tuominen et al. 2022; Tavera Busso et al. 2021; Moustris et al. 2012; Lin et al. 2021; Siqueira et al. 2023; Song et al. 2022; Cengiz and Terzi 2012; Chou-Chen and Barboza 2024; Zhong et al. 2024; Yang et al. 2025; Kurucz et al. 2024; Cappelli et al. 2024) adopted a simple train and test split. Six (Sciannameo et al. 2022; Stonko et al. 2018; de Souza Tadano et al. 2016; Meng et al. 2015; de Souza Tadano et al. 2024; Mutlu et al. 2024) used a train, validation, and test split. Ten (Jing et al. 2020; Chen et al. 2019; Ram et al. 2015;

Huang et al. 2020; Medina Quero et al. 2018; Hu et al. 2020; Ogata et al. 2021; Qiu et al. 2020; Lu and Qiu 2023; Gafni-Pappas and Khan 2023) used k-fold cross-validation. The predominant choice for the number of validation folds in the studies was 5 and 10. Study (Qiu et al. 2020) used Stratified Random Sampling when using cross-validation. Seven studies (Sciannameo et al. 2022; Duwalage et al. 2020; Tuominen et al. 2022; Medina Quero et al. 2018; Mossop et al. 2023; Guo et al. 2024; Susmann et al. 2025) reported using rolling origin forecasting as a training strategy. This approach consists of successively updating the point from which forecasts are generated, also known as the forecasting origin, to produce forecasts from the new origin (Tashman 2000). Among them, Medina Quero et al. (2018) adopted a sliding window of size 30 days.

Regarding the results presented in each study, comparing the different models used was impossible due to differences in the datasets, train/test/validation splits adopted, FHs, study population sizes, and evaluation metrics. However, in Fig. 10, we present the models most frequently selected in each study for studies with daily granularity, along with their respective FHs. For longer FHs, RF and GLM were commonly selected. For FHs between 32 and 365 days, GRU and FNN appeared frequently. For FHs between 8 and 31 days, FNN and LR were often selected. Finally, for FHs of 7 days or less, LR and Ensemble models were the most selected.

Interpretation and discussion

Fifteen studies (Hu et al. 2020; de Souza Tadano et al. 2016; Siqueira et al. 2023; Song et al. 2022; Jing et al. 2020; Stonko et al. 2018; Zhang et al. 2018; Hovland et al. 2013; Ram et al. 2015; Medina Quero et al. 2018; Lin et al. 2021; Mossop et al. 2023; Gafni-Pappas and Khan 2023; Cengiz and Terzi 2012; Zhong et al. 2024) did not discuss comparisons with other studies. Among them, five (Hu et al. 2020; de Souza Tadano et al. 2016; Siqueira et al. 2023; Song et al. 2022; Cengiz and Terzi 2012) did not discuss study limitations.

Availability of code and dataset

The lack of research reproducibility in various domains of study poses a challenge to generating reliable knowledge bases to support decision-making. This section lists the code and dataset availability for each study included in this systematic review.

Only seven studies (Table 2) made their code available (Duwalage et al. 2020; Tuominen et al. 2022; Ogata et al. 2021; Siqueira et al. 2023; Chou-Chen and Barboza 2024; Guo et al. 2024; Susmann et al. 2025). The link provided by Duwalage et al. (2020) to access the code implementation is no longer accessible.

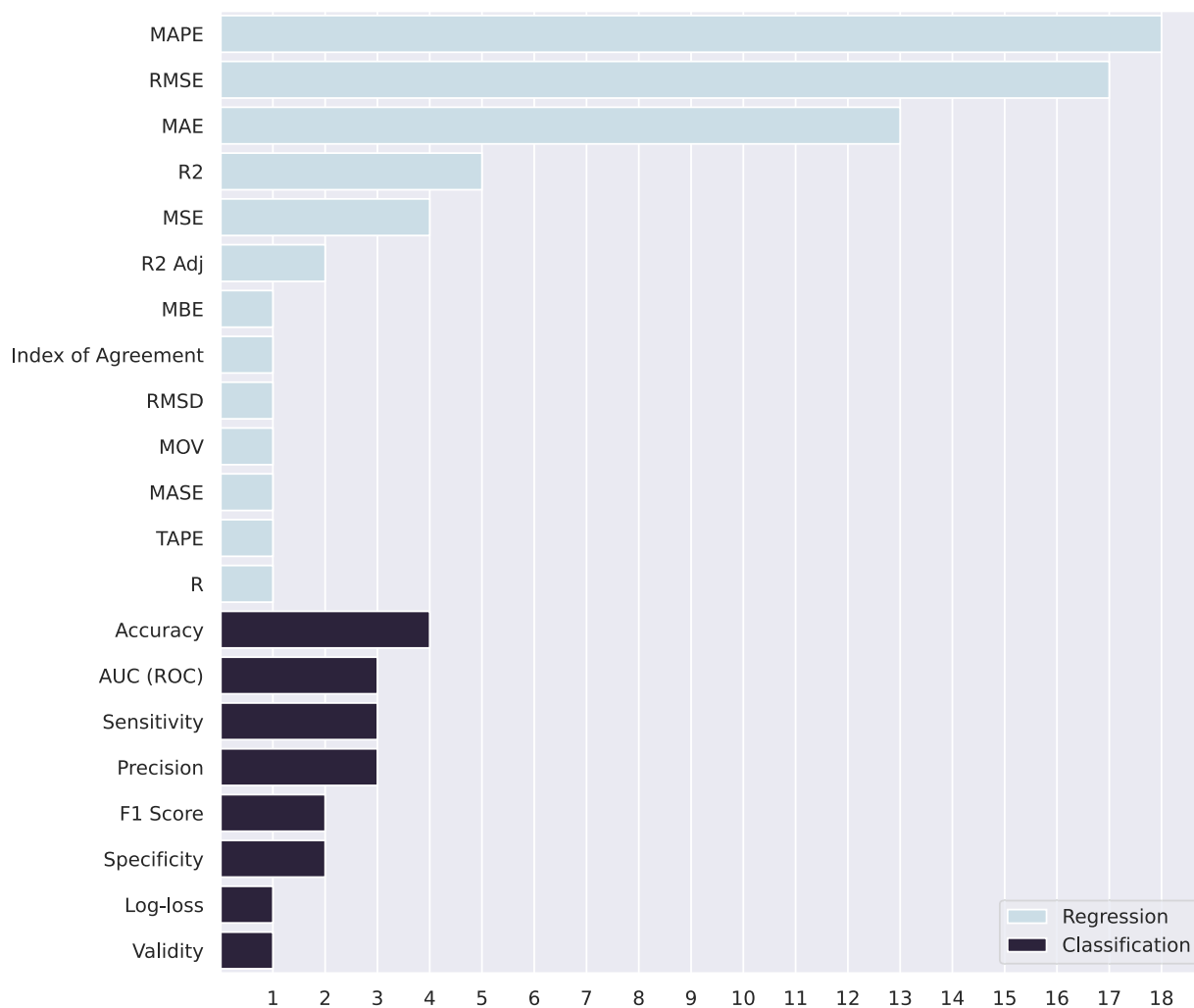


Fig. 9 Total count of evaluation metrics used across studies. Root-mean-square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), mean squared error (MSE), mean bias error (MBE), residual mean square deviation (RMSD), mean observed value (MOV), mean absolute scaled error (MASE), total absolute percentage error (TAPE), correlation coefficient (R), R-squared (R^2), and adjusted R-squared (R^2_{adj})

Five of the studies (Moustris et al. 2012; Lin et al. 2021; Song et al. 2022; Cengiz and Terzi 2012; Mutlu et al. 2024) did not provide the air pollution datasets used. One study (Chou-Chen and Barboza 2024) provided direct access to air pollution data. The remaining studies shared information about public data providers, allowing for extraction or requesting access to the air pollution datasets. It should be noted that the air pollution datasets provided by Usmani et al. (2021); de Souza Tadano et al. (2016) are no longer accessible through the links provided by the authors.

Among the studies, eight (Chen et al. 2019; Meng et al. 2015; Moustris et al. 2012; Lin et al. 2021; Mossop et al. 2023; Song et al. 2022; Yang et al. 2025; Mutlu et al. 2024) did not share the atmospheric data used,

and the atmospheric data provided by de Souza Tadano et al. (2016) was no longer accessible. Five studies (Ogata et al. 2021; Siqueira et al. 2023; de Souza Tadano et al. 2024; Kurucz et al. 2024; Cappelli et al. 2024) stated that their atmospheric data are accessible on reasonable request to authors. All the other studies provided public data provider information, enabling the extraction or request for accessing the atmospheric dataset.

Regarding hospital data, only two studies (Tuominen et al. 2022; Chou-Chen and Barboza 2024) shared their dataset with direct access. Six studies (Ogata et al. 2021; Qiu et al. 2020; Siqueira et al. 2023; de Souza Tadano et al. 2024; Kurucz et al. 2024; Cappelli et al. 2024) mentioned that their hospital datasets are available upon reasonable

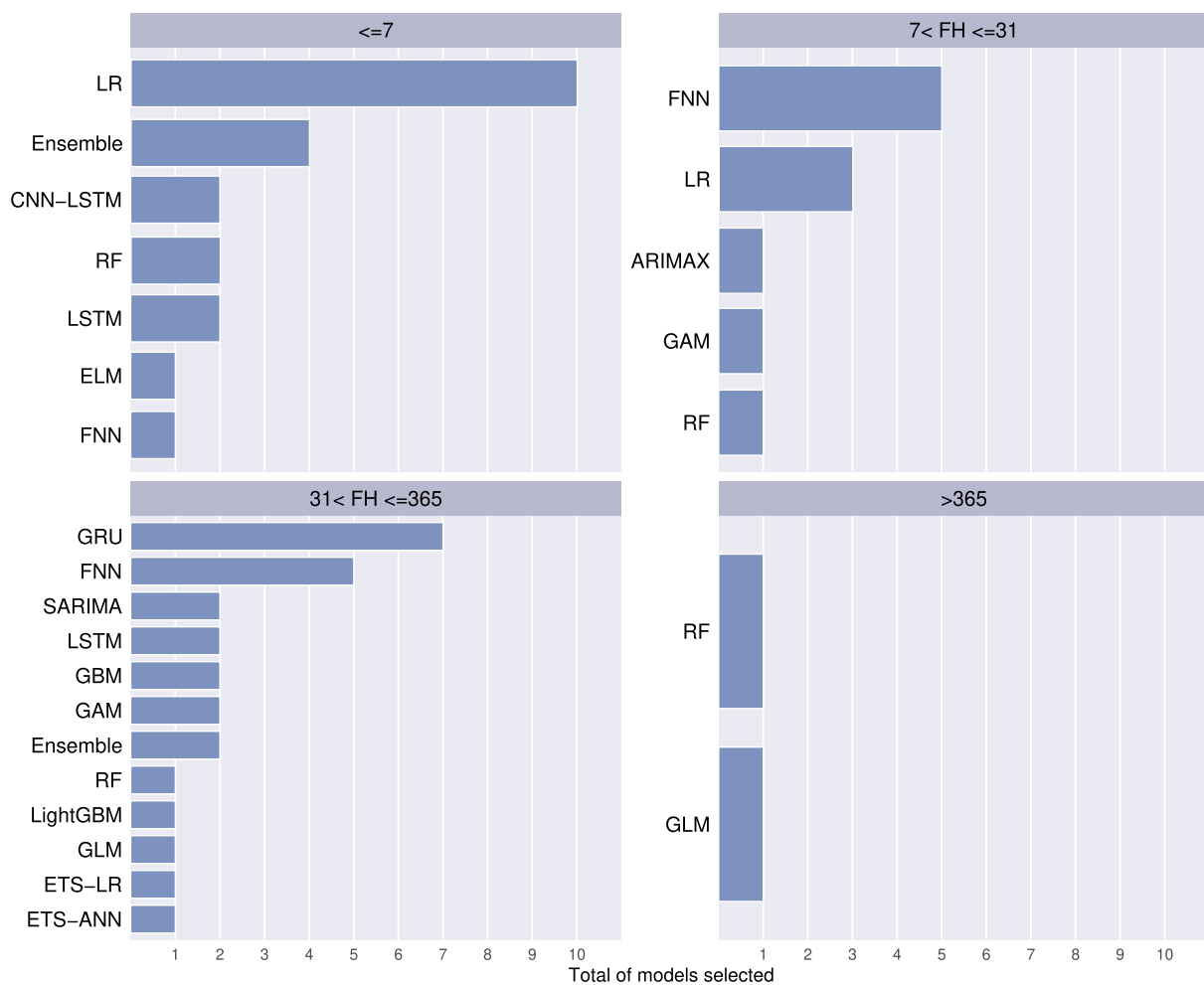


Fig. 10 Models most frequently selected in each study and their respective FHs, for studies having daily granularity

request. The other three studies (de Souza Tadano et al. 2016; Lin et al. 2021; Lu and Qiu 2023) provided public data provider information, allowing for the extraction or requesting access to the hospital data.

