



---

*Research article*

## Forecasting hospital discharges for respiratory conditions in Costa Rica using climate and pollution data

Shu Wei Chou-Chen<sup>1,2,\*</sup>, and Luis A. Barboza<sup>1,3,\*</sup>

<sup>1</sup> Centro de Investigación en Matemática Pura y Aplicada, Universidad de Costa Rica, Costa Rica

<sup>2</sup> Escuela de Estadística, Universidad de Costa Rica, Costa Rica

<sup>3</sup> Escuela de Matemática, Universidad de Costa Rica, Costa Rica

\* **Correspondence:** Email: [shuwei.chou@ucr.ac.cr](mailto:shuwei.chou@ucr.ac.cr), [luisalberto.barboza@ucr.ac.cr](mailto:luisalberto.barboza@ucr.ac.cr)

**Abstract:** Respiratory diseases represent one of the most significant economic burdens on health-care systems worldwide. The variation in the increasing number of cases depends greatly on climatic seasonal effects, socioeconomic factors, and pollution. Therefore, understanding these variations and obtaining precise forecasts allows health authorities to make correct decisions regarding the allocation of limited economic and human resources. We aimed to model and forecast weekly hospitalizations due to respiratory conditions in seven regional hospitals in Costa Rica using four statistical learning techniques (Random Forest, XGboost, Facebook’s Prophet forecasting model, and an ensemble method combining the above methods), along with 22 climate change indices and aerosol optical depth as an indicator of pollution. Models were trained using data from 2000 to 2018 and were evaluated using data from 2019 as testing data. During the training period, we set up 2-year sliding windows and a 1-year assessment period, along with the grid search method to optimize hyperparameters for each model. The best model for each region was selected using testing data, based on predictive precision and to prevent overfitting. Prediction intervals were then computed using conformal inference. The relative importance of all climatic variables was computed for the best model, and similar patterns in some of the seven regions were observed based on the selected model. Finally, reliable predictions were obtained for each of the seven regional hospitals.

**Keywords:** temperature; precipitation; climate change, aerosol optical depth; statistical learning; forecasting; hospitalization; respiratory diseases

---

### 1. Introduction

Respiratory diseases are conditions that affect the organs and tissues within the lungs and airway systems, leading to difficulty in breathing. These illnesses are currently the leading contributors to

the global burden of diseases when assessed through disability-adjusted life-years. Moreover, the rising healthcare costs associated with these diseases are placing a growing strain on the economies of nations worldwide, particularly due to the “big five” conditions: Chronic obstructive pulmonary disease, asthma, acute respiratory infections, tuberculosis, and lung cancer [1].

To maintain a balanced health economy, it is crucial to understand the distribution and the incidence of these diseases to optimize healthcare systems, ensure efficient resource allocation, and enhance healthcare access and quality. Therefore, our focus is to provide a reliable forecast for respiratory hospitalizations at the hospital level in Costa Rica by using potential climatic factors as inputs in predictive models, enabling health authorities to anticipate future hospital needs. Specifically, three different competitive predictive models (Random Forest, XGBoost, and PROPHET) along with an ensemble method are used and compared.

Extensive research indicates that climate influences respiratory health, resulting in higher rates of hospitalization and mortality in varying seasons [2–6]. Furthermore, it is well known that human activities significantly impact the short-term and long-term climate by releasing greenhouse gases and other pollutants into the atmosphere, leading to more extreme weather events in recent years. As a consequence, researchers worldwide have been studying the effects of pollution and climatological factors on human health in different regions of the world [7–11].

Specifically with respiratory diseases, climate factors are associated with extreme weather [5,6,12], and pollution, such as particulate matter (PM) [4,13]. Airborne PM is a complex mixture of different chemical species originating from numerous sources. These particles can be a product of combustion, suspension of soil materials, suspension of substances from the sea, and can also be formed by chemical reactions in the atmosphere [14]. Furthermore, elevated concentrations of PM not only substantially heighten health risks but also contribute significantly to the challenges associated with climate change. Therefore, the monitoring of PM is essential for comprehending and mitigating air pollution, aiming to enhance public health. Nevertheless, the measurement of PM presents challenges in developing countries due to the constrained monitoring infrastructure, resulting in gaps in data analysis both temporally and spatially.

On the other hand, an aerosol is defined as a stable suspension of solid and liquid particles in gas. It is possible to obtain related measurements with the facilitation of remote sensing techniques. Aerosol optical depth (AOD), a satellite-derived metric that quantifies the presence of aerosols across the entire atmospheric column, is a good option. Several researchers around the world have found that the driving factors of AOD are related to urban and economic development, agricultural activities, industrialization, landscape aggregation, and regional transportation [15,16]. Moreover, AOD is associated with socio-economic factors, such as GDP, industry, and vehicle density [17].

Costa Rica, a country with an area of 51,179  $Km^2$ , is located in Central America with an estimated population of approximately 5,003,402 inhabitants in 2018. It is a stable democratic country with a socioeconomic development model based on the opening and economic liberalization, upholding human rights, and ensuring universal basic access to essential goods and services, such as education and health system, while preserving extensive natural resources. [18].

Although Costa Rica is renowned for its biodiversity, with most environmental policies focusing on the protection and conservation of natural resources, unregulated land use has been observed, resulting in significant environmental impacts. This issue is not limited to the Great Metropolitan Area, the country’s most urban and densely populated region, but is also prevalent in other parts of the national

territory. Consequently, this form of human development can intensify vulnerability to natural disasters and extreme weather conditions [18].

Regarding to its health system, since the 1940s Costa Rica prompted its national health system and social investment by creating the Costa Rican Social Security Fund (CCSS, from its Spanish acronym *Caja Costarricense de Seguro Social*). The basis of the health system in Costa Rica is universal and public, which covers 95% of the Costa Rican population [18, 19]. The CCSS carries all health care functions in the country and divides the territory into 7 Health Regions (*región sanitaria* in Spanish), and at the same time, it is divided into 104 Health Areas (AS, from its Spanish acronym *áreas de salud*) and then, 1045 Basic Team for Comprehensive Health Care (EBAIS, from its Spanish acronym *Equipos Básicos de Atención Integral de Salud*). Each AS caters to a population ranging from 15,000 to 40,000 individuals in rural areas and 30,000 to 60,000 individuals in urban areas. Subsequently, these ASs are subdivided into 1,045 segments, each overseen by an EBAIS. Each segment, in turn, serves approximately 4,000 people.

The health service in Costa Rica operates on three distinct levels of attention, interconnected through referral and counter-referral mechanisms. The initial level of service acts as the primary entry point to the healthcare system, providing essential services within each EBAIS at the community level. Additionally, there are two possibilities in which a person can access the health system: (1) an insured person can be attended by a private doctor who is registered by the CCSS and access reference to more specialized attention, and (2) an insured person is attended by a doctor hired by the company. The second level supports the first level through a network of 10 main clinics, 13 peripheral hospitals, and 7 regional hospitals. It provides specialized outpatient services, interventions, hospitalization, and medical-surgical treatments for medical specialties and high-demand areas like dermatology, urology, and ophthalmology, all of low complexity. Finally, the third level includes outpatient and inpatient services of greater complexity and specialization that require high technology and level of specialization. It is provided through 3 national hospitals and 6 specialized hospitals; the area of influence of this level covers the territory of several provinces.

Due to this health administrative flow, the accurate diagnostics of respiratory conditions that require hospitalization are registered in hospital discharge events (second level), instead of hospital entrance (first level). As a consequence, the limitation of this data lies in the presence of a lagged and unknown effect, necessitating consideration in the analysis. Beyond the delayed impact associated with hospital admissions and discharges, there is also a temporal lag effect pertaining to climate and pollution variables.

In the literature, most related studies are linked to statistical association between climate and hospitalizations [2–6, 20]. Additionally, some concentrate on forecasting hospitalizations specifically within emergency departments and respiratory diseases [21–29].

Our main goal is to predict hospital discharges due to respiratory diseases at the secondary health-care level of the system, in the hope of enhancing the country's healthcare economy. Specifically, by utilizing pollution and climatic indices, we aim to obtain accurate forecasts of respiratory hospitalizations in seven regions by comparing four different predictive models. Our goal is to provide easily implemented models to health authorities in order to make accurate predictions and informed budget decisions in the near future. To the best of our knowledge, this approach is the first predictive study undertaken in the region with these distinctive characteristics. Moreover, this study may serve as a potential tool to be replicated in other countries with similar diverse climatic characteristics around the

world.

This paper is divided as follows: Section 2 describes the data and the implemented methodologies. Section 3 presents the analysis and forecasting results. Finally, Section 4 discusses the prediction performance, limitations, and future work.

## 2. Materials and methods

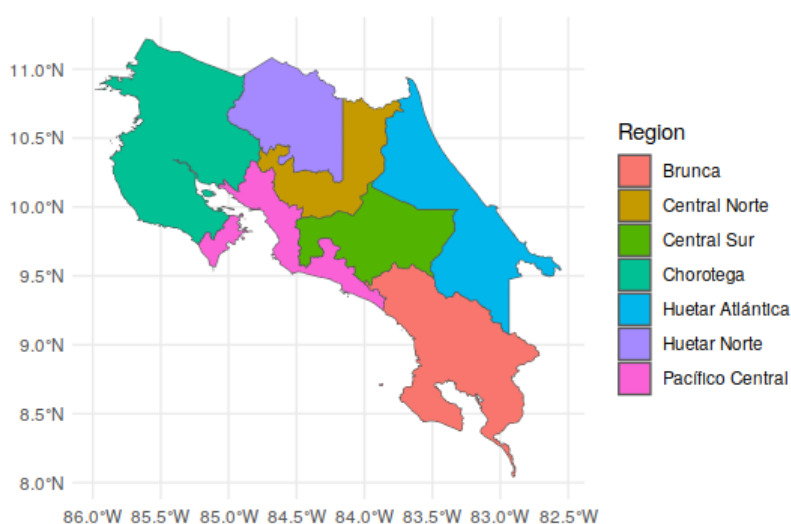
This section first provides an overview of the primary datasets used in our study. Then, the statistical methods applied to predict the hospital discharge are outlined.

### 2.1. Data

We describe the main variable of interest, respiratory hospitalizations, followed by the input data comprising climate and pollution data, along with their sources.

#### 2.1.1. Respiratory hospitalization

Weekly data on hospital discharges due to all respiratory diseases (J00-J99 using ICD10 [30]) from 2000 to 2019 for seven regions (Brunca, Central Norte, Central Sur, Chorotega, Huetar Atlántica, Huetar Norte and Pacífico Central), were obtained from the Health Statistics Area of the CCSS. The choice of using 'discharge' instead of 'hospital entrance' is based on the accurate registration of diagnostics at the second level of the health system in Costa Rica. According to the official statistics from the CCSS, the average hospitalization duration in regional hospitals in 2019 due to respiratory diseases was 5.97 days [31]. For a visual representation of the geographic distribution of these regions, refer to Figure 1.



**Figure 1.** Health administrative division of Costa Rica.

#### 2.1.2. Input data: Climate, pollution, and lagged hospital discharge

We consider distinct climatic indices related to precipitation and temperature extremes, as well as the aerosol index and previous weekly hospital discharges for each Health Region as input variables to

predict weekly hospital discharges.

Daily precipitation estimates from 2000 to 2019, sourced from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPs) [32], were used to measure the land surface rainfall. Furthermore, the Geophysical Research Center (CIGEFI) of the University of Costa Rica supplied maximum and minimum daily temperatures across the country, employing the same estimation procedure as the Climate Hazards Center Infrared Temperature with Stations (CHIRTS) [33], with an extension of the time period until 2019. Both data sources offer high spatial resolution at 5km by 5km.

After reviewing the 27 core indices\* related to climate extremes, as recommended by the Climate and Ocean: Variability, Predictability, and Change (CLIVAR)<sup>†</sup> on Climate Change Detection [34], we adapted them for tropical countries and computed 22 climate change indices for each region related to cumulative rainfall estimates and temperature for each region  $i = 1, \dots, 7$  and week  $t = 1, \dots, T$  (See Table 1).

Specifically for the temperature indices, we eliminated those indices related to extreme winter conditions (numbers of days in which maximum and minimum temperatures are below 0C). Moreover, since we are dealing with weekly data, those indices related to the number of days greater than the 90th percentile and below the 10th percentiles, respectively, are zero for the majority of some periods of the year. To maintain the variability of these indices, we modified them to the 75th and 25th percentiles to retain the indices' variability throughout the year. We also incorporated weekly maximum, minimum, and average of daily amplitude, as well as the number of days in which the amplitude is greater than the 75th and 90th percentiles, to account for significant temperature changes within a day.

On the other hand, for precipitation indices, we eliminated the number of consecutive dry days and consecutive wet days since Costa Rica has distinct rainy and dry seasons, and these indices are basically zero in the dry season and 7 days a week in the rainy season. Furthermore, we incorporated the maximum, average, and minimum of maximum daily precipitation across the region, and modified the yearly day count with precipitation greater than the 95th and 99th percentiles to weekly day counts with precipitation greater than the 75th and 90th percentiles to account for the variability of the input variable.

In relation to pollution data, daily measurements of AOD from 2000 to 2019 were obtained from the MODIS Atmosphere L3 Daily Product (MOD08\_D3) [35] from the National Aeronautics and Space Administration (NASA). We obtained daily data points at 12 locations across the country using a spatial resolution of  $1 \times 1$  degree grid. The health regions do not naturally align with the  $1 \times 1$  grid, making it necessary to interpolate the variables to estimate a representative value for each region at a fixed point in time. Therefore, we calculated the weighted average based on the Euclidean distance between the 12 data points and the centroid of each region  $i$ . We believe this is a reasonable method for estimating a representative value for each region, making the best possible use of the geographic information available. Finally, weekly AOD values were derived by averaging the daily information.

The selection of the above environmental covariates is based on the results cited in the introduction, where previous studies have found a significant relationship between extreme weather data, pollutants, and hospital admissions. Due to the lack of admission information in the CCSS, we consider the dependent variable of hospital discharges to be a fairly close proxy for the former variable. All precipitation indices were logarithmically transformed due to their asymmetry where the main purpose of this trans-

\*See [https://etccdi.pacificclimate.org/list\\_27\\_indices.shtml](https://etccdi.pacificclimate.org/list_27_indices.shtml)