While making the code and data available to the public may not be a sufficient guarantee of reproducibility and transparency in research, this can still contribute to mitigating the problem of research reproducibility.

Model taxonomy

The models found in this systematic review were listed in the CHARMS section (3.1.4). This section presents their advantages and limitations and offers recommendations for their application.

Simple forecasting methods

The use of simpler forecasting methods such as the Naïve and SNaïve, as adopted by Tuominen et al. (2022), was

mainly to establish a baseline when conducting model benchmarking. Both methods are univariate and consist of the direct use of past values as the forecast. These are methods commonly used for non-stationary time series, i.e., a time series whose statistical properties depend on the time the series is observed. In the Naïve method, also known as the random walk model, all forecasts are set to the value of the last observation. Similarly, the SNaïve method sets the forecast values to be equal to the last value observed from the same season, being mostly recommended when data is highly seasonal (Rob J Hyndman and George Athanasopoulos 2018d). These approaches incorporate the notion of volatility, given that their forecasts have the same level of variation as the time series used for forecasting. Their use is recommended when the successive observations are similar. However, a limitation of these methods is that they do not consider other factors that could influence the volatility (Morlidge 2013).

Table 2 Availability of codes and datasets

Study	Code	Air pollution	Atmospheric	Hospital
Sciannameo et al. (2022)	No	Yes [^]	Yes [^]	No
Hu et al. (2020)	No	Yes [^]	Yes [^]	No
Usmani et al. (2021)	No	Yes [~]	–	No
Jing et al. (2020)	No	Yes [^]	Yes [^]	No
Stonko et al. (2018)	No	–	Yes [^]	No
Zhang et al. (2018)	No	–	Yes [^]	No
Khatri and Tamil (2017)	No	Yes [^]	Yes [^]	No
Hovland et al. (2013)	No	Yes [^]	Yes [^]	No
Duwalage et al. (2020)	Yes [~]	–	Yes [^]	No
Tuominen et al. (2022)	Yes	–	Yes [^]	Yes
Ogata et al. (2021)	Yes	–	Yes [*]	Yes [*]
Qiu et al. (2020)	No	Yes [^]	Yes [^]	Yes [*]
Chen et al. (2019)	No	Yes [^]	No	No
Meng et al. (2015)	No	–	No	No
Tavera Busso et al. (2021)	No	Yes [^]	Yes [^]	No
Ram et al. (2015)	No	Yes [^]	–	No
Huang et al. (2020)	No	Yes [^]	–	No
Medina Quero et al. (2018)	No	Yes [^]	Yes [^]	No
Moustris et al. (2012)	No	No	No	No
de Souza Tadano et al. (2016)	No	Yes [~]	Yes [~]	Yes [^]
Lin et al. (2021)	No	No	No	Yes [^]
Lu and Qiu (2023)	No	Yes [^]	Yes [^]	Yes [^]
Mossop et al. (2023)	No	–	No	No
Siqueira et al. (2023)	Yes [*]	Yes [*]	Yes [*]	Yes [*]
Gafni-Pappas and Khan (2023)	No	Yes [^]	Yes [^]	No
Song et al. (2022)	No	No	No	No
Cengiz and Terzi (2012)	No	No	–	No
Chou-Chen and Barboza (2024)	Yes	Yes	Yes	Yes
Guo et al. (2024)	Yes	Yes [^]	Yes [^]	No
Zhong et al. (2024)	No	Yes [^]	Yes [^]	No
Yang et al. (2025)	No	–	No	No
de Souza Tadano et al. (2024)	No	Yes [*]	Yes [*]	Yes [*]
Mutlu et al. (2024)	No	No	No	No
Kurucz et al. (2024)	No	–	Yes [*]	Yes [*]
Susmann et al. (2025)	Yes	Yes [^]	Yes [^]	No
Cappelli et al. (2024)	No	Yes [*]	Yes [*]	Yes [*]

(~) The link provided in the paper no longer works

(*) Available on request

(-) Not used in the study

(^) The data is accessible through online extraction or a request to be made through data provider platforms

Another forecasting method found in this systematic review is the Error-Trend-Seasonal Model (ETS), also known as Exponential Smoothing. In this method, forecasts are weighted combinations of past observations, with recent observations given relatively more weight than older observations (Hyndman et al. 2008). In addition, this method varies according to the combination of the time series components: error, trend,

and seasonal, which can be either additive, multiplicative or none. One of the advantages of using ETS models is their ability to generate prediction intervals (Rob J Hyndman and George Athanasopoulos 2018c) and their flexibility to define how each component interacts with each other (Svetunkov 2024). Nevertheless, these approaches are limited when dealing with time series with multiple seasonal patterns, usually

observed in hospital visits and admissions. For example, daily data might exhibit both a weekly pattern and an annual pattern (Hyndman et al. 2008; Rob J Hyndman and George Athanasopoulos 2018b). Furthermore, ETS can face numerical difficulties depending on the combination of components. This may occur when ETS models with additive errors and either a multiplicative trend or multiplicative seasonality require division by a state component, which could lead to a division by zero (Hyndman et al. 2008).

Furthermore, two adaptive (time-varying) signal processing methods were found, LMS and RLS. They were used due to the non-stationary nature of the data. The LMS is a stochastic gradient descent method known for being simple yet capable of achieving good prediction performance. However, this approach has limitations, such as a slow rate of convergence and sensitivity to the variations in the condition number of the correlation matrix of the input signal vector (Haykin 2002). The RLS is based on the method of least squares, which can be formulated with recursive estimation. RLS overcomes the convergence time problem of LMS. Both approaches are linear and can yield suboptimal solutions when input data is non-linear. When deciding which adaptive signal processing method to use, it is recommended to start with LMS (Haykin 2002).

Regression models

In time series regression, we assume a linear relationship between the forecast variable and the predictor variables. This type of model can capture important relationships between the forecast variable and the predictor variable. In this systematic review, different regression models were found. One was the LR model, which predicts the target variable as a weighted sum of the predictors. The advantage of using LR is that interpretation is easy, and the estimation procedure is simple. However, LR models assume that the target outcome follows a normal distribution. Therefore, if the outcome does not follow a normal distribution, it would require transformation.

Many regression models are available to deal with the violations of assumptions imposed by LR models. Deciding which one to use will depend on the data used, its distribution, relationship type among predictors and target variables, dimensionality, and the nature of the problem (classification/regression). This study found the use of GLM, GAM, MARS, QR, QGAM, and LWRL. GLM enables the modeling of a Poisson distribution, which is usually more appropriate for modeling count data than a normal distribution. In addition, logistic regression, a GLM model, can be used when the outcome follows a Bernoulli distribution. Furthermore, when modeling more

complex relationships, spline-based approaches such as MARS and GAM can be used. QR and QGAM are regression models that, instead of focusing only on the mean relationship between variables, examine how different parts of a distribution are affected by other factors (Koenker 2005; Fasiolo et al. 2020).

MARS models are adapted to deal with non-linear relationships in the data. These models assess each data point for each predictor as a knot and create a linear regression model with the candidate features (Boehmke and Greenwell 2019). Unlike GAM, which models the outcome as the sum of smooth functions of each feature using a fixed spline design with pre-specified knot locations, MARS is non-parametric. It automatically models variable interactions with automatic knot location selection (Hastie and Tibshirani 1986). In addition, MARS may offer advantages when working with high-dimensional data compared to GAM (Stone et al. 1997; Leathwick et al. 2006), and it can be faster to train (Moisen and Frescino 2002).

Another forecasting method found in this review was the locally weighted linear regression (LWLR). The study presented by Huang et al. (2020) used LWLR and other variations and concluded that LWLR provides better local feature selectivity and higher prediction accuracy when compared to the LR model. In addition, LWLR was found to be competitive with FNN (Atkeson et al. 1997). The LWLR is a non-parametric memory-based k-nearest-neighbor-based model, i.e., the training data must be retained and used each time a prediction needs to be made. This method performs a regression around a point of interest using only the data that are “local” to that point (Cohn et al. 1996). One limitation of LWLR is that it can be computationally expensive when a huge amount of data is used for training. However, variations of LWLR, such as the one proposed by Huang et al. (2020), can be used to tackle this problem.

Autoregressive models

Autoregressive models aim to describe the autocorrelation present in the data. They are based on the idea that the current value of the series can be explained as a linear combination of past values. This model type is commonly used for time series forecasting, which is why some authors used it. In this study, the autoregressive models used were the ARIMA(X), SARIMA, and Vector Autoregression (VAR).

These models are known for being flexible at handling different time series patterns (Rob J Hyndman and George Athanasopoulos 2018a). ARIMA and SARIMA are univariate models, i.e., they deal with a sequence of measurements of the same variable collected over time, with the latter including a seasonal component.

However, ARIMA and SARIMA are limited when considering the effects of other confounding variables. To

address this limitation, Tuominen et al. (2022) used ARI-MAX, a multivariate extension of ARIMA, for non-seasonal data with exogenous variables. In addition, we found the use of the VAR model for seasonal data. Both ARI-MAX and VAR are multivariate models capable of handling multiple seasonal cycles (e.g., a time series with daily data can have weekly and annual patterns) (Hyndman et al. 2008; Rob J Hyndman and George Athanasopoulos 2018b). Besides, ARIMA(X) and SARIMA are unidirectional, i.e., the forecast variable is influenced by the predictor variables, but not vice versa. In certain cases, variables may affect each other. More advanced autoregressive methods, like VAR, may be recommended in these cases. VAR is an extension of the univariate autoregressive model and is an n -equation, n -variable linear model, where each variable is explained by its lagged values, plus the current and past values of the remaining $n - 1$ variables (Stock and Watson 2001). For instance, in cases where the variables are related to other hospital activities, such as the count of patients waiting for or receiving service in the ED (ED census) and the count of laboratory batteries ordered in the ED (laboratory volumes). A surge in the ED census is followed by a surge in laboratory volumes. This may lead to another increase in the ED census as the longer ED stay is due to longer laboratory turnaround times (Jones et al. 2009). However, when forecasting hospital visits with environmental predictors, we would need to assume, for example, that the number of visits affects the environmental predictors, which is not true. Additionally, some of the limitations of the VAR model are that it fails to capture nonlinearities, conditional heteroskedasticity, and drifts or breaks in parameters (Stock and Watson 2001). Furthermore, using the conventional VAR model may involve estimating too many parameters, particularly when the number of time series and lag order is moderately large (Wang et al. 2021).

Machine learning models

In this systematic review, all studies focused on a supervised learning problem, i.e., model learning relies on labeled training examples to predict volumetric hospital visits and admissions (Murphy 2012).

Some studies used DT, a non-parametric tree-like model used for classification and regression (Breiman et al. 2017). When used for classification, the predictions concerned mainly the levels of expected visits, e.g.(low, medium, or high). In the DT structure, the path from the root to the leaf represents the classification rules, which are composed of a set of if-then-else decision rules. The DT used for the decision can be visualized, making it easy to understand and interpret results. However, they are not good at extrapolation, i.e., they don't perform well when making predictions beyond the training samples and can create

complex trees if using a large number of nodes and depth. Ram et al. (2015) did not specify the reason for using DT. Nonetheless, it is known that DT is popular due to its relative ease of applicability and interpretability (Song and Ying 2015). Furthermore, we found the use of the RF model, an ensemble learning method that uses multiple learning models to obtain better results. RF uses a combination of tree predictors. Each tree is created using a bootstrapped dataset and a random subset of predictor variables (Breiman 2001). Each tree casts a unit vote for the classification task, and the most popular class is the final output.

For the regression task, the predicted output is the number of visits, and the final result is the average among all the trees. The authors decided to use RF mainly because of its capability to deal well with non-linearities and its internal variable scoring method. RF provides the importance score of each variable, which can be helpful for feature selection and model explainability. Nonetheless, like DT, attention must be paid when using a large number of trees and depth since it can lead to computational complexity.