<sup>†</sup><https://www.clivar.org/>

**Table 1.** Climate change indices related to rainfall and temperature.

Variable	Description
1. Tmax_max	Maximum (across the region) of maximum weekly temperature.
2. Tmax_mean	Average (across the region) of maximum weekly temperature.
3. Tmax_min	Minimum (across the region) of maximum weekly temperature.
4. n_Tmax_Q3	Average (across the region) of the number of days in which the maximum temperature is higher than the percentile 75 of maximum temperature.
5. Tmin_min	Minimum (across the region) of minimum weekly temperature.
6. Tmin_mean	Average (across the region) of minimum weekly temperature.
7. Tmin_max	Maximum (across the region) of minimum weekly temperature.
8. n_Tmin_Q1	Average (across the region) of the number of days in which the minimum temperature is lower than the percentile 25 of minimum temperature.
9. amplitude_max_max	Maximum of the maximum daily amplitude of temperature (difference of daily maximum and minimum temperature).
10. amplitude_max_mean	Average of the maximum daily amplitude of temperature.
11. amplitude_max_min	Minimum of the maximum daily amplitude of temperature.
12. amplitude_min_max	Maximum of the minimum daily amplitude of temperature.
13. amplitude_min_mean	Average of the minimum daily amplitude of temperature.
14. amplitude_min_min	Minimum of the minimum daily amplitude of temperature).
15. n_amplitude_Q3	Average of the number of days in which the amplitude of temperature is higher than the percentile 75 of daily amplitude of temperature.
16. n_amplitude_P90	Average of the number of days in which the amplitude of temperature is greater than the percentile 90 of daily amplitude of temperature.
17. precip_max_max	Maximum (across the region) of the maximum daily precipitation.
18. precip_max_mean	Average (across the region) of the maximum daily precipitation.
19. precip_max_min	Minimum (across the region) of the maximum daily precipitation.
20. precip_mean_mean	Average (across the region) of the average daily precipitation.
21. n_precip_max_Q3	Average (across the region) of the number of days in which the maximum precipitation is higher than the percentile 75 of daily maximum precipitation.
22. n_precip_max_P90	Average (across the region) of the number of days in which the maximum precipitation is higher than the percentile 90 of daily maximum precipitation.

formation is to ensure stability in calculating the importance factors of these variables, rather than to improve the predictive power of the models. Furthermore, we computed the sample cross-correlation function between all covariates and hospital discharges. We observed that the highest correlation between all covariates and hospital discharges occurs in the same week, that is, at 0 lag, except for AOD, where the highest sample cross-correlation is detected with a lag of 10 weeks, namely AOD<sub>.10</sub>, thus we incorporate this lagged covariate instead of the same-week AOD in the model. Finally, hospital discharges from both the previous week and the week before are used in the model.

## 2.2. Statistical Learning Methods

We employed a supervised statistical learning approach to forecast hospital discharges based on the environmental information in each region. The prediction model for hospital discharges at time  $t$  ( $HD_t$ ) in a fixed region  $i$  is defined as follows:

$$HD_t \sim f(X_t, AOD_{t-10}, HD_{t-1}, HD_{t-2}),$$

where  $f(\cdot)$  represents the applied machine learning technique,  $X_t$  encompasses all 22 climate covariates at time  $t$  (see Table 1), and  $AOD_{t-10}$  denotes the AOD at time  $t - 10$ . The variables  $HD_{t-1}$  and  $HD_{t-2}$  represent the lags of one and two weeks respectively of the discharge series. These lags are considered as predictors because they allow to capture the non-stationarity of the original series.

For  $f(\cdot)$ , we apply four different machine learning methods:

**Random Forest (RF)** [36, 37] is a bootstrapped ensemble method that combines results of regression trees to improve the prediction.

**XGBoost (XGBOOST)** [38] is a gradient-boosting algorithm that employs bagging and trains multiple decision trees in order to produce forecasts.

**Facebook's Prophet Forecasting Model (PROPHET)** [39] is an additive-based model that fits non-linear trends with multiple seasonality plus holiday factors.

**Ensemble method (ENSEMBLE)** is a combination of the predictions of the three aforementioned models with equal weights.

We selected these methods because they are widely known in the literature and extensively used in the modeling and prediction of admission variables or hospital discharges; see [40–48] for examples of recent uses. We acknowledge the potential for employing hybrid or more complex methodologies (as exemplified in [49, 50]); however, for this article, we opt for straightforward and purely predictive techniques. We favor simplicity in implementation and interpretation, particularly considering the potential utilization of our methods by non-expert users of the CCSS. For example, any of the models used in this article are implemented in R libraries such as *tidymodels* [51], *modeltime* [52], as well as in *scikit-learn* [53] in Python.

To train the model, we split the data into training (2001-2018) and testing sets (2019). We selected the testing period with the intention of encompassing a complete cycle (annual) of observed behavior in regional discharges, where there was no presence of external shocks such as the COVID-19 pandemic that began in the year 2020. The training process involves a systematic approach, where the training set is divided into sliding windows. Each sliding window comprises a 2-year analysis set and a 1-year

assessment set, accounting for the temporal annual seasonality inherent in the data. Following this, we proceed with model tuning utilizing the grid search method employing a Latin hypercube approach with 20 combinations of hyperparameters, all based on the default tuning procedures provided by the `tune_grid` function in the `tidymodels` package. Subsequently, we refine this process through a grid search over a regular grid with a smaller number of combinations. In both instances, our objective remains consistent: to minimize the mean squared error (MSE). Once we identified the optimized hyperparameters based on both the aforementioned criterion and the tuning procedure (see Table 2), we predict hospital discharges and their prediction interval using conformal inference [54] in 2019 to compare with the testing set. The use of conformal inference ensures that prediction intervals are free

**Table 2.** Optimal hyperparameters found using grid search for each method and region. For the RF and XGBOOST methods, the number of trees was held constant at 1000.

Method	Hyperparameter	Brunca	Central Norte	Central Sur	Chorotega	Huetar lantica	At-Huetar Norte	Pacifico Central
RF	mtry	25	20	17	15	12	20	22
	min_n	40	22	30	30	40	40	37
	mtry	12	14	14	12	12	10	10
	min_n	40	8	8	40	40	35	35
XGBOOST	tree_depth	13	11	11	13	13	12	12
	learn_rate	0.007	0.078	0.078	0.007	0.007	0.003	0.003
	loss_reduction	0.001	0.018	0.018	0.001	0.001	0.000	0.000
	sample_size	0.898	0.828	0.828	0.898	0.898	0.589	0.589
PROPHET	prior_scale_changepoints	1.000	1.778	1.000	1.000	3.162	1.778	1.000
	prior_scale_seasonality	3.162	1.778	10.000	1.000	1.778	5.623	1.000

from any assumptions about the distribution of our outcome, and it also allows for achieving interesting coverage properties even for small samples (see [54]).

Finally, we compute three metrics to compare the predictive performance of the applied methods for a fixed region. The first one, MSE [55] is defined as follows:

$$\text{MSE} = \frac{1}{m} \sum_{t=1}^m (HD_t - \widehat{HD}_t)^2,$$

where  $m$  is the number of weeks in the testing period,  $HD_t$  is the observed hospital discharge at week  $t$ , and  $\widehat{HD}_t$  is the estimated hospital discharge according to each model at time  $t$ . The Mean Absolute Error (MAE) [55] is computed as follows:

$$\text{MAE} = \frac{1}{m} \sum_{t=1}^m |HD_t - \widehat{HD}_t|,$$

and finally, the Interval Score (IS) [56] is defined as

$$\text{IS}_\alpha = \frac{1}{m} \sum_{t=1}^m \left[ (U_t - L_t) + \frac{2}{\alpha} (L_t - HD_t) \cdot 1_{HD_t < L_t} + \frac{2}{\alpha} (HD_t - U_t) \cdot 1_{HD_t > U_t} \right],$$



where  $U_t$  and  $L_t$  are the upper and lower limits of the prediction interval of  $(1 - \alpha)\%$ , respectively.

MSE and MAE evaluate the point forecasts with the observed hospital discharges. On the other hand,  $IS_\alpha$  assesses a  $(1 - \alpha)\%$ -prediction interval by comparing its upper and lower limits against the observed values. This metric is more comprehensive than MSE for evaluating the models' predictive capacity when uncertainty is summarized through a predictive interval [56]. While it is not the only measure that achieves this, it is one of the simplest to compute and easiest to interpret. Finally, to quantify the degree of overfitting in each of the models, we compare the MSE and the IS obtained in the training set versus those obtained in the testing period through the relative MSE and relative IS as follows:

$$MSE_{rel} = \frac{MSE_{test}}{MSE_{train}}, \quad \text{and} \quad IS_{rel} = \frac{IS_{test}}{IS_{train}}.$$

We also calculate the contribution of each covariate in predicting the dependent variable. In particular, we are interested in quantifying the global contribution through the aggregation of local contributions based on Shapley values. In other words, we calculate the importance of each covariate through its mean absolute Shapley value using the `vip_shap` function from the `vip` R package [57]. This allows us to compare which covariates have a greater impact on the prediction process for each specific region.

All calculations were performed in the statistical software R [58] and the following packages `tidymodels` [51], `modeltime` [52], `ranger` [59], `xgboost` [60], `modeltime.ensemble` [61], and `prophet` [62]. The block diagram depicted in Figure 2 provides a concise overview of the methodological steps outlined in this article. The code used in this application, as well as the computational details of this article, can be found at [https://github.com/shuwei325/pollution\\_hosp](https://github.com/shuwei325/pollution_hosp).

### 3. Results

In Table 3, we provide a comparison of the four employed methods using the metrics detailed in Section 2. It is crucial to highlight that our objective is to assess the predictive performance of the four models during the testing period. This assessment involves comparing the observed outcomes with the projected ones, utilizing the set of covariates observed within the same period.

Due to differences in the behavior of the series in each region, it is confirmed that there is no method that is optimal for all cases. Likewise, the ensemble method is not optimal according to the calculated metrics. In order to choose the proposed models, we sought to have the smallest IS in the testing period, and we also ensured an  $IS_{rel}$  ratio higher than 80%, with the goal of selecting alternatives with potential overfitting. The other metrics were used to verify that the selected methods are also competitive compared to alternatives. In general terms, all selected methods successfully predict the testing period across the regions and capture both high and low-frequency characteristics of the original discharge series in a descriptive manner, as seen in Figure 3.

Figure 3 illustrates that the 95% predictive intervals successfully capture the observed series in most regions, except for Huetar Atlantica and Norte regions, and partially in the Pacifico Central region, which showed greater difficulty in their forecasts.

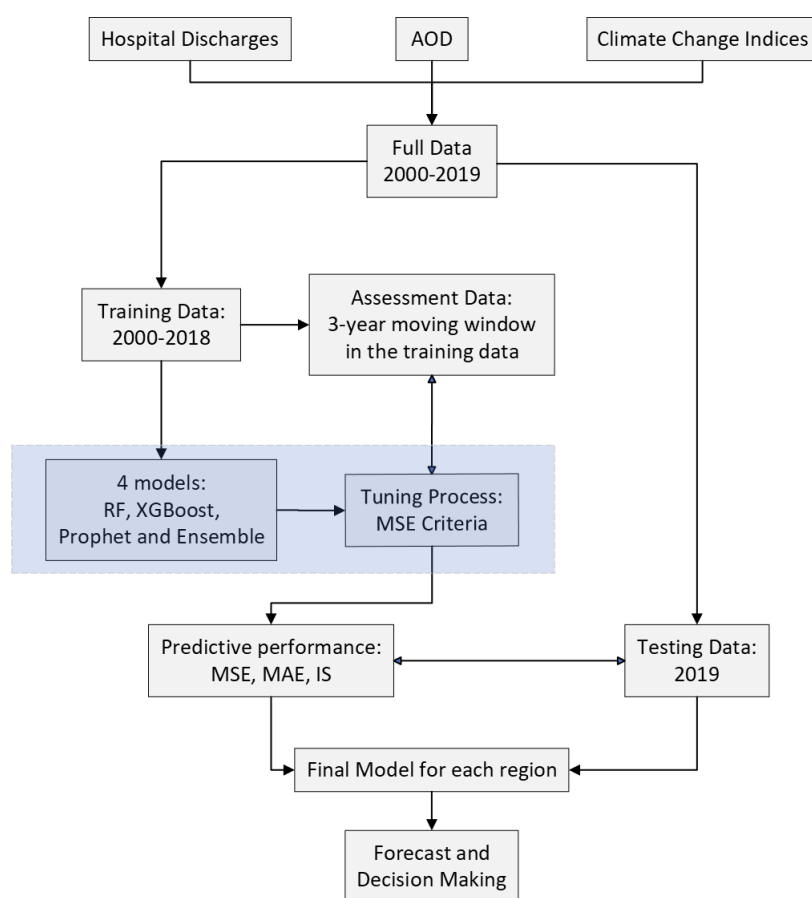
For these last regions, the PROPHET method constitutes an interesting alternative, since this method also allows incorporating a time index that can explain those non-stationary components not being

**Table 3.** Metric comparison by region and method (all covariates). The best method for each region is shown in bold.

Region	Model	MSE	MAE	IS <sub>0.05</sub>	MSE (train)	MAE (train)	IS <sub>0.05</sub> (train)	MSE <sub>rel</sub>	IS <sub>rel</sub>
Brunca	RF	188.55	10.18	71.79	73.10	7.00	60.09	68.78	83.69
	XGBOOST	253.34	11.74	79.12	74.65	6.94	69.71	59.13	88.10
	<b>PROPHET</b>	237.86	13.29	63.68	55.41	6.08	59.29	45.77	93.10
	ENSEMBLE	205.36	11.29	70.83	58.74	6.25	34.67	55.32	48.94
Central Norte	<b>RF</b>	222.95	12.01	69.65	97.77	7.80	56.27	64.89	80.79
	XGBOOST	281.86	13.70	75.23	0.00	0.03	71.80	0.21	95.44
	PROPHET	2947.93	48.38	210.24	99.48	7.80	191.61	16.12	91.14
	ENSEMBLE	540.34	18.11	112.19	38.07	4.89	29.89	27.00	26.64
Central Sur	<b>RF</b>	407.37	16.63	85.27	166.16	10.01	76.44	60.21	89.64
	XGBOOST	429.42	17.28	86.63	0.00	0.03	79.24	0.17	91.47
	PROPHET	660.35	21.00	110.24	131.06	9.23	89.29	43.97	81.00
	ENSEMBLE	408.17	17.28	89.01	57.41	6.14	35.92	35.52	40.36
Chorotega	RF	90.16	7.34	44.12	24.29	3.87	41.31	52.71	93.64
	<b>XGBOOST</b>	84.14	7.11	42.12	32.86	4.49	42.01	63.18	99.75
	PROPHET	96.64	8.09	47.00	29.61	4.22	35.05	52.15	74.58
	ENSEMBLE	81.11	7.18	42.19	25.60	3.94	24.81	54.92	58.81
Huetar Atlántica	<b>RF</b>	142.02	8.77	58.77	32.12	4.36	57.38	49.73	97.63
	XGBOOST	156.60	9.13	62.70	34.71	4.50	58.33	49.28	93.03
	PROPHET	237.39	12.56	66.79	22.09	3.72	63.98	29.61	95.79
	ENSEMBLE	160.25	9.61	61.59	25.67	3.90	24.55	40.55	39.87
Huetar Norte	RF	158.97	10.13	55.13	27.52	4.12	49.76	40.70	90.26
	XGBOOST	180.60	10.85	60.19	48.21	5.42	54.86	49.94	91.13
	<b>PROPHET</b>	163.56	10.15	58.24	34.34	4.72	55.85	46.52	95.89
	ENSEMBLE	161.28	10.20	56.61	33.23	4.51	27.50	44.21	48.58
Pacífico Central	RF	53.64	5.89	34.90	35.48	4.73	31.01	80.21	88.87
	XGBOOST	47.99	5.46	32.65	56.81	5.83	35.97	106.65	110.16
	<b>PROPHET</b>	85.38	8.18	32.11	34.52	4.67	32.05	57.06	99.82
	ENSEMBLE	55.21	6.20	31.03	37.39	4.82	28.68	77.67	92.42

**Table 4.** Relative Importance factors for the best methods (in percentage) by means of the Shapley values. The *date* covariate corresponds to the time index, needed for the PROPHET algorithm.