Moreover, other variations of DT were used, such as distributed random forest (DRF), which consists of a parallelized implementation of RF (LeDell et al. 2020), and XRT (Geurts et al. 2006), which consists of totally randomized trees resulting from the randomization of both attribute and cut-point choice while splitting a tree node. The first provides an implementation that can handle more data and computing resources. The second has been shown to be competitive with RF in terms of regression problems when considering accuracy. However, XRT presented more computational advantages for very large numbers of attributes than RF. Another DRF variant, which uses a different splitting criterion to detect heterogeneity in multivariate distributions, was also used (Cevic et al. 2022). Additionally, generalized random forests (GRF) was used. The GRF is a model used for nonparametric statistical estimation based on RFs and can be used to fit any quantity of interest (Athey et al. 2019).

Additionally, some studies used the SVM model. SVM is primarily used for binary classification, but it can also handle multi-class problems by transforming them into multiple binary classification problems. Furthermore, it can also be used for regression problems through SVR to predict quantitative outputs. This model implements the idea that the input vectors are non-linearly mapped to a very high-dimensional feature space. In addition, a linear decision surface is constructed in this feature space, and its special properties ensure high generalization. To construct optimal hyperplanes, i.e., the linear decision function with the maximal margin between the vectors of two classes, SVM considers a small amount of the training data to determine this margin (Cortes and Vapnik

1995). This approach can be computationally expensive, especially if using large datasets. The authors chose SVM mainly because of its popularity and ability to handle small datasets well.

We also found the use of stacking ensemble models. In contrast to RF, which uses bootstrap aggregating (bagging) by generating random samples from the training observations and fitting the same base model to each sample, stacking models use different base models combined into the ensemble model. Since the base models differ, they can gain insights into the training data from various perspectives, eventually increasing the accuracy and improving the model's generalization ability (Chen et al. 2022). This is why some authors have adopted this approach.

Another model used was the naive bayes (NB), which is commonly used for binary classification but can also be adjusted for multi-class problems. NB assumes that the attributes used to make predictions are independent of each other, given the predicted value. This model performs well on most classification tasks. However, it does not perform as well in the context of regression (Frank et al. 2000). Ram et al. (2015) used this approach based on previous use in similar works and found poor results. In addition, we found the use of the KNN model, a simple but effective classification method where its results are given by the majority of votes among the data points in the neighborhood. Nevertheless, KNN does not use the training data to make generalizations, meaning it does not learn rules. Instead, it retains the training data during the testing phase. This can cause problems when using a large dataset. Additionally, its efficiency relies on selecting a good k value (Song et al. 2017; Guo et al. 2003). The authors who used KNN did not specify the reason for it. We assume it was because of its simplicity.

Gradient boosting models

Gradient boosting decision trees (GBDT), also known as gradient boosting machines (GBM), is an ensemble learning model. It consists of an ML technique that combines the predictions of multiple models (Mienye and Sun 2022). In this approach, each base model is added to correct the error produced by its previous model (Friedman 2001). This approach is known for its efficiency, accuracy, and interpretability. In addition, other implementations of this approach were used, including XGBoost and LightGBM. The first algorithm is well-known for its scalability and computational efficiency due to numerous algorithmic optimizations (Chen and Guestrin 2016). The

second algorithm is known for significantly outperforming XGBoost regarding computational speed and memory consumption (Ke et al. 2017). The authors used these approaches due to their high performance and ability to handle nonlinear relationships within the data. Besides, some authors used XGBoost to improve model heterogeneity by combining it with different models, creating a more robust model capable of capturing various patterns in the data.

Deep learning models

DL methods require little human intervention and can process huge amounts of raw data to discover the representations needed for detection or classification automatically. They are well-known for their exceptional capability to uncover intricate structures within high-dimensional data. This differs from ML, which faces limitations in processing raw data, usually requiring manual engineering of a feature extractor to transform raw data into a feature vector to be used as input for a learning system (LeCun et al. 2015).

One of the most common families of models used when modeling sequential data is the RNN. RNNs process an input sequence of data one slice at a time while maintaining information about the history of past elements of the sequence (LeCun et al. 2015). However, RNNs face limitations when modeling long-term dependencies, usually leading to gradient vanishing and exploding problems. This is why many authors opted for using other variants of RNN that address these limitations through gated network signals that control how the present input and previous memory update the current activation and produce the current state through their gating network (Dey and Salem 2017).

One study used a gated recurrent unit (GRU), which has two gates, reset and update. These gates control the flow of information within the network. The update gate determines the amount of past information to carry into the future, while the reset gate determines the amount of past information to forget (Cho et al. 2014). In addition, LSTM was used. Unlike GRU, LSTM has a memory cell and three gates: input, forget, and output. The gates control the information flow in the memory cell: the input gate adds information, the forget gate removes information, and the output gate manages the output. This enables handling long-term dependencies, which is impossible when using standard RNNs (Hochreiter and Schmidhuber 1997, 1996).

Furthermore, we found the use of bidirectional LSTM (BiLSTM), which enables training by traversing the input from left to right and right to left. This approach

can present a better performance than LSTMs when using text data. However, it remains unclear whether this approach would be beneficial when using numerical time series (Siami-Namini et al. 2019). In the study presented by Mossop et al. (2023), the results of LSTM and Bi-LSTM were quite similar. However, when combined with CNN, Bi-LSTM performed better.

Hybrid models

In this review, we found that the use of hybrid models was primarily motivated by the potential to improve the overall modeling and forecasting performance. This is due to the model's heterogeneity, especially when modeling linear and nonlinear patterns separately (Zhang 2003).

In the study conducted by Zhang et al. (2018), the authors built hybrid models by fitting the ETS residuals as input to other models. An approach similar to the gradient boosting method. Besides, we found the use of a hybrid DL model presented by Sciannameo et al. (2022). The authors used a CNN-LSTM, a choice adapted to the data type, which consisted of grid satellite images captured over time. For this type of data, using CNNs to model spatial relations can be advantageous, especially because each neuron processes the inputs belonging to a specific bounded area, enabling forecasting at the grid level. In addition, using LSTM with CNN made processing a sequence of images one slice at a time possible while handling long-term dependencies presented in the data. In addition, Mossop et al. (2023) presented other variations of CNN-LSTM. The authors did not use remote sensing data in their study yet still achieved good results. They used causal convolutions, a variation of CNN that preserves causality by preventing information "leakage" from future to past, considering only the past historical input. This approach may outperform other recurrent architectures and exhibit longer memory, making it more suitable for capturing long-term dependencies (Bai et al. 2018).

Risk of bias

Figure 11 summarizes the PROBAST assessment conducted, and Table 3 presents more details. Among the studies, twenty-two were identified as possessing a high RoB, while thirteen studies were determined to have low RoB, and one study had unclear RoB. Four studies displayed a high RoB in the participant domain, indicating concerns regarding the data sources utilized. One study

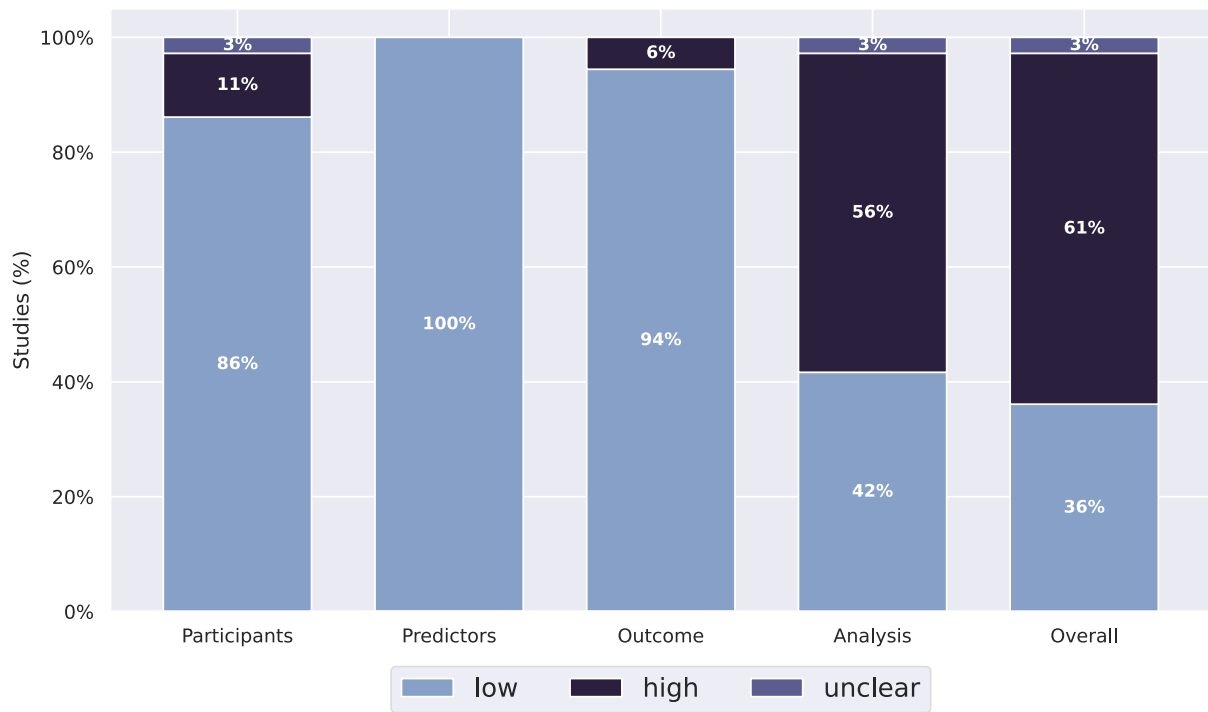
had unclear RoB in the participant domain, raising concerns regarding the inclusion or exclusion of participants. Additionally, two studies exhibited a high RoB in the outcome domain, raising concerns about the definition and determination of the outcome. Furthermore, twenty studies demonstrated a high RoB in the analysis domain, suggesting inadequate analysis methods or the omission of crucial statistical considerations. One study had an unclear RoB in the analysis domain, indicating insufficient information to assess whether missing data were handled appropriately. These factors can introduce bias and undermine the reliability of the estimated predictive performance of the model. In addition, two studies were identified as having a high concern regarding their applicability. Specifically, these studies exhibited high concerns in the predictor domain, indicating inconsistencies between the predictors' definition, assessment, or timing and the review question. Three studies had unclear concerns regarding applicability, both with unclear concerns in the participant domain. Notably, thirteen studies were classified as having a low RoB and low applicability concerns, suggesting adequate data collection methods and study designs.

Discussion

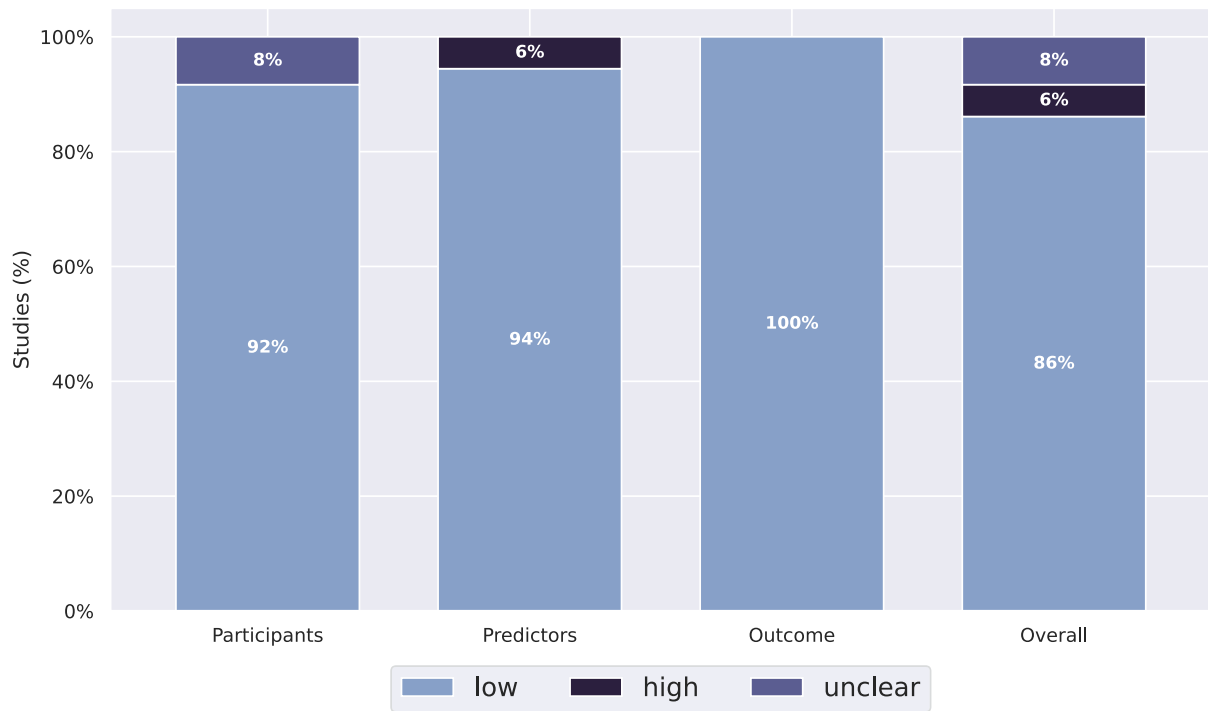
In this review, we systematically searched studies using machine learning methods to predict hospital visits and admissions from meteorological or air pollution data.