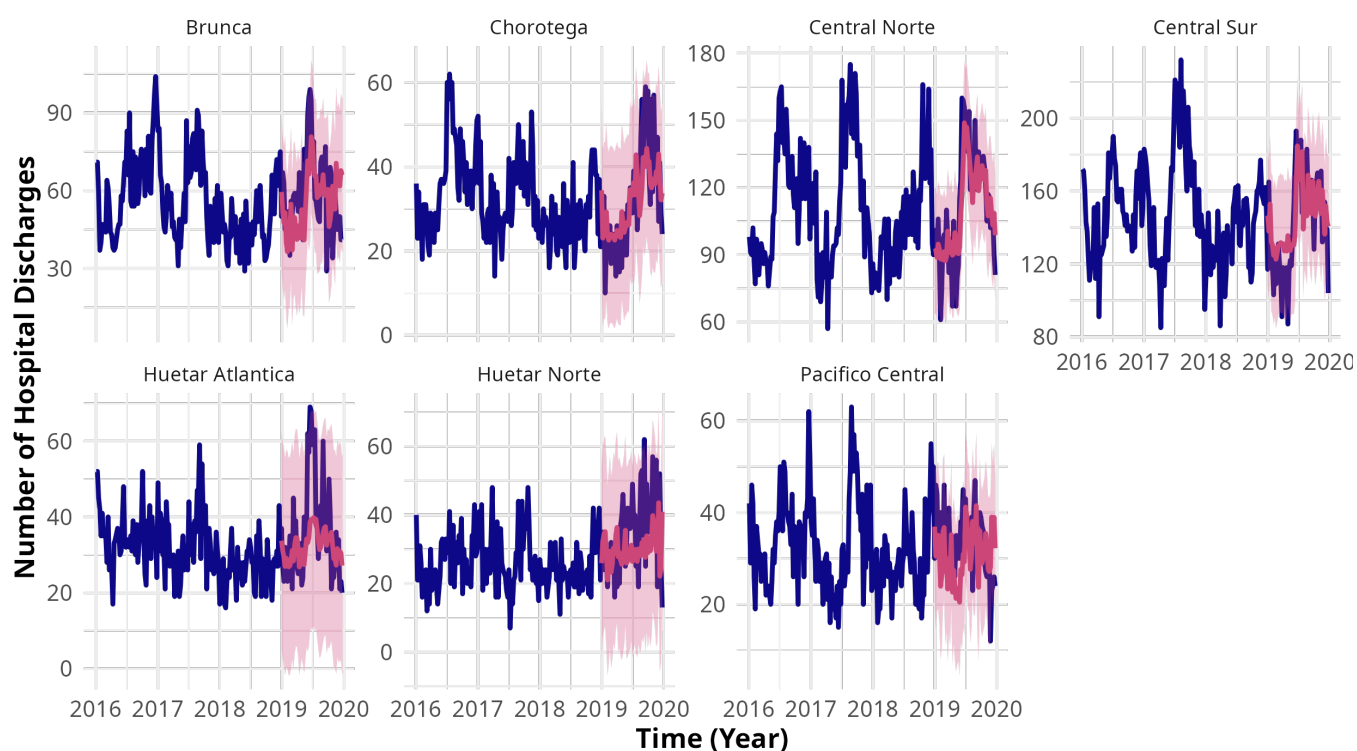
Covariate	Brunca	Central Norte	Central Sur	Chorotega	Huetar Atlantica	Huetar Norte	Pacifico Central
date	21.1					18.4	23.2
n_amplitude_Q3	11.0		2.0		4.4	7.5	7.6
n_amplitude_P90	9.9				6.7		
amplitude_max_max	9.9					6.1	7.4
egreso_2	9.3	21.6	18.9	36.1	31.0		8.8
precip_max_min	9.1			3.1	3.1		6.6
precip_mean_mean	9.0	3.0	3.0			7.7	
Tmax_max	7.9					13.1	
n_precip_max_Q3	6.6	2.3		3.0	3.6		
Tmax_mean	6.1		1.5		4.0	10.9	
egreso_1		46.5	50.3	16.8	24.5		
AOD_10		15.6	16.5	17.1	14.0		
AOD		4.3		4.5			
precip_max_max		2.2	1.6	3.8			
precip_max_mean		1.6	2.3			6.1	
Tmin_min		1.5		2.8		7.3	7.5
Tmin_mean		1.4				13.8	13.0
amplitude_min_min			2.1				
amplitude_min_mean			1.8				5.7
amplitude_max_min				9.3		9.2	
n_Tmin_Q1				3.5			
amplitude_max_mean					5.8		10.4
amplitude_min_max					2.9		
Tmin_max							10.0



**Figure 2.** Block diagram for developing the forecasting models for hospital discharges at regional level using environmental factors. The blue shaded blocks are specified in greater detail in Figure 4 in the Supplementary section.

explained by the covariates. Its flexibility in modeling trends, seasonal components, and trend changes makes it particularly well-suited for these complex-behaved regions, as it was purposefully designed to accommodate such complexities (see [39]). This method has been used in situations where the complexity of the patterns observed in the series makes the use of classical time series models not as effective (see [63,64]). For example, the  $IS_{rel}$  reflects the model's ability to provide accurate predictions in the testing set, even in the Huetar Atlantica region where it is slightly outperformed by the RF. This is largely due to its balanced adaptability to the series. Also, in Table 4, we show the Shapley values, all in percentage terms to facilitate interpretation.

It is worth noting that the PROPHET method performs better in three regions (Brunca, Huetar Norte, and Pacifico Central), where the date is the most important feature, while lagged HD and AOD are not. It is worth noting that for other four regions where the RF and XGBOOST methods are the best, the covariates with the greatest contribution to the prediction of hospital discharges tend to be the lagged discharge variables (orders 1 and 2) along with the lagged AOD variable. This indicates that for these regions, the climate and pollution features, as well as lagged HD, are important inputs to predict HD, while for those that PROPHET are best models, temporal information such as nonstationarity and seasonality provides better information to forecast the HD.



**Figure 3.** Forecast comparison of hospital discharges, by region. Dark blue line: observed series, pink line: predicted series and pink shaded area: 95% prediction interval.

On the other hand, depending on the region, there are environmental variables that contribute to predictions to varying degrees once the contribution of lagged variables has been taken into account. For example, it is noteworthy that for the Chorotega and Huetar Norte regions, the amplitude between maximum and minimum temperature is important, as well as in the latter region, the regional minimum of minimum temperature and the regional maximums of maximum temperature.

In general terms, a considerable number of the covariates we used show significant contribution in the prediction models. The difficulty in determining an optimal fitting method may be due to the high variability in weekly discharge data, and the existence of complex seasonality components among different health regions, which may largely stem from the combination of epidemiological cycles of respiratory diseases. These situations of high variability and seasonality in hospital admission data (which we are approximating through discharges) have already been evidenced in previous studies [23, 25, 65, 66]. It is also worth mentioning that as we are choosing models where  $IS_{rel}$  is less than 80%, we then directly minimize the occurrence of overfitting, which as far as we know, was not directly addressed in these same articles previously cited.