Participants and outcome

We found one study (Jing et al. 2020) where the definition of the data sources was unclear. The authors acknowledged the utilization of non-original data but failed to provide a clear definition of its sources. This omission poses a potential risk of bias in the participant domain. Also, a high risk of bias was assigned in the participant domain to Mossop et al. (2023); Yang et al. (2025) and Cappelli et al. (2024). These studies used datasets that combined pre-pandemic and pandemic periods. The inclusion of the COVID-19 pandemic period, likely negatively impacted model performance as the pandemic led to a global surge in hospital demand and influenced air pollution emission levels (Muhammad et al. 2020). In addition, one study (Tuominen et al. 2022) reported a reorganization of the ED during the study period, causing patients to be redirected to another ED, which may result in biased estimates of the models. Furthermore, most studies use hospital administrative databases. Those data sources are not designed for research purposes, and



(a) RoB



(b) Applicability concerns

Fig. 11 RoB and applicability concerns

Table 3 PROBABST assessment of the included studies

Study	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Sciannameo et al. (2022)	+	+	+	-	+	-	+	-	-
Hu et al. (2020)	+	+	+	+	+	+	+	+	+
Usmani et al. (2021)	+	+	+	-	+	+	+	-	+
Jing et al. (2020)	-	+	+	+	+	-	+	-	-
Stonko et al. (2018)	+	+	+	-	+	+	+	-	+
Zhang et al. (2018)	+	+	+	-	+	+	+	-	+
Khatri and Tamil (2017)	+	+	+	-	+	+	+	-	+
Hovland et al. (2013)	+	+	+	+	+	+	+	+	+
Duwalage et al. (2020)	+	+	+	+	+	+	+	+	+
Tuominen et al. (2022)	?	+	+	-	+	+	+	-	+
Ogata et al. (2021)	+	+	+	+	+	+	+	+	+
Qiu et al. (2020)	+	+	+	+	+	+	+	+	+
Chen et al. (2019)	+	+	+	+	+	+	+	+	+
Meng et al. (2015)	+	+	-	-	+	+	+	-	+
Tavera Busso et al. (2021)	+	+	+	-	+	+	+	-	+
Ram et al. (2015)	+	+	-	-	+	+	+	-	+
Huang et al. (2020)	+	+	+	+	+	+	+	+	+
Medina Quero et al. (2018)	+	+	+	-	+	+	+	-	+
Moustris et al. (2012)	+	+	+	-	+	+	+	-	+
de Souza Tadano et al. (2016)	+	+	+	-	+	+	+	-	+
Lin et al. (2021)	+	+	+	-	+	+	+	-	+
Lu and Qiu (2023)	+	+	+	+	+	+	+	+	+
Mossop et al. (2023)	-	+	+	-	?	+	+	-	?
Siqueira et al. (2023)	+	+	+	-	+	+	+	-	+
Gafni-Pappas and Khan (2023)	+	+	+	+	+	+	+	+	+
Song et al. (2022)	+	+	+	-	+	+	+	-	+
Cengiz and Terzi (2012)	+	+	+	-	+	+	+	-	+
Chou-Chen and Barboza (2024)	+	+	+	+	+	+	+	+	+
Guo et al. (2024)	+	+	+	-	+	+	+	-	+
Zhong et al. (2024)	+	+	+	-	+	+	+	-	+
Yang et al. (2025)	-	+	+	-	?	+	+	-	?
de Souza Tadano et al. (2024)	+	+	+	+	+	+	+	+	+
Mutlu et al. (2024)	+	+	+	?	+	+	+	?	+
Kurucz et al. (2024)	+	+	+	+	+	+	+	+	+
Susmann et al. (2025)	+	+	+	+	+	+	+	+	+
Cappelli et al. (2024)	-	+	+	+	?	+	+	-	?

+ Indicates low ROB/low concern regarding applicability;

- Indicates high ROB/high concern regarding applicability;

? Indicates unclear ROB/concern regarding applicability

the reasons that motivated their creation and the way the information was collected may strongly influence the accuracy and reliability of the data set (van Walraven and Austin 2012). Studies that target some particular cause of visit or admission (e.g., cardiovascular disease) often rely on ICD-10 classification, and so they rely on the thoroughness of the coding. As suggested by several studies, administrative databases may contain coding errors in the diagnostic (Khatri and Tamil 2017; Medina Quero et al. 2018; Moustris et al. 2012), and may be inconsistently collected over time or across different physicians and institutions.

The question of readmissions was rarely raised. Tavera Busso et al. (2021) mentioned being aware of multiple entries per patient but chose to ignore this point. For a large dataset, this is unlikely to affect the predictive performance. However, in some cases, ignoring those dependencies may lead to prediction and variance biases.

The primary source of concern related to outcome definition stemmed from the use of non-standard definitions, often relying on the authors' subjective judgment and expertise, and the lack of information provided regarding the authors' approach to converting hospital visits and admissions into distinct classes (e.g., high, medium, low).

Predictors and sample size

We did not identify high bias regarding predictors. Most studies used PM_{2.5}, PM₁₀, CO, O₃, NO₂, SO₂, Temperature, and RH. The World Health Organization (2021) recognizes these air pollutants as harmful to health and establishes recommended thresholds to guide global air quality standards. A classical way of defining the pollution exposure of the patients is to use their residential address (Usmani et al. 2021; Ram et al. 2015). Pollution data may be an average of measures (Khatri and Tamil 2017) coming from stations whose location (e.g. close to a road) may not be accurately representative of the real exposure of patients. Moreover, those measures may not be continuous and result in missing data. Pollution models may be more accurate to describe the exposure but may suffer from their own bias (Stonko et al. 2018). Some authors have raised concerns about the excessive use of calendar-based predictor variables (e.g., school holidays, events, sporting events, music festivals, etc.), which could lead to overfitting (Stonko et al. 2018; Tuominen et al. 2022). In addition, they raised concerns about the possibility of leakage of weekly seasonality into public event predictors. This issue arises due to the prevalence of public events occurring on weekends, making it challenging

to distinguish the effects of the weekly seasonality from the impact of the public events. Gafni-Pappas and Khan (2023) suggested that much of the patient volume could be predicted from calendar-based predictor variables alone. Using this type of variable can be beneficial for capturing long-term trends. In addition, Hu et al. (2020) found a slightly higher feature importance for time-related variables when compared to environmental predictors. Furthermore, the incorporation of additional environmental exposures, such as arsenic and cadmium metals, along with factors related to influenza, seasonal flu, and circulating aero-allergens, could potentially contribute to improving forecasting capabilities (Qiu et al. 2020; Medina Quero et al. 2018; Moustris et al. 2012).

In some studies, the data exhibited noise in the form of daily random fluctuations. To mitigate this, some studies suggested employing smoothing techniques to reduce the noise. Although this approach can help address model overfitting, it also carries the potential risk of losing valuable information if these fluctuations contain significant insights.

Concerning the sample size, we found that EPV is poorly reported across all studies. According to van der Ploeg et al. (2014), predictive models using ML often require an EPV greater than 200. However, based on the e-EPV, only a small portion of the studies reviewed adhered to this recommendation. The use of a low EPV increases the risk of model overfitting.

Model analysis

This study found the most important sources of bias from model analysis. The main source of bias was related to the way of handling missing data. Several studies did not mention the proportion of missing data or the way of handling it. In some cases, the authors chose to address this issue by removing missing values or replacing them with zeros. Multiple imputation is recommended but has been rarely used. The observation regarding the poor management and reporting of missing data in the prediction models has already been noted (Nijman et al. 2022). Furthermore, another source of bias was the use of unconventional metrics to assess the model's predictive performance Stonko et al. (2018); Meng et al. (2015).

The correlation between predictor variables was rarely discussed in the reviewed articles despite being inherent in pollution measures (e.g., PM_{2.5} and PM₁₀). Studies that did not mention the problem of multicollinearity but employed a predictor selection technique may have indirectly mitigated this problem.

Most studies used an out-of-sample with fixed-origin procedure approach to evaluate the predictive ability of their models. This means that the data set is split into a train and a test set and that the predictions on the test set have a single forecasting origin. The fixed-origin procedure has several shortcomings, and rolling-origin procedures were advocated instead (Tashman 2000). Several studies resorted to cross-validation to choose the hyperparameters. However, with data that are not independent and identically distributed like in time series, cross-validation may not be appropriate (Cerqueira et al. 2020). Some variants have been designed for dependent data but were specifically stated in the articles of this review. In general, the reporting and conducting of hyperparameter selection in the studies lacked details.

Applicability

The main limitation for the applicability of the models based on weather and pollution predictors is that these data need to be predicted along with the model's predictions. Weather and pollution forecasting may be inaccurate and reduce the predictive performance of the models (Stonko et al. 2018; Duwalage et al. 2020). Additionally, the study period in some cases included the pre-pandemic and pandemic periods (Mossop et al. 2023; Yang et al. 2025; Cappelli et al. 2024), and the use of a proprietary algorithm to define predictors (Sciannameo et al. 2022), combined with insufficient information about the predictors and the use of unspecified data (Jing et al. 2020) undermines the study's generalizability and applicability. In addition, population characteristics and differences in study locations may affect the generalizability and applicability of the studies. Overall, the reviewed studies do not discuss the practical application of their models for hospital visits and admissions management.

Limitations—study level

This study overviewed the various machine learning models recently used in the literature. However, there was a large heterogeneity between the selected studies regarding population, forecast horizons, granularity (e.g., hourly, daily, weekly), evaluation metrics adopted, and the way to enter the variables (e.g., pollution or weather variables) in the models. Hence, it was impossible to compare their performances or make definitive recommendations regarding the optimal forecast horizon and models to be used.

Nonetheless, we identified the most commonly selected models across studies and their corresponding

forecast horizons. While these findings do not indicate the optimal models for a given forecast horizon, they provide insight into the models most frequently used for each forecast horizon. Models reported in the literature review show good performances. Those results may be prone to a publication bias (Boulesteix et al. 2015).

Limitations—review level

When using the PROBAST tool, it is customary to assess personalized predictions and apply the PROBAST evaluation to each model used in a study. In the present work, we used the PROBAST tool to evaluate ecological retrospective longitudinal studies at the study level rather than applying it to each model within a study. This decision was based on the fact that the data sources, participants, definition and assessment of predictors, and outcome definition remained consistent across all models investigated in each study. Consequently, employing a standardized tool for assessing the risk of bias in ecological retrospective longitudinal studies may prove more appropriate and yield better outcomes. Although considerable efforts were made to use comprehensive search terms, some studies may have been published under terms not included in this review. In addition, while EMBASE, PubMed, and IEEE Xplore cover a wide range of relevant studies, relying on these sources may have excluded studies indexed in other databases. Adding more databases could have improved the comprehensiveness of this review, though the current selection still ensures a solid representation of the key literature in the field.

When drawing conclusions based on the e-EPV values, it is crucial to exercise caution, as the authors of the included studies did not explicitly provide the EPV ratio. Instead, we estimated the ratio using the available information within each study. Furthermore, the lack of information regarding missing data may have directly influenced the number of outcome events used to calculate the e-EPV ratio.

In this review, study selection was done by two authors: author LMM conducted 100% of the study selection, and author EC contributed to 20% of the study selection. Furthermore, the PROBAST application was performed by one author (LMM) and overseen by EC. DMM and MS have provided significant intellectual contributions in revising the manuscript, thoroughly reviewing it, and granting approval for the final version.