#### 4. Conclusion and recommendations

To summarize, we used climatic and pollution data to model and predict weekly hospital discharges due to respiratory conditions, a leading contributor to the global healthcare burden in terms of costs, at regional hospital levels in Costa Rica. Four statistical learning approaches: RF, XGBOOST,

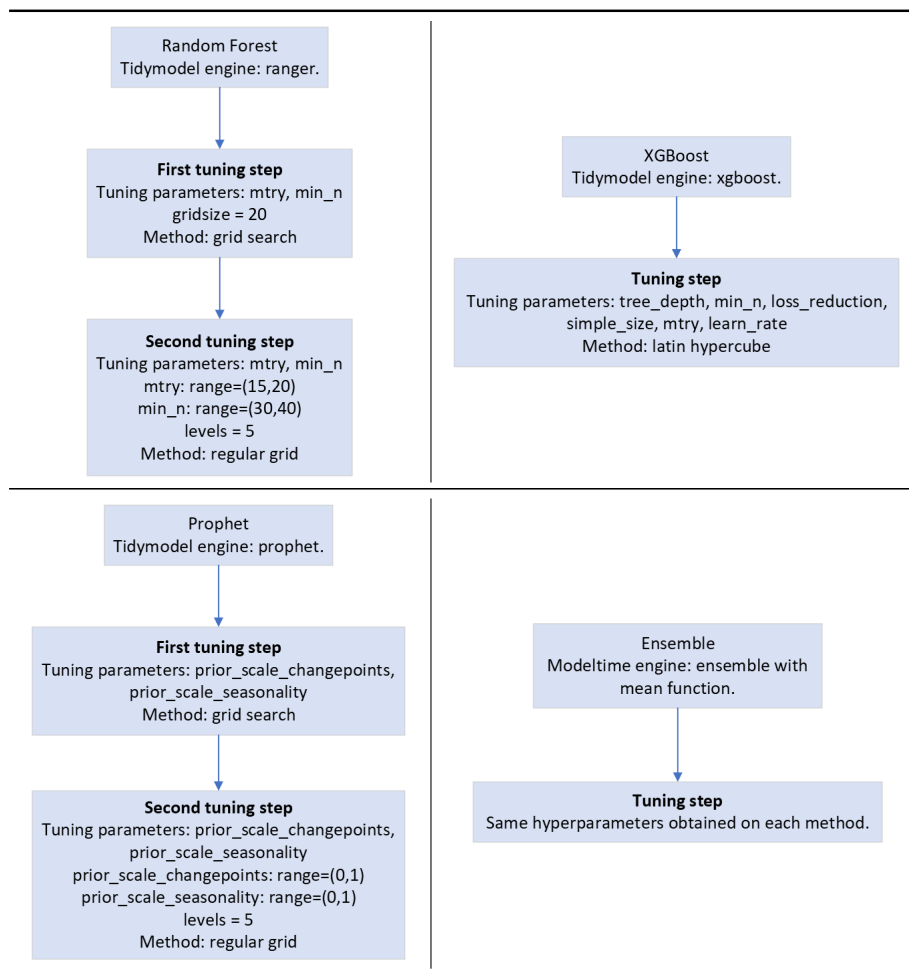
PROPHET, and the ENSEMBLE method, were applied to each of the seven regions. According to the predictive metrics for each region, reliable forecasts are obtained, and specific conclusions can be drawn. In Brunca, Huetar Norte, and the Pacífico Central region, PROPHET obtained better forecasts, and the temporal input was detected as the most important predictor. For other regions (Central Norte, Central Sur, Chorotega, and Huetar Atlántico), where other methods obtained better results, temporal information was not included as a predictor, and other climatic indices and lagged hospitalization and AOD were found to be the most important factors.

These conclusions are obtained from comparing their predictive performance in both the training set (2001-2018) and the testing set (2019). Initially, the training procedure utilized sliding windows of a 2-year training period and a 1-year assessment period, allowing for the optimization of the model's hyperparameters while considering yearly seasonality. Subsequently, their predictive capacity was evaluated in the testing set using metrics such as MSE, MAE, and IS, as well as  $MSE_{rel}$  and  $IS_{rel}$  to assess overfitting. We conclude that the optimal model for each region depends on specific environmental and pollution variables, which do not align for all cases. However, at least for most regions, the lagged variables of hospital discharges and the lagged AOD are of the highest relative importance. In other cases, the non-stationary component is strong enough for a time-indexed model to be preferable. Despite all of the above, the selected model for each region demonstrates its ability to generate reliable forecasts. Although no previous predictive studies were found in Costa Rica, we can conclude that the MSE and MAE metrics of our study are lower compared to a similar study of emergency admissions in the city of Santiago, Chile (see Table 3 in [25]), which reports MSE and MAE for their forecasts.

As limitations, it is important to recognize that other significant factors, such as the socio-economic conditions of each region, could serve as potential predictors for hospitalizations due to respiratory conditions. However, many socio-economic factors, including GDP, population density, and healthcare facilities, among others, are not available on a weekly basis or remain constant over time. We argue that, given the established association in the literature between social factors and AOD, we indirectly incorporate this information into the model, recognizing it as a proxy for the anthropogenic effect on climate.

Due to the organizational structure of the CCSS, where the decentralization of services through regional hospitals and clinics by health region is encouraged, decision-making at the hospital level in Costa Rica is crucial. This study enables health authorities to anticipate and visualize the hospital requirements for serving the population with respiratory problems in Costa Rican Health Regions. Additionally, it empowers authorities to administer and allocate limited human and economic resources effectively and efficiently, fostering a balanced national health economy.

As future work, two directions may be considered. First, since our focus is on forecasting respiratory hospitalization using climatic and pollution data, the implicit consideration of nonlinear associations in these models raises questions about the non-linear relationships and long/short-term associations among them, which need further analysis in different microclimates in Costa Rica. Second, while climatic and pollution data are key variables for forecasting respiratory hospitalization, they are not the only factors influencing respiratory health. Socioeconomic data, albeit limited, could be directly or indirectly included to enhance predictions.



**Figure 4.** Block diagrams for the model definition and hyperparameter tuning stages for each used model.

## Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

Luis A. Barboza and Shu Wei Chou Chen received funding through V.I. C3195, V.I. C4196, and V.I. C3175. We also extend our gratitude to the Health Statistics Area of the CCSS and CIGEFI, UCR, for providing the data used in this manuscript.

## Conflict of interest

The authors declare there is no conflict of interest.