Recommendations

The heterogeneity across studies poses challenges when comparing and recommending a specific model architecture for patient volume forecasting. However, we observed that most studies used at least both relatively simple and more complex models. The simple model serves as a baseline model. For some studies, this model may be univariate, autoregressive, and able to capture seasonality such as SARIMA. This baseline model can be a good starting point for capturing the seasonality present in hospital visits and admissions. More complex models, on the other hand, are usually multivariate. We could consider using ML models such as RF if we have a considerably large dataset, and using DL models such as LSTM or causal CNN if we are using high-dimensional data. Furthermore, hybrid deep learning approaches such as CNN-LSTM may be worth considering in cases where remote sensing and high-dimensional data are used. Although the number of studies investigating ensemble learning found in this review is limited, these findings suggest a promising avenue for enhancing model performance in future applications, and they emphasize the potential benefits of ensemble learning when integrating multiple types of data.

Many studies have incorporated calendar predictors such as year, month, day of the week, and holidays in conjunction with environmental predictors. We suggest that these predictors be taken into account when forecasting hospital visits and admissions. However, it is important for authors not to limit themselves to only these predictors. Very few studies have used land use, remote sensing, demographics, and socioeconomic predictors. This presents an opportunity for further exploration in future studies.

Furthermore, using longer time periods and larger cohorts is recommended to minimize the likelihood of systematic errors introduced by physicians during the diagnostic process (Moustris et al. 2012).

Regarding study reproducibility, most studies did not share their source codes and datasets. Sharing source codes and datasets would enhance transparency in

generating computational results and facilitate future benchmarking studies.

Future work

In this literature review, we observed some lack in the way of reporting the methods (e.g., missing data). Moreover, some authors did not sufficiently discuss their results. More care should be addressed in future research to improve the validity of the studies. We also observed that studies often did not explore the use of social and demographic variables, which can play a role in determining the individuals' susceptibility to health issues caused by air pollution and their utilization of healthcare services (Jiao et al. 2018). Additionally, validation studies regarding the applicability of the models for daily use in hospital management would be of interest for future research. Finally, the geographic distribution of the studies in this review reveals an opportunity to study air pollution and weather variables in forecasting hospital visits and admissions in previously unstudied locations.

Conclusion

The study comprehensively synthesized evidence from previous studies to elucidate how ML has been applied and which environmental predictors have been used to forecast hospital visits and admissions. Results were obtained following standard guidelines to select and extract data from 36 papers. Overall, studies included in this work reported good predictive model performances. However, a significant portion of the studies exhibited a high risk of bias. This is primarily due to the lack of standard outcome definitions, inadequate handling of missing values, small sample sizes, and a low number of events per variable. In addition, it is crucial to adopt standardized reporting guidelines to enhance the quality and reliability of future studies in this field. Further research is necessary to evaluate and determine the optimal model architecture for patient volume forecasting when incorporating environmental predictors.

Appendix A Search Queries

Table 4 Search queries deployed in the PubMed, Embase and IEEE Xplore databases

	PubMed	Embase	IEEE Xplore
1	"Hospitalization"[Mesh] OR "Emergency Service, Hospital"[Mesh] OR "hospital admission"*[all] OR "emergency department visits"*[all] OR "emergency department admission"*[all]	("hospital admission"/exp OR "hospital admission" OR "emergency department admission" OR "emergency department visit"/exp OR "emergency department visit")	(("Mesh Terms":"Hospitalization") OR ("Full Text & Metadata":"hospital admission") OR ("Mesh Terms":"Emergency Ser- vice, Hospital") OR ("Full Text & Metadata":"emergency department admission*") OR ("Full Text & Metadata":"emergency department visit*")) AND (("Mesh Terms":"Machine Learn- ing") OR ("Full Text & Metadata":"machine learning") OR ("Mesh Terms":"Artificial Intel- ligence") OR ("Full Text & Metadata":"artificial intelligence") OR ("Mesh Terms":"Deep Learning") OR ("Full Text & Metadata":"deep learning") OR ("Full Text & Metadata":"Istm") OR ("Full Text & Metadata":"neural network")) AND (((("Mesh Terms":"Air Pollution") OR ("Mesh Terms":"Air Pollutants") OR ("Mesh Terms":"Particulate Mat- ter") OR ("Full Text & Metadata":"air qual- ity")) OR (("Mesh Terms":"Weather") OR ("Mesh Terms":"Atmosphere") OR ("Full Text & Metadata":"weather variables") OR ("Full Text & Metadata":"temperature") OR ("Full Text & Metadata":"wind")))) AND (("Mesh Terms":"Forecasting") OR ("Full Text & Metadata":"prediction") OR ("Full Text & Metadata":"forecast*")) Manual settings: Filters: Journals and Conferences; Period: 2012 to 2025.

Table 4 (continued)

	PubMed	Embase	IEEE Xplore
2	"Machine Learning"[Mesh] OR "Artificial Intelligence"[Mesh] OR "Deep Learning"[Mesh] OR "Machine Learning"[all] OR "Artificial Intelligence"[all] OR "Deep Learning"[all] OR "ml"[all] OR "ai"[all] OR "naive bayes"[all] OR "linear regression"[all] OR "logistic regression"[all] OR "genetic algorithm"[all] OR "svm"[all] OR "support vector machine"[all] OR "random forest"[all] OR "cluster analysis"[all] OR "decision tree"[all] OR "XGBoost"[all] OR "gradient boost"[all] OR "adaboost"[all] OR "mlp"[all] OR "multilayer perceptron"[all] OR "Istm"[all] OR "long short-term memory"[all] OR "long short term memory"[all] OR "long short-term memory network"[all] OR "rnn"[all] OR "recurrent neural network"[all] OR "cnn" OR "convolutional neural network"[all] OR "expert system"[all] OR "supervised machine learning"[all] OR "unsupervised machine learning"[all] OR "semi-supervised learning"[all] OR "reinforcement learning"[all] OR "computational intelligence"[all] OR "computer reasoning"[all] OR "nlp"[all] OR "natural language processing"[all] OR "ann"[all] OR "artificial neural network"[all] OR "nn"[all] OR "neural network"[all] OR "feedforward neural network"[all] OR "computer neural network"[all]	("machine learning"/exp OR "machine learning" OR "artificial intelligence"/exp OR "artificial intelligence" OR "deep learning"/exp OR "deep learning" OR "ml" OR "ai" OR "naive bayes" OR "linear regression" OR "logistic regression" OR "genetic algorithm" OR "svm" OR "support vector machine" OR "random forest" OR "cluster analysis" OR "decision tree" OR "XGBoost" OR "gradient boost" OR "adaboost" OR "mlp" OR "multilayer perceptron" OR "Istm" OR "long short term memory" OR "rnn" OR "recurrent neural network" OR "cnn" OR "convolutional neural network" OR "expert system" OR "supervised machine learning" OR "unsupervised machine learning" OR "semi-supervised learning" OR "reinforcement learning" OR "computational intelligence" OR "computer reasoning" OR "nlp" OR "natural language processing" OR "ann" OR "artificial neural network" OR "nn" OR "neural network" OR "feedforward neural network" OR "computer neural network")	IEEE Xplore
3	"Air Pollution"[Mesh] OR "Air Pollutants"[Mesh] OR "Particulate Matter"[Mesh] OR "air quality"[all]	("air pollution"/exp OR "air pollutants"/exp OR "particulate matter"/exp OR "air quality")	IEEE Xplore
4	"Weather"[Mesh] OR "Atmosphere"[Mesh] OR "weather variables"[all] OR "temperature"[all] OR "wind"[all] OR "meteorological variables"[all]	("weather"/exp OR "atmosphere"/exp OR "weather variables" OR "meteorological variables" OR "wind")	IEEE Xplore
5	"Forecasting"[Mesh] OR "prediction"[all] OR "forecast"[all]	("prediction"/exp OR "prediction" OR "forecast"/exp OR "forecast")	IEEE Xplore
6	(2012:2025[mdat])	[2012-2025]/py	IEEE Xplore
7	#1 AND #2 AND (#3 OR #4) AND #5 AND #6	#1 AND #2 AND (#3 OR #4) AND #5 AND #6	IEEE Xplore

Table 5 (continued)

Study	Sciannaneo et al. (2022)	Hu et al. (2020)	Usmani et al. (2021)	Jing et al. (2020)	Stonko et al. (2018)	Zhang et al. (2018)	Khatri and Tamil (2017)	Hovland et al. (2013) ⁽¹⁾	Duwalage et al. (2020)
Avg. of the number of days in which the amplitude of temperature is >90 th percentile of daily amplitude temperature									
Max. of the max. daily precipitation									
Avg. of the max. daily precipitation									
Avg. of the avg. daily precipitation									
Avg. of the number of days in which the max. precipitation is >75th of daily max. precipitation									
Avg. of the number of days in which the max. precipitation is >90th of daily max. precipitation									
Max. level 1 soil temperature									
Mean level 1 soil temperature									
Min. level 1 soil temperature									
Max. vapor pressure									
Mean vapor pressure									
Min. vapor pressure									
Evaporation									
Daily fire count									
Temperature difference									
Temperature (Difference, 12h to previous 24h)									
Heat Index (Apparent Temperature)									
Thermal Amplitude									
Air/Atmospheric/Surface pressure									
Air/Atmospheric/Surface pressure mean									
Air/Atmospheric/Surface pressure min.									
Air/Atmospheric/Surface pressure max.									
Air/Atmospheric/Surface pressure difference									
Dew Point									
Max. dew point									
Mean dew point				▲					
Min dew point									
Relative Humidity	▲	▲					▲		

Table 5 (continued)

Study	Sciannaeo et al. (2022)	Hu et al. (2020)	Usmani et al. (2021)	Jing et al. (2020)	Stonko et al. (2018)	Zhang et al. (2018)	Khatri and Tamil (2017)	Hovland et al. (2013) ⁽¹⁾	Duwalage et al. (2020)
Mean atmospheric humidity									
Max atmospheric humidity									
Min atmospheric humidity									
Mean Relative Humidity									
Min Relative Humidity									
Mean absolute humidity									
Atmospheric humidity difference									
Max absolute humidity									
Rain Intensity									
Precipitation					▲				
Precipitation duration									
Max. Precipitation									
Mean wind speed									
Max. wind speed									
Min. wind speed									
Max. wind temperature									
Max. wind gust									
Vector average wind direction									
Vector average wind speed									
Wind Speed	▲			▲			▲		
Wind Direction				▲					
Visibility				▲					
Wet Bulb (Global) Temperature									
Solar Radiation	▲								
Downward Solar Radiation									
Sunshine duration									
Daily potential sunshine duration (%)									
Snow Depth									
Cloud Count									
Weather Type									
Cloud Cover				▲					
Significant Tornado Parameter				▲					
SFC Pressure (?)									
Daily Change in SFC Pressure (?)									
Sea-level Pressure				▲					
Artificially Observed Weather				▲					

Table 5 (continued)

Study	Sciannaneo et al. (2022)	Hu et al. (2020)	Usmani et al. (2021)	Jing et al. (2020)	Stonko et al. (2018)	Zhang et al. (2018)	Khatri and Tamil (2017)	Hovland et al. (2013) ⁽¹⁾	Duwalage et al. (2020)
Website visits									
Tweets (w/ word "Asthma")									
Google search trends									
Predictor/Type	Calendar								
Date Index									
Time of the day									
Year		▲							▲
Month		▲							▲
Day Of the Week		▲			▲				▲
Day of Month									
Day of the Year					▲				
Working Day									
Holidays		▲				▲			▲
Holiday length		▲							
Vacations						▲			
Public Events (minor, major, all, and specific)									
Work day (last and first)		▲							
Season								△	
Period (Warm, Cold)									
Rainy Season									
Difference between the corresponding day and the last day of the rainy season of each year									
Day of the Quarter									
Modulus for Quad-weekly frequency									
Week Number of the Year (Monday and Sunday start)									
Indicator for any Friday immediately following a national holiday									
Indicator for days following a national holiday									
Holiday Indicator: July 14									
Holiday Indicator: Christmas									
Holiday Indicator: New Years Eve									
Holiday Indicator: New Years Day									
Predictor/Type	Time series features								
Difference value (previous 1-27 days)		△							
Moving Average									
Standard deviation		△							
Kurtosis		△							
Skewness		△							

Table 5 (continued)

Study	Tuominen et al. (2022) ⁽¹⁾	Ogata et al. (2021)	Qiu et al. (2020)	Chen et al. (2019)	Meng et al. (2015)	Tavera Busso Ram et al. (2015)	Huang et al. (2020)	Medina Quero et al. (2018)
Avg. of the number of days in which the min. temperature is < 25th of min. temperature								
Min. of Min. Temperature								
Max. of Min. Temperature								
Max. of Max. daily amplitude of temperature (difference of daily max. and min)								
Avg. of Max. daily amplitude of temperature								
Min. of max. daily amplitude of temperature								
Max. of the min. daily amplitude of temperature								
Avg. of min. daily amplitude of temperature								
Min. of Min. daily amplitude of temperature								
Avg. of the number of days in which the amplitude of temperature is >75th percentile of daily amplitude of temperature								
Avg. of the number of days in which the amplitude of temperature is >90 th percentile of daily amplitude temperature								
Max. of the max. daily precipitation								
Avg. of the max. daily precipitation								
Avg. of the avg. daily precipitation								
Avg. of the number of days in which the max. precipitation is >75th of daily max. precipitation								
Avg. of the number of days in which the max. precipitation is >90th of daily max. precipitation								
Max. level 1 soil temperature								
Mean level 1 soil temperature								
Min. level 1 soil temperature								
Max. vapor pressure								
Mean vapor pressure								
Min. vapor pressure								
Evaporation								
Daily fire count								
Temperature difference								
Temperature (Difference, 12h to previous 24h)		▲*						
Heat Index (Apparent Temperature)		△						
Thermal Amplitude						△		

Table 5 (continued)

Study	Tuominen et al. (2022) ⁽¹⁾	Ogata et al. (2021)	Qiu et al. (2020)	Chen et al. (2019)	Meng et al. (2015)	Tavera Busso Ram et al. (2015)	Huang et al. (2020)	Medina Quero et al. (2018)
Time of the day		▲						
Year			▲					
Month	△	▲	▲			△		
Day Of the Week	△		▲			△		△
Day of Month			▲					
Day of the Year								
Working Day	△							
Holidays	△	▲	▲					
Holiday length								
Vacations								
Public Events (minor, major, all, and specific)	△							
Work day (last and first)								
Season								
Period (Warm, Cold)						△		
Rainy Season		▲						
Difference between the corresponding day and the last day of the rainy season of each year		▲						
Day of the Quarter								
Modulus for Quad-weekly frequency								
Week Number of the Year (Monday and Sunday start)								
Indicator for any Friday immediately following a national holiday								
Indicator for days following a national holiday								
Holiday Indicator: July 14								
Holiday Indicator: Christmas								
Holiday Indicator: New Years Eve								
Holiday Indicator: New Years Day								
Predictor/Type	Time series features							
Difference value (previous 1-27 days)								
Moving Average								
Standard deviation								
Kurtosis								
Skewness								
Predictor/Type	Demographic/ Socioeconomic							
Land Value (median and mean)						△		
Population Density (Mean)						△		
Households with at Least One Unsatisfied Basic Need (Mean)						△		
Household Head with Unfinished High School Studies (Mean)						△		
Unemployment rate								
Violent crime rate								
Median Age of a city		▲						

Table 5 (continued)

Study	Moustris et al. (2012)	de Souza Tadano et al. (2016)	Lin et al. (2021)	Lu and Qiu Mossop et al. (2023)	Siqueira et al. (2023)	Gafni-Pappas and Khan (2023)	Song et al. (2022)	Cengiz and Terzi (2012)
Min. of Min. daily amplitude of temperature								
Avg. of the number of days in which the amplitude of temperature is >75th percentile of daily amplitude of temperature								
Avg. of the number of days in which the amplitude of temperature is >90 th percentile of daily amplitude temperature								
Max. of the max. daily precipitation								
Avg. of the max. daily precipitation								
Avg. of the avg. daily precipitation								
Avg. of the number of days in which the max. precipitation is >75th of daily max. precipitation								
Avg. of the number of days in which the max. precipitation is >90th of daily max. precipitation								
Max. level 1 soil temperature								
Mean level 1 soil temperature								
Min. level 1 soil temperature								
Max. vapor pressure								
Mean vapor pressure								
Min. vapor pressure								
Evaporation								
Daily fire count								
Temperature difference								
Temperature (Difference, 12h to previous 24h)								
Heat Index (Apparent Temperature)			▲					
Thermal Amplitude								
Air/Atmospheric/Surface pressure								
Air/Atmospheric/Surface pressure mean								
Air/Atmospheric/Surface pressure min.								
Air/Atmospheric/Surface pressure max.								
Air/Atmospheric/Surface pressure difference								
Dew Point								
Max. dew point								
Mean dew point								
Min dew point								
Relative Humidity	▲			▲	▲	△	△	

Table 5 (continued)

Study	Moustris et al. (2012)	de Souza Tadano et al. (2016)	Lin et al. (2021)	Lu and Qiu Mossop et al. (2023)	Siqueira et al. (2023)	Gafni-Pappas and Khan (2023)	Song et al. (2022)	Cengiz and Terzi (2012)
Musculoskeletal disease								
Congenital Abnormalities								
Oral Diseases								
Unintentional injuries								
Intentional injuries								
Ill-defined diseases								
Ill-defined injuries/accidents								
Factors influencing health status and contact with health services								
Predictor/Type		Land Use						
Land Use								
Urban Fragmentation								
Built Surface								
Roads Lineal Distance (avenues, street other, total)								
Energy (distance to the nearest thermal plants, and gas stations)								
Open-Air Dumps (Distance to the Nearest)								
Productive Units - Amount								
Energy - Number of Gas Stations								
Open-Air Dumps - Amount								
Green Area Proportion (% area of parks, forests, and arable land divided by city area)								
Predictor/Type			Remote Sensing					
Crowding Index								
Elevation - Mean								
Main Land Cover								
Land Surface Temperature								
Normalized Difference Vegetation Index								
Aerosol Optical Density								
Predictor/Type						Website (visits/searches/text posts)		
Ekstrom (visits, and ratios)								
Website visits								
Tweets (w/ word "Asthma")								
Google search trends								
Predictor/Type							Calendar	
Date Index								△
Time of the day								
Year				▲				
Month	▲			▲				△
Day Of the Week		▲		▲	▲			▲
Day of Month								
Day of the Year								△
Working Day				▲				

Table 5 (continued)

Study	Chou-Chen and Barboza (2024)	Guo et al. (2024)	Zhong et al. (2024)	Yang et al. (2025)	de Souza Tadano et al. (2024)	Mutlu et al. (2024)	Kurucz et al. (2024)	Susmann et al. (2025)	Cappelli et al. (2024)
Avg. of the number of days in which the amplitude of temperature is >75th percentile of daily amplitude of temperature	▲								
Avg. of the number of days in which the amplitude of temperature is >90 th percentile of daily amplitude temperature	▲								
Max. of the max. daily precipitation	▲								
Avg. of the max. daily precipitation	▲								
Avg. of the avg. daily precipitation	▲								
Avg. of the number of days in which the max. precipitation is >75th of daily max. precipitation	▲								
Avg. of the number of days in which the max. precipitation is >90th of daily max. precipitation	▲								
Max. level 1 soil temperature		▲							
Mean level 1 soil temperature		▲							
Min. level 1 soil temperature		▲							
Max. vapor pressure		▲							
Mean vapor pressure		▲							
Min. vapor pressure		▲							
Evaporation							▲		
Daily fire count					▲				
Temperature difference (Difference, 12h to previous 24h)							▲		
Heat Index (Apparent Temperature)									
Thermal Amplitude									
Air/Atmospheric/Surface pressure									
Air/Atmospheric/Surface pressure mean		▲		▲		▲	▲		▲
Air/Atmospheric/Surface pressure min.		▲		▲			▲		
Air/Atmospheric/Surface pressure max.		▲	▲	▲			▲		
Air/Atmospheric/Surface pressure difference							▲		
Dew Point									
Max. dew point		▲							
Mean dew point		▲							△
Min dew point		v							
Relative Humidity									
Mean atmospheric humidity							▲		
Max atmospheric humidity							▲		

Table 5 (continued)

Study	Chou-Chen and Barboza (2024)	Guo et al. (2024)	Zhong et al. (2024)	Yang et al. (2025)	de Souza Tadano et al. (2024)	Mutlu et al. (2024)	Kurucz et al. (2024)	Susmann et al. (2025)	Cappelli et al. (2024)
NO									
NO2		▲	▲						▲
Nox			▲						
Sox									
SO2		▲				▲			
CO		▲							▲
CO2									
O3		▲	▲						▲
Black Carbon (BC)					▲				
Lead (Pb)									
Air Quality Index (AQI)									
Pollen - Olea (max)									
Pollen - Amaranta (max)									
Pollen - Cupra (max)									
Pollen - Graminea (max)									
Max. value of the European Regional Pollution Index (ERPI)									
Number of days with air pollution episode ERPI value >= 50									
Number of stations that showed air pollution episode									
Predictor/Type	Public Health Trends								
Residuals of Influenza model									
Influenza visits (%)								▲	
Acute diarrhea								▲	
Chickenpox								▲	
Infectious and Parasitic Diseases		▲							
Not Categorised		▲							
Respiratory Infection		▲							
Maternal conditions		▲							
Perinatal conditions		▲							
Nutritional deficiencies		▲							
Malignant neoplasms		▲							
Other neoplasms		▲							
Diabetes mellitus		▲							
Endocrine disorders		▲							
Mental disorders		▲							
Neurological and sense disorders		▲							
Cardiovascular disease		▲							
Chronic respiratory disease		▲							
Digestive disease		▲							
Genitourinary disorders		▲							
Skin diseases		▲							
Musculoskeletal disease		▲							
Congenital Abnormalities		▲							
Oral Diseases		▲							
Unintentional injuries		▲							
Intentional injuries		▲							

Table 5 (continued)

Study	Chou-Chen and Barboza (2024)	Guo et al. (2024)	Zhong et al. (2024)	Yang et al. (2025)	de Souza Tadano et al. (2024)	Mutlu et al. (2024)	Kurucz et al. (2024)	Susmann et al. (2025)	Cappelli et al. (2024)
Mean of all available Hospital Beds									
Sum of all available Hospital Beds									
Daily Admissions (avg/std)	▲		▲			▲	▲	▲	
Predictor/Type	Not reported								
MV, ALT, and Other factors (?)									

This table presents the base predictors, i.e., not including the lagged predictors

△ Candidate Predictor

▲ Final predictor used

(1) Did not report the final predictors used for each model

(?) Indicates that the authors did not provide a definition for this variable

* Indicates the use of avg, max, and min values

** Indicates the use of AQI values

Abbreviations

AI	Artificial intelligence
ANN	Artificial neural network
AUC	Area under the curve
AR	Autoregressive model
ARIMA	Autoregressive integrated moving average
ARIMAX	Autoregressive integrated moving average with exogenous features
BJML	Daily minimal bed requirement (<i>Besoin Journalier Minimal en Lits</i>)
BiLSTM	Bidirectional long short-term memory
CD	Cerebrovascular disease
CHARMS	Checklist for critical appraisal and data extraction for systematic reviews of prediction modeling studies
CI	Cerebral Infarction
CNN	Convolutional neural network.
CNN-LSTM	Hybrid convolutional neural network + long-short term memory
COPD	Chronic obstructive pulmonary disease
CV	Cross-validation
CVD	Cardiovascular diseases
DRF	Distributed random forest
DL	Deep learning
DT	Decision trees
ED	Emergency department
ELSTM	enhanced long short-term memory
EPV	events-per-variable
e-EPV	estimated events-per-variable
ELM	Extreme learning machines
ESN	Echo state networks
ENMD	Endocrine, nutritional, and metabolic disorder
ETS	Error-trend-seasonal
FNN	feedforward neural networks
FH	forecast horizon
FS	Floating search
GAM	Generalized additive model
GBDT	Gradient boosting decision tree
GBM	Gradient boosting machine
GLM	Generalized linear model
GRF	Generalized random forests
GRU	Gated recurrent unit
ICH	Intracerebral hemorrhage
IHD	Ischemic heart disease
KNN	k-nearest neighbors algorithm
LASSO	least absolute shrinkage and selection operator
LMS	Least mean squares
LR	Linear regression

LSTM	Long short-term memory
LWLR	Locally weighted linear regression
MLWLR	Multiple Locally weighted linear regression
MAE	Mean absolute error
MASE	Mean absolute scaled error
MAPE	Mean absolute percentage error
MARS	Multivariate adaptive regression splines
MBE	Mean bias error
ML	Machine learning
MLP	Multilayer perceptron
MLR	Multiple linear regression
MOV	Mean observed value
MSE	Mean square error
NB	Naive bayes
NR	Not reported
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PROBAST	Prediction model risk of bias assessment tool
PROSPERO	International prospective register of systematic reviews
QGAM	Quantile generalized additive model
QR	Quantile regression
R	correlation coefficient
RMSD	Residual mean square deviation
RBF	Radial basis function network
RoB	Risk of bias
ROC	Receiver operating characteristic curve
RF	Random forest
RH	Relative humidity
RG	Ridge regression
RLS	Recursive least squares
RMSE	Root-mean-square error
RNN	Recurrent neural network1
SA	Simulated annealing
SAH	Subarachnoid hemorrhage
SARIMA	Seasonal autoregressive integrated moving average
SCNN-BiLSTM	Stacked-dilated-causal BiLSTM
Snaive	Seasonal naive
SFFS	Sequential forward floating selection
STEMI/NSTEMI	non-ST-elevated myocardial infarction and ST-elevated myocardial infarction
SVM	Support vector machine
SVR	Support vector regression
TAPE	Total absolute percentage error
TLRN	Time lagged recurrent networks
TF-IDF	Term frequency-inverse document frequency algorithm
VAR	Vector autoregression
XGBoost	eXtreme gradient boosting

XRT Extremely randomized trees

Acknowledgements

We express our gratitude to Professor Christophe Garcia, Professor Engelbert Mephu Nguifo, and Professor Stefan Duffner for their valuable contributions in reviewing this work.