## References

1. Forum of International Respiratory Societies, *Respiratory diseases in the world. Realities of today – opportunities for tomorrow*, Technical Report 9, 2014.
2. Y. Chen, D. Kong, J. Fu, Y. Zhang, Y. Zhao, Y. Liu, et al., Associations between ambient temperature and adult asthma hospitalizations in Beijing, China: A time-stratified case-crossover study, *Respir. Res.*, **23** (2022), 38. <https://doi.org/10.1186/s12931-022-01960-8>
3. S. Lin, M. Luo, R. J. Walker, X. Liu, S.-A. Hwang, R. Chinery, Extreme high temperatures and hospital admissions for respiratory and cardiovascular diseases, *Epidemiology*, **20** (2009), 738–746. <https://doi.org/10.1097/EDE.0b013e3181ad5522>
4. H. Pedder, T. Kapwata, G. Howard, R. N. Naidoo, Z. Kunene, R. W. Morris, et al., Lagged Association between Climate Variables and Hospital Admissions for Pneumonia in South Africa, *Int. J. Env. Res. Pub. He.*, **18** (2021). <https://doi.org/10.3390/ijerph18126191>
5. Q. A. Tran, V. T. H. Le, V. T. Ngo, T. H. Le, D. T. Phung, J. D. Berman, et al., The Association Between Ambient Temperatures and Hospital Admissions Due to Respiratory Diseases in the Capital City of Vietnam, *Front. Pub. Hea.*, **10** (2022). <https://doi.org/10.3389/fpubh.2022.903623>
6. H. Tsangari, A. K. Paschalidou, A. P. Kassomenos, S. Vardoulakis, C. Heaviside, K. E. Georgiou, et al., Extreme weather and air pollution effects on cardiovascular and respiratory hospital admissions in Cyprus, *Sci. Total Environ.*, **542** (2016), 247–253. <https://doi.org/10.1016/j.scitotenv.2015.10.106>
7. Y. Guo, A. Gasparrini, B. G. Armstrong, B. Tawatsupa, A. Tobias, E. Lavigne, et al., Temperature Variability and Mortality: A Multi-Country Study, *Environ. Health Persp.*, **124** (2016), 1554–1559. <https://doi.org/10.1289/ehp149>
8. H. Kan, R. Chen, S. Tong, Ambient air pollution, climate change, and population health in China, *Environ. Int.*, **42** (2012), 10–19. <https://doi.org/10.1016/j.envint.2011.03.003>
9. C. Liu, R. Chen, F. Sera, A. M. Vicedo-Cabrera, Y. Guo, S. Tong, et al., Ambient Particulate Air Pollution and Daily Mortality in 652 Cities, *New Engl. J. Med.*, **381** (2019), 705–715. <https://doi.org/10.1056/NEJMoa1817364>
10. J. Lou, Y. Wu, P. Liu, S. H. Kota, L. Huang, Health Effects of Climate Change Through Temperature and Air Pollution, *Curr. Poll. Rep.*, **5** (2019), 144–158. <https://doi.org/10.1007/s40726-019-00112-9>
11. S. G. E. K. Miraglia, P. H. N. Saldiva, G. M. Böhm, An Evaluation of Air Pollution Health Impacts and Costs in São Paulo, Brazil, *Environ. Manage.*, **35** (2005), 667–676. <https://doi.org/10.1007/s00267-004-0042-9>
12. F. Makrufardi, A. Manullang, D. Rusmawatingtyas, K. F. Chung, S.-C. Lin, H.-C. Chuang, Extreme weather and asthma: A systematic review and meta-analysis, *Eur. Respir. Rev.*, **32**, <https://doi.org/10.1183/16000617.0019-2023>
13. G. B. Hamra, N. Guha, A. Cohen, F. Laden, O. Raaschou-Nielsen, et al., Outdoor particulate matter exposure and lung cancer: A systematic review and meta-analysis, *Environ. Health Persp.*, **122** (2014), 906–911. <https://doi.org/10.1289/ehp/1408092>
14. I. Colbeck, M. Lazaridis, Aerosols and environmental pollution, *Naturwissenschaften*, **97** (2010), 117–131. <https://doi.org/10.1007/s00114-009-0594-x>



15. S. Provençal, P. Kishcha, A. M. d. Silva, E. Elhacham, P. Alpert, AOD distributions and trends of major aerosol species over a selection of the world's most populated cities based on the 1st Version of NASA's MERRA Aerosol Reanalysis, *Urban climate*, **20** (2017), 168, <https://doi.org/10.1016/j.uclim.2017.04.001>
16. W. Zhang, Q. He, H. Wang, K. Cao and S. He, Factor analysis for aerosol optical depth and its prediction from the perspective of land-use change, *Ecol. Indic.*, **93** (2018), 458–469, <https://doi.org/10.1016/j.ecolind.2018.05.026>
17. L. Li, Y. Wang, What drives the aerosol distribution in Guangdong - the most developed province in Southern China?, *Sci. Rep-UK*, **4** (2014), 5972, <https://doi.org/10.1038/srep05972>
18. Organización Panamericana de la Salud, Perfil del sistema y servicios de salud de Costa Rica con base al marco de monitoreo de la Estrategia Regional de Salud Universal, Place: San José.
19. M. Baird, (In)equity and primary health care: The case of Costa Rica and Panama, *Int. J. Soc. Det. He. Ser.*, **53** (2023), 122–129. <https://doi.org/10.1177/27551938231152991>
20. H. Jia, J. Xu, L. Ning, T. Feng, P. Cao, S. Gao, et al., Ambient air pollution, temperature and hospital admissions due to respiratory diseases in a cold, industrial city, *J. Glob. Health*, **12** (2022), 4085. <https://doi.org/10.7189/jogh.12.04085>
21. J. Peng, C. Chen, M. Zhou, X. Xie, Y. Zhou, C.-H. Luo, Peak outpatient and emergency department visit forecasting for patients with chronic respiratory diseases using machine learning methods: Retrospective cohort study, *JMIR Medical Inform.*, **8** (2020), e13075. <https://doi.org/10.2196/13075>
22. Y. de Souza Tadano, H. V. Siqueira, T. A. Alves, Unorganized machines to predict hospital admissions for respiratory diseases, in *2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, (2016), 1–6. <https://doi.org/10.1109/LA-CCI.2016.7885699>
23. I. N. Soyiri, D. D. Reidpath, C. Sarran, Forecasting asthma-related hospital admissions in london using negative binomial models, *Chron. Resp. Dis.*, **10** (2013), 85–94. <https://doi.org/10.1177/1479972313482847>
24. P. Kassomenos, C. Papaloukas, M. Petrakis, S. Karakitsios, Assessment and prediction of short term hospital admissions: the case of athens, greece, *Atmos. Environ.*, **42** (2008), 7078–7086. <https://doi.org/10.1016/j.atmosenv.2008.06.011>
25. M. Becerra, A. Jerez, B. Aballay, H. O. Garcés, A. Fuentes, Forecasting emergency admissions due to respiratory diseases in high variability scenarios using time series: A case study in chile, *Sci. Total Environ.*, **706** (2020), 134978, <https://doi.org/10.1016/j.scitotenv.2019.134978>
26. E. K. Poon, V. Kitsios, D. Pilcher, R. Bellomo, J. Raman, Projecting future climate impact on national australian respiratory-related intensive care unit demand, *Heart Lung Circ.*, **32** (2023), 95–104. <https://doi.org/10.1016/j.hlc.2022.12.001>
27. K. Ravindra, S. S. Bahadur, V. Katoch, S. Bhardwaj, M. Kaur-Sidhu, M. Gupta, et al., Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections, *Sci. Total Environ.*, **858** (2023), 159509. <https://doi.org/10.1016/j.scitotenv.2022.159509>