Author contributions

Levi Monteiro Martins: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing—original draft, Writing—Review & Editing, Visualization. Elsa Coz: Validation, Writing—Review & Editing. Delphine Maucort-Boulch: Conceptualization, Resources, Writing—Review & Editing, Supervision, Project administration, Funding acquisition. Mohand-Saïd Hacıd: Conceptualization, Resources, Writing—Review & Editing, Supervision, Project administration, Funding acquisition.

Funding

This study is fruit of the IADoc@UdL *Programme de l'Université de Lyon*. LMM is funded by the Agence Nationale de la Recherche (ANR) under the grant ANR-20-THIA-0007-01 and by the Hospices Civils de Lyon (HCL). The sponsor did not participate in the design, data collection, analysis, and interpretation of this study, nor in the writing of the report or the decision to submit the article for publication.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Received: 10 February 2025 Accepted: 12 April 2025

Published online: 05 June 2025

References

- Athey S, Tibshirani J, Wager S (2019) Generalized random forests. *Ann Stat* 47(2):1148–1178. <https://doi.org/10.1214/18-AOS1709>
- Atkeson CG, Moore AW, Schaal S (1997) Locally weighted learning. In: Aha DW (ed) *Lazy Learning*. Springer, Dordrecht, pp 11–73. https://doi.org/10.1007/978-94-017-2053-3_2
- Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *ArXiv*. <https://doi.org/10.48550/arXiv.1803.01271>
- Boehmke B, Greenwell BM (2019) *Hands-on machine learning with R*. Chapman and Hall/CRC, New York
- Boulesteix AL, Stierle V, Hapfelmeier A (2015) Publication Bias in methodological computational research. *Cancer Inform*. <https://doi.org/10.4137/CIN.530747>
- Breiman L (2001) Random Forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman L, Friedman J, Olshen R et al (2017) *Classification and regression trees*. Routledge, New York
- Cappelli F, Castronuovo G, Gimaldi S et al (2024) Random forest and feature importance measures for discriminating the most influential environmental factors in predicting cardiovascular and respiratory diseases. *Int J Environ Res Public Health* 21(7):867. <https://doi.org/10.3390/ijerph21070867>
- Cengiz MA, Terzi Y (2012) Comparing models of the effect of air pollutants on hospital admissions and symptoms for chronic obstructive pulmonary disease. *Central Eur J Public Health*. 20(4): 282–286. <https://doi.org/10.21101/cejpha3757>
- Cerqueira V, Torgo L, Mozetic I (2020) Evaluating time series forecasting models: an empirical study on performance estimation methods. *Mach Learn* 109(11):1997–2028. <https://doi.org/10.1007/s10994-020-05910-7>
- Cevid D, Michel L, Näf J et al (2022) Distributional random forests: heterogeneity adjustment and multivariate distributional regression. *J Mach Learn Res* 23(333):1–79
- Chen J, Li H, Luo L et al (2019) Machine learning-based forecast of hemorrhagic stroke healthcare service demand considering air pollution. *J Healthcare Eng* 2019:7463242. <https://doi.org/10.1155/2019/7463242>
- Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, pp 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen W, Linthicum B, Argon NT et al (2020) The effects of emergency department crowding on triage and hospital admission decisions. *Am J Emergency Med* 38(4):774–779. <https://doi.org/10.1016/j.ajem.2019.06.039>
- Cho K, van Merriënboer B, Bahdanau D et al (2014) On the properties of neural machine translation: encoder-decoder approaches. *arXiv*. <https://doi.org/10.3115/v1/W14-4012>
- Chou-Chen SW, Barboza LA (2024) Forecasting hospital discharges for respiratory conditions in Costa Rica using climate and pollution data. *Math Biosci Eng* 21(7):6539–6558. <https://doi.org/10.3934/mbe.2024285>
- Cohn DA, Ghahramani Z, Jordan MI (1996) Active learning with statistical models. *J Artif Intell Res* 4:129–145. <https://doi.org/10.1613/jair.295>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1007/BF00994018>
- de Souza Tadano Y, Siqueira HV, Alves TA (2016) Unorganized machines to predict hospital admissions for respiratory diseases. In: 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCL), Cartagena, pp 1–6. <https://doi.org/10.1109/LA-CCL.2016.7885699>
- de Souza Tadano Y, Potgieter-Vermaak S, Siqueira HV et al (2024) Predicting health impacts of wildfire smoke in Amazonas basin, Brazil. *Chemosphere* 367:143688. <https://doi.org/10.1016/j.chemosphere.2024.143688>
- Dey R, Salem FM (2017) Gate-variants of gated recurrent unit (GRU) neural networks. In: 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS). IEEE, Boston, pp 1597–1600. <https://doi.org/10.1109/MWSCAS.2017.8053243>
- Duwalage KI, Burkett E, White G et al (2020) Forecasting daily counts of patient presentations in Australian emergency departments using statistical models with time-varying predictors. *Emerg Med Australas* 32(4):618–625. <https://doi.org/10.1111/1742-6723.13481>
- Fasiolo M, Wood SN, Zaffran M et al (2020) Fast calibrated additive quantile regression. *J Am Stat Assoc* 116(535):1402–1412. <https://doi.org/10.1080/01621459.2020.1725521>
- Feng C, Wang H, Lu N et al (2014) Log-transformation and its implications for data analysis. *Shanghai Archiv Psychiatry* 26(2):105–109. <https://doi.org/10.3969/j.issn.1002-0829.2014.02.009>
- Frandsen TF, Eriksen MB, Hammer DMG et al (2021) Using Embase as a supplement to PubMed in Cochrane reviews differed across fields. *J Clin Epidemiol* 133:24–31. <https://doi.org/10.1016/j.jclinepi.2020.12.022>
- Frank E, Trigg L, Holmes G et al (2000) Technical note: naive bayes for regression. *Mach Learn* 41(1):5–25. <https://doi.org/10.1023/A:1007670802811>
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann stat* 29(5):1189–1232
- Gafni-Pappas G, Khan M (2023) Predicting daily emergency department visits using machine learning could increase accuracy. *Am J Emerg Med* 65:5–11. <https://doi.org/10.1016/j.ajem.2022.12.019>
- Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. *Mach Learn* 63(1):3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Guo G, Wang H, Bell D, et al (2003) KNN Model-Based Approach in Classification. In: Meersman R, Tari Z, Schmidt DC (eds) *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*. Lecture Notes in Computer Science, vol 2888. Springer, Berlin, Heidelberg, pp 986–996. https://doi.org/10.1007/978-3-540-39964-3_62

- Guo P, Tewari P, Wen Choo EL, et al (2024) Forecasting infectious and parasitic disease emergency department attendances using high-dimensional time series data. In: 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, pp 603–610. <https://doi.org/10.1109/CAI59869.2024.00119>
- Hastie T, Tibshirani R (1986) Generalized additive models. *Stat Sci* 1(3):297–310. <https://doi.org/10.1214/ss/1177013604>. (publisher: Institute of Mathematical Statistics)
- Haykin SS (2002) Adaptive filter theory. Pearson Education, India
- Hochreiter S, Schmidhuber J (1996) LSTM can solve hard long time lag problems. *Advances in neural information processing systems* 9
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hovland AE, Wagner AM, Pierce KM, et al (2013) Environmental surveillance modeling: A predictive respiratory alert model for the Shenandoah Valley, Virginia. In: 2013 IEEE Systems and Information Engineering Design Symposium, pp 167–172. <https://doi.org/10.1109/SIEDS.2013.6549513>
- Hu Z, Qiu H, Su Z et al (2020) A stacking ensemble model to predict daily number of hospital admissions for cardiovascular diseases. *IEEE Access* 8:138719–138729. <https://doi.org/10.1109/ACCESS.2020.3012143>
- Hyndman R, Koehler AB, Ord JK, et al (2008) Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media
- Jiang S, Liu Q, Ding B (2022) A systematic review of the modelling of patient arrivals in emergency departments. *Quant Imaging Med Surg* 13(3):1957971. <https://doi.org/10.2103/qjms-22-2687>
- Jiao K, Xu M, Liu M (2018) Health status and air pollution related socioeconomic concerns in urban China. *Int J Equity Health* 17(1):18. <https://doi.org/10.1186/s12939-018-0719-y>
- Jing LI, Yu LB, Jian WZ et al (2020) Application Research on Gated Recurrent Unit Deep Learning Prediction and Graded Early Warning of Emergency Department Visits Based on Meteorological Environmental Data. *Biomedical and Environmental Sciences* 33(10):817–820. <https://doi.org/10.3967/bes2020.111>
- Jones SS, Evans RS, Allen TL et al (2009) A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of Biomedical Informatics* 42(1):123–139. <https://doi.org/10.1016/j.jbi.2008.05.003>
- Ke G, Meng Q, Finley T, et al (2017) Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30
- Khatri KL, Tamil LS (2017) Early Detection of Peak Demand Days of Chronic Respiratory Diseases Emergency Department Visits Using Artificial Neural Networks. *IEEE Journal of Biomedical and Health Informatics* 22(1):285–290. <https://doi.org/10.1109/JBHI.2017.2698418>
- Koenker R (2005) Quantile regression, vol 38. Cambridge University Press
- Kurucz VC, Schenk J, Veelo DP et al (2024) Prediction of emergency department presentations for acute coronary syndrome using a machine learning approach. *Scientific Reports* 14:23125. <https://doi.org/10.1038/s41598-024-73291-1>
- Lam MT, McDiarmid M (2016) Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014. *Journal of the Medical Library Association* 104(4):284–289. <https://doi.org/10.5195/jmla.2016.141>
- Leathwick JR, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling* 199(2):188–196. <https://doi.org/10.1016/j.ecolmodel.2006.05.022>
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- LeDell E, Gill N, Aiello S, et al (2020) h2o: R interface for the “H2O” Scalable machine learning platform. R package version 3(0.1):2
- Lin YC, Tsai CH, Hsu HT, et al (2021) Using Machine Learning to Analyze and Predict the Relations Between Cardiovascular Disease Incidence, Extreme Temperature and Air Pollution. In: 2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), pp 234–237. <https://doi.org/10.1109/ECBIOS51820.2021.9510479>
- Lq Chen, Wang C, Si Song (2022) Software defect prediction based on nested-stacking and heterogeneous feature selection. *Complex Intell Syst* 8(4):3333–3348. <https://doi.org/10.1007/s40747-022-00676-y>
- Lu X, Qiu H (2023) Explainable prediction of daily hospitalizations for cerebrovascular disease using stacked ensemble learning. *BMC Medical Informatics and Decision Making* 23(1):59. <https://doi.org/10.1186/s12911-023-02159-7>
- Medina Quero J, Lopez Medina MA, Salguero Hidalgo A et al (2018) Predicting the Urgency Demand of COPD Patients From Environmental Sensors Within Smart Cities With High-Environmental Sensitivity. *IEEE Access* 6:25081–25089. <https://doi.org/10.1109/ACCESS.2018.2828652>
- Mendo IR, Marques G, de la Torre Díez I et al (2021) Machine Learning in Medical Emergencies: a Systematic Review and Analysis. *Journal of Medical Systems* 45:88. <https://doi.org/10.1007/s10916-021-01762-3>
- Meng G, Tan Y, Fang M, et al (2015) Meteorological Factors Related to Emergency Admission of Elderly Stroke Patients in Shanghai: Analysis with a Multilayer Perceptron Neural Network. *Medical Science Monitor* 21:3600–3607. <https://doi.org/10.12659/MSM.895334>
- Mienye ID, Sun Y (2022) A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access* 10:99129–99149. <https://doi.org/10.1109/ACCESS.2022.3207287>
- Ministère de la Transition écologique et de la Cohésion des territoires (2022) Pollution de l'air : origines, situation et impacts. <https://www.ecologie.gouv.fr/pollution-lair-origines-situation-et-impacts>, (In French)
- Moisen GG, Frescino TS (2002) Comparing five modelling techniques for predicting forest characteristics. *Ecological modelling* 157(2–3):209–225. [https://doi.org/10.1016/S0304-3800\(02\)00197-7](https://doi.org/10.1016/S0304-3800(02)00197-7)
- Moons KGM, Groot JAHd, Bouwmeester W et al (2014) Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLOS Medicine* 11(10):e1001744. <https://doi.org/10.1371/journal.pmed.1001744>
- Morlidige S (2013) How good is a “good” forecast? Forecast errors and their avoidability. *Foresight: The International Journal of Applied Forecasting* 30:5–11
- Mossop B, Bismil R, Rahman QA (2023) Forecasting Hospital Mental-Health Admissions with a Novel Hybrid Deep Learning Architecture. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Istanbul, pp 4093–4100. <https://doi.org/10.1109/BIBM58861.2023.10385786>
- Moustris KP, Douros K, Nastos PT et al (2012) Seven-days-ahead forecasting of childhood asthma admissions using artificial neural networks in Athens, Greece. *International Journal of Environmental Health Research* 22(2):93–104. <https://doi.org/10.1080/09603123.2011.605876>
- Muhammad S, Long X, Salman M (2020) COVID-19 pandemic and environmental pollution: A blessing in disguise? *Science of The Total Environment* 728:138820. <https://doi.org/10.1016/j.scitotenv.2020.138820>
- Murphy KP (2012) Machine learning: a probabilistic perspective. MIT press
- Mutlu A, Aydin Keskin G, Çıldır I, (2024) Predicting hospital admissions for upper respiratory tract complaints: An artificial neural network approach integrating air pollution and meteorological factors. *Environmental Monitoring and Assessment* 196:759. <https://doi.org/10.1007/s10661-024-12908-4>
- Nestor Maslej, Loredana Fattorini, Raymond Perrault, et al (2024) The AI Index 2024 Annual Report. <https://hai.stanford.edu/ai-index/2024-ai-index-report>
- Nijman S, Leeuwenberg A, Beekers I et al (2022) Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *Journal of Clinical Epidemiology* 142:218–229. <https://doi.org/10.1016/j.jclinepi.2021.11.023>
- Ogata S, Takegami M, Ozaki T et al (2021) Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. *Nature Communications* 12:4575. <https://doi.org/10.1038/s41467-021-24823-0>
- Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. <https://doi.org/10.1136/bmj.n71>
- Pearce S, Marchand T, Shannon T et al (2023) Emergency department crowding: an overview of reviews describing measures causes, and harms. *Internal and Emergency Medicine* 18:1137–1158. <https://doi.org/10.1007/s11739-023-03239-2>
- Peduzzi P, Concato J, Kemper E et al (1996) A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49(12):1373–1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)

- van der Ploeg T, Austin PC, Steyerberg EW (2014) Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology* 14(1):137. <https://doi.org/10.1186/1471-2288-14-137>
- Qiu H, Luo L, Su Z et al (2020) Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. *BMC Medical Informatics and Decision Making* 20:83. <https://doi.org/10.1186/s12911-020-1101-8>
- Ram S, Zhang W, Williams M et al (2015) Predicting Asthma-Related Emergency Department Visits Using Big Data. *IEEE Journal of Biomedical and Health Informatics* 19(4):1216–1223. <https://doi.org/10.1109/JBHI.2015.2404829>
- Ravi D, Wong C, Deligianni F et al (2017) Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics* 21(1):4–21. <https://doi.org/10.1109/JBHI.2016.2636665>
- Rob J Hyndman, George Athanasopoulos (2018a) *Autoregressive models (Forecasting: Principles and Practice)*, 2nd edn. OTexts, Melbourne, Australia, <https://otexts.com/fpp2/AR.html>, accessed 19 Oct 2021
- Rob J Hyndman, George Athanasopoulos (2018b) *Complex seasonality (Forecasting: Principles and Practice)*, 2nd edn. OTexts, Melbourne, Australia, <https://otexts.com/fpp2/complexseasonality.html>, accessed 19 Oct 2021
- Rob J Hyndman, George Athanasopoulos (2018c) *Forecasting with ETS models (Forecasting: Principles and Practice)*, 2nd edn. OTexts, Melbourne, Australia, <https://otexts.com/fpp2/ets-forecasting.html>, accessed 19 Oct 2021
- Rob J Hyndman, George Athanasopoulos (2018d) *Some simple forecasting methods (Forecasting: Principles and Practice)*, 2nd edn. OTexts, Melbourne, Australia, <https://otexts.com/fpp2/simple-methods.html>, accessed 19 Oct 2021
- Schneider SM, Gallery ME, Schafermeyer R et al (2003) Emergency department crowding: a point in time. *Annals of emergency medicine* 42(2):167–172. <https://doi.org/10.1067/mem.2003.258>
- Sciannameo V, Goffi A, Maffei G et al (2022) A deep learning approach for Spatio-Temporal forecasting of new cases and new hospital admissions of COVID-19 spread in Reggio Emilia, Northern Italy. *Journal of Biomedical Informatics* 132:104132. <https://doi.org/10.1016/j.jbi.2022.104132>
- Siami-Namini S, Tavakoli N, Namin AS (2019) The performance of LSTM and BiLSTM in forecasting time series. In: 2019 IEEE International conference on big data (Big Data). IEEE, Los Angeles, pp 3285–3292. <https://doi.org/10.1109/BigData47090.2019.9005997>
- Silva E, Pereira MF, Vieira JT et al (2023) Predicting hospital emergency department visits accurately: A systematic review. *The International Journal of Health Planning and Management* 38(4):904–917. <https://doi.org/10.1002/hpm.3629>
- Siqueira HV, Bacalhau ET, Casacio L et al (2023) Hybrid unorganized machines to estimate the number of hospital admissions caused by pm10 concentration. *Environmental Science and Pollution Research* 30:113175–113192. <https://doi.org/10.1007/s11356-023-30180-w>
- Song X, Zhang X, Wang X, et al (2022) Prediction of daily patient visits in respiratory department using deep learning. In: 2022 18th International Conference on Computational Intelligence and Security (CIS), Chengdu, pp 69–73. <https://doi.org/10.1109/CIS58238.2022.00022>
- Song Y, Liang J, Lu J et al (2017) An efficient instance selection algorithm for k nearest neighbor regression. *Neurocomputing* 251:26–34. <https://doi.org/10.1016/j.neucom.2017.04.018>
- Song YY, Ying LU (2015) Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry* 27(2):130. <https://doi.org/10.11919/j.jissn.1002-0829.215044>
- Stock JH, Watson MW (2001) Vector autoregressions. *Journal of Economic perspectives* 15(4):101–115. <https://doi.org/10.1257/jep.15.4.101>
- Stone CJ, Hansen MH, Kooperberg C et al (1997) Polynomial splines and their tensor products in extended linear modeling: 1994 Wald memorial lecture. *The Annals of statistics* 25(4):1371–1470. <https://doi.org/10.1214/aos/1031594728>
- Stonko DP, Dennis BM, Betzold RD et al (2018) Artificial intelligence can predict daily trauma volume and average acuity. *Journal of Trauma and Acute Care Surgery* 85(2):393. <https://doi.org/10.1097/TA.0000000000001947>
- Susmann H, Chambaz A, Josse J et al (2025) Probabilistic prediction of arrivals and hospitalizations in emergency departments in Île-de-France. *International Journal of Medical Informatics* 195:105728. <https://doi.org/10.1016/j.jimedinf.2024.105728>
- Svetunkov I (2024) 4.1 ETS taxonomy | Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM). <https://openforecast.org/adam/ETSTaxonomy.html>, accessed 04 Oct 2024
- Tashman LJ (2000) Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 16(4):437–450. [https://doi.org/10.1016/S0169-2070\(00\)00065-0](https://doi.org/10.1016/S0169-2070(00)00065-0)
- Tavera Busso I, Rodríguez Núñez M, Amarillo AC et al (2021) Modeling air pollution-related hospital admissions employing remote sensing and geographical information systems. *Atmospheric Environment* 261:118502. <https://doi.org/10.1016/j.atmosenv.2021.118502>
- Tomaszewski R (2021) A study of citations to STEM databases: ACM Digital Library, Engineering Village, IEEE Xplore, and MathSciNet. *Scientometrics* 126:1797–1811. <https://doi.org/10.1007/s11192-020-03795-w>
- Tuominen J, Lomio F, Oksala N et al (2022) Forecasting daily emergency department arrivals using high-dimensional multivariate data: a feature selection approach. *BMC Medical Informatics and Decision Making* 22:134. <https://doi.org/10.1186/s12911-022-01878-7>
- Usmani RSA, Pillai TR, Hashem IAT et al (2021) Air pollution and cardiorespiratory hospitalization, predictive modeling, and analysis using artificial intelligence techniques. *Environmental Science and Pollution Research* 28:56759–56771. <https://doi.org/10.1007/s11356-021-14305-7>
- Vassar M, Yerokhin V, Sinnott PM et al (2017) Database selection in systematic reviews: an insight through clinical neurology. *Health Information & Libraries Journal* 34(2):156–164. <https://doi.org/10.1111/hir.12176>
- van Walraven C, Austin P (2012) Administrative database research has unique characteristics that can risk biased results. *Journal of Clinical Epidemiology* 65(2):126–131. <https://doi.org/10.1016/j.jclinepi.2011.08.002>
- Wang D, Zheng Y, Lian H et al (2021) High-Dimensional Vector Autoregressive Time Series Modeling via Tensor Decomposition. *Journal of the American Statistical Association* 117(539):1338–1356. <https://doi.org/10.1080/01621459.2020.1855183>
- Wargon M, Guidet B, Hoang TD et al (2009) A systematic review of models for forecasting the number of emergency department visits. *Emergency Medicine Journal* 26:395–399. <https://doi.org/10.1136/emj.2008.062380>
- Wolff RF, Moons KG, Riley RD et al (2019) PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine* 170(1):51–58. <https://doi.org/10.7326/M18-1376>
- Woods D, Trewheellar K (1998) Medline and Embase complement each other in literature searches. *BMJ: British Medical Journal* 316(7138):1166. <https://doi.org/10.1136/bmj.316.7138.1166>
- World Health Organization (2021) Ambient (outdoor) air pollution. [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health), accessed 07 Dec 2022
- Yang Y, Zhang M, Zhang J et al (2025) Medical meteorological forecast for ischemic stroke: random forest regression vs long short-term memory model. *International Journal of Biometeorology* 69:397–402. <https://doi.org/10.1007/s00484-024-02818-y>
- Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- Zhang Q, Wang K, Guo H, et al (2018) Base on ETS model for Forecasting Emergency Department Visits. In: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, pp 2148–2151. <https://doi.org/10.1109/IAEAC.2018.8577491>
- Zhong J, Abreu T, Heidet M, et al (2024) Machine Learning for Emergency Service Optimization: A Real-World Application. In: 2024 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Kingston, pp 387–391. <https://doi.org/10.1109/CCECE59415.2024.10667106>
- Zy Huang, Lin S, Li Long et al (2020) Predicting the morbidity of chronic obstructive pulmonary disease based on multiple locally weighted linear regression model with K-means clustering. *Int J Med Inform* 139:104141. <https://doi.org/10.1016/j.jimedinf.2020.104141>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.