28. Y. Odeyemi, A. Lal, E. Barreto, A. LeMahieu, H. Yadav, O. Gajic, et al., Early machine learning prediction of hospitalized patients at low risk of respiratory deterioration or mortality in community-acquired pneumonia: Derivation and validation of a multivariable model, *Biomol. Biomed.*, **24** (2024), 337–345. <https://doi.org/10.17305/bb.2023.9754>
29. J. Yang, X. Xu, X. Ma, Z. Wang, Q. You, W. Shan, et al., Application of machine learning to predict hospital visits for respiratory diseases using meteorological and air pollution factors in linyi, china, *Environ. Sci. Pollut. R.*, **30** (2023), 88431–88443, <https://doi.org/10.1007/s11356-023-28682-8>
30. ICD-10-CM, J00-J99: Diseases of the respiratory system, Available online at <https://www.icd10data.com/ICD10CM/Codes/J00-J99> (accessed on September 10, 2022).
31. CCSS, Estadística en salud, Available online at <https://www.ccss.sa.cr/estadisticas-salud> (accessed on October 16, 2023).
32. C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, et al., The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, *Sci. Data*, **2** (2015), 150066–150066. <https://doi.org/10.1038/sdata.2015.66>
33. C. Funk, P. Peterson, S. Peterson, S. Shukla, F. Davenport, J. Michaelsen, et al., A high-resolution 1983–2016 tmax climate data record based on infrared temperatures and stations by the climate hazard center, *J. Climate*, **32** (2019), 5639–5658. <https://doi.org/10.1175/JCLI-D-18-0698.1>
34. T. R. Karl, N. Nicholls, A. Ghazi, *CLIVAR/GCOS/WMO Workshop on Indices and Indicators for Climate Extremes Workshop Summary*, 3–7, Springer Netherlands, Dordrecht, (1999). [https://doi.org/10.1007/978-94-015-9265-9\\_2](https://doi.org/10.1007/978-94-015-9265-9_2)
35. MODIS Atmosphere Science Team, Modis/terra aerosol cloud water vapor ozone daily 13 global 1deg cmg, 2017.
36. L. Breiman, Random forests, *Mach. learning.*, **45** (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
37. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition, Springer Series in Statistics, Springer (2009). <https://doi.org/10.1007/978-0-387-84858-7>
38. T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, Association for Computing Machinery, New York, NY, USA, (2016) 785–794. <https://doi.org/10.1145/2939672.2939785>
39. S. J. Taylor, B. Letham, Forecasting at scale, *Am. Stat.*, **72** (2018), 37–45. <https://doi.org/10.1080/00031305.2017.1380080>
40. A. Gramaje, F. Thabtah, N. Abdelhamid, S. K. Ray, Patient discharge classification using machine learning techniques, *Ann. Data Sci.*, **8** (2021), 755–767. <https://doi.org/10.1007/s40745-019-00223-6>
41. J. H. Chung, D. Cannon, M. Gulbrandsen, D. Yalamanchili, W. P. Phipatanakul, J. Liu, et al., Random forest identifies predictors of discharge destination following total shoulder arthroplasty, *JSES Int.*, **8** (2024), 317–321. <https://doi.org/10.1016/j.jseint.2023.04.003>

42. R. Chen, S. Zhang, J. Li, D. Guo, W. Zhang, X. Wang, et al., A study on predicting the length of hospital stay for chinese patients with ischemic stroke based on the XGBoost algorithm, *BMC Med. Inform. Decis.*, **23** (2023). <https://doi.org/10.1186/s12911-023-02140-4>
43. Y.-T. Lo, J. C. Liao, M.-H. Chen, C.-M. Chang, C.-T. Li, Predictive modeling for 14-day unplanned hospital readmission risk by using machine learning algorithms, *BMC Med. Inform. Decis.*, **21** (2021). <https://doi.org/10.1186/s12911-021-01639-y>
44. T. H. McCoy Jr, A. M. Pellegrini, R. H. Perlis, Assessment of time-series machine learning methods for forecasting hospital discharge volume, *JAMA Netw. Open*, **1** (2018), e184087. <https://doi.org/10.1001/jamanetworkopen.2018.4087>
45. H. Álvarez-Chaves, P. Muñoz, M. D. R-Moreno, Machine learning methods for predicting the admissions and hospitalisations in the emergency department of a civil and military hospital, *J. Intell. Inf. Syst.*, **61** (2023), 881–900. <https://doi.org/10.1007/s10844-023-00790-4>
46. J. Wolff, A. Klimke, M. Marscholke, T. Kacprowski, Forecasting admissions in psychiatric hospitals before and during covid-19: A retrospective study with routine data, *Sci. Rep.*, **12** (2022), 1–9. <https://doi.org/10.1038/s41598-022-20190-y>
47. S. A. Suha, T. F. Sanam, A machine learning approach for predicting patient’s length of hospital stay with random forest regression, in *2022 IEEE Region 10 Symposium (TENSYMP)*, (2022) 1–6. <https://doi.org/10.1109/TENSYMP54529.2022.9864447>
48. H. Yun, J. Choi, J. H. Park, Prediction of critical care outcome for adult patients presenting to emergency department using initial triage information: An XGboost algorithm analysis, *JMIR Med. Inform.*, **9** (2021), e30770. <https://doi.org/10.2196/30770>
49. L. C. P. Velasco, N. R. Estoperez, R. J. R. Jayson, C. J. T. Sabijon, V. C. Sayles, Day-ahead base, intermediate, and peak load forecasting using k-means and artificial neural networks, *IJACSA*, **9** (2018). <https://doi.org/10.14569/IJACSA.2018.090210>
50. L. C. P. Velasco, D. L. L. Polestico, G. P. O. Macasieb, M. B. V. Reyes, F. B. Vasquez Jr, A hybrid model of autoregressive integrated moving average and artificial neural network for load forecasting, *IJACSA*, **10** (2019), 14–16. <https://doi.org/10.14569/IJACSA.2019.0101103>
51. M. Kuhn, H. Wickham, *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles.*, (2020).
52. M. Dancho, *modeltime: The Tidymodels Extension for Time Series Modeling*, (2023). R package version 1.2.8.
53. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, **12** (2011), 2825–2830.
54. J. Lei, M. G’Sell, A. Rinaldo, R. J. Tibshirani, L. Wasserman, Distribution-free predictive inference for regression, *J. Am Stat. Assoc.*, **113** (2018), 1094–1111. <https://doi.org/10.1080/01621459.2017.1307116>
55. R. Hyndman, G. Athanasopoulos, *Forecasting: Principles and practice, 3rd edition.*, 3rd edition, OTexts, Melbourne, Australia, (2021).
56. T. Gneiting, A. E. Raftery, Strictly Proper Scoring Rules, Prediction, and Estimation, *J. Am Stat. Assoc.*, **102** (2007), 359–378. <https://doi.org/10.1198/016214506000001437>

57. B. M. Greenwell, B. C. Boehmke, Variable importance plots—an introduction to the vip package, *R J.*, **12** (2020), 343–366. <https://doi.org/10.32614/RJ-2020-013>
58. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, (2023).
59. M. N. Wright, A. Ziegler, ranger: A fast implementation of random forests for high dimensional data in C++ and R, *J. Stat Softw.*, **77** (2017), 1–17. <https://doi.org/10.18637/jss.v077.i01>
60. T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, et al., *xgboost: Extreme Gradient Boosting*, (2023). R package version 1.7.5.1.
61. M. Dancho, *modeltime.ensemble: Ensemble Algorithms for Time Series Forecasting with Modeltime*, (2023). R package version 1.0.3.
62. S. Taylor, B. Letham, *prophet: Automatic Forecasting Procedure*, (2021). R package version 1.0.
63. S. Chaturvedi, E. Rajasekar, S. Natarajan, N. McCullen, A comparative assessment of SARIMA, LSTM RNN and fb prophet models to forecast total and peak monthly energy demand for india, *Energ. Policy*, **168** (2022), 113097. <https://doi.org/10.1016/j.enpol.2022.113097>
64. M. Elseidi, A hybrid facebook Prophet-ARIMA framework for forecasting high-frequency temperature data, *Model. Earth Syst. Environ.*, **10** (2024), 1855–1867. <https://doi.org/10.1007/s40808-023-01874-4>
65. D. Royé, J. J. Taboada, A. Martí, M. N. Lorenzo, Winter circulation weather types and hospital admissions for respiratory diseases in galicia, Spain, *Int. J. Biometeorol.*, **60** (2016), 507–520. <https://doi.org/10.1007/s00484-015-1047-1>
66. R. Navares, J. Díaz, C. Linares, J. L. Aznarte, Comparing ARIMA and computational intelligence methods to forecast daily hospital admissions due to circulatory and respiratory causes in madrid, *Stoch. Env. Res. Risk A.*, **32** (2018), 2849–2859. <https://doi.org/10.1007/s00477-018-1519-z>



AIMS Press

© 2024 the author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)