Healthcare Analytics 3 (2023) 100169

Contents lists available at ScienceDirect

Healthcare Analytics

journal homepage: www.elsevier.com/locate/health

Near real-time prediction of urgent care hospital performance metrics using scalable random forest algorithm: A multi-site development

Theresia A. Budiman [a], Charlotte R. James [a,b], Nicholas C. Howlett [a], Richard M. Wood [a,c,*]

[a] Modelling and Analytics, UK National Health Service (BNSSG ICB), United Kingdom

[b] NIHR Bristol Biomedical Research Centre (BRC), University Hospitals Bristol and Weston NHS Foundation Trust and University of Bristol, United Kingdom

[c] Centre for Healthcare Innovation and Improvement (CHI^2), School of Management, University of Bath, United Kingdom

ARTICLE INFO

Keywords:
Machine learning
Random forest
Forecasting
Urgent care
Predictive analytics
Time series

ABSTRACT

While previous studies have shown the potential value of predictive modelling for emergency care, few models have been implemented for producing near real-time predictions across various demand, utilisation and performance metrics. In this study, 33 independent Random Forest (RF) algorithms were developed to forecast 11 urgent care metrics over a 24-hour period across three hospital sites in an Integrated Care System (ICS) in South West England. Metrics included: ambulance handover delay; emergency department occupancy; and patients awaiting admission. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Symmetric Mean Absolute Percentage Error (SMAPE) were used to assess the performance of RF and compare it to two alternative models: naïve baseline (NB) and Auto-Regressive Integrated Moving Average (ARIMA). Using these measures, RF outperformed NB and ARIMA in 76% (N = 25/33) of urgent care metrics according to SMAPE, 88% (N = 29/33) according to MAE and 91% (N = 30/33) according to RMSE. The RFs developed in this study have been implemented within the local ICS, providing predictions on an hourly basis that can be accessed by local healthcare planners and managers.

1. Introduction

Predicting healthcare demand, utilisation and performance can help hospitals to optimise the allocation of resources and the forward planning of services. Over the longer term, predictions can support strategic decision making in areas such as elective care, where future waitlist projections can help managers appraise capacity sufficiency and the necessity for any alternative treatment [1]. Over the shorter term, predictions can be useful to support operational and tactical decisions. In this study, we focus on the shorter term, in aiming to develop and implement a solution for near real-time predictions of various urgent care performance metrics across three hospitals in South West England.

Meeting the demand for urgent care is a challenge for many healthcare systems across the world, and England's National Health Service (NHS) is no exception. The constitutional target to treat or admit 95% of Accident and Emergency (A&E) attendances within 4 h has not been met since 2012, with performance deteriorating rapidly to 57% in the two years to 2022 [2]. This is a symptom of a wider system-wide flow issue, that has been exacerbated by COVID-19 demand and workforce recruitment and retention issues [3]. Greater acute 'bed blocking' caused by inadequate social care capacity pushes up acute bed occupancy, making it harder to admit patients from A&E [4]. This increases A&E waits (so reducing 4-hour performance) and leads to A&E crowding, resulting in more delays for ambulances handing over patients to A&E, and longer ambulance response times [5]. It is in these areas where patients most in need of life-saving treatment are exposed to the greatest risk of harm.

While the ultimate solution to this problem is a strategic one, there are operational and tactical responses that can alleviate some of the transient pressures facing urgent care services. In this respect, readily available short term predictions of demand, utilisation and performance can be an asset [6]. For instance, if ambulance staff are forewarned of an impending spike in A&E handover delays, then attempts can be made to ensure extra resources are in place and response times are not significantly impaired. Better still, if hospital managers are alerted to a forthcoming increase in A&E patients requiring hospital admission, then additional efforts can be made to proactively discharge any medically fit patients from the wards; thus averting A&E crowding and the knock-on effects for the ambulance service.

Many investigators have previously considered short term predictions of urgent care related metrics [7–22]. Time series approaches to forecasting A&E attendances have been applied as early as 1988 [7]. Since then, there have been many studies comparing various time series approaches to forecasting attendances. Champion et al. [8] used exponential smoothing and autoregressive integrated moving average

* Correspondence to: 360 Bristol, Marlborough Street, Bristol, BS1 3NX, United Kingdom.
E-mail address: richard.wood16@nhs.net (R.M. Wood).

https://doi.org/10.1016/j.health.2023.100169
Received 13 January 2023; Received in revised form 21 March 2023; Accepted 3 April 2023

(ARIMA) methods to forecast demand for one A&E department in Australia. Boyle et al. [9] used a regression model as well as exponential smoothing and ARIMA methods to forecast monthly, daily, and four hourly demand at 27 Australian A&E departments. Napoli et al. [10] compared four different exponential smoothing methods to forecast emergency activity in four A&E departments in the USA, noting that time series methods, since the COVID-19 pandemic, have performed significantly worse in terms of forecast accuracy due to their sensitivity to previous seasonality.

Other studies have focused on comparing machine learning algorithms to traditional time series methods [11–22]. Jones et al. [11] compared multiple time series algorithms and an artificial neural network approach against a benchmark multiple linear regression method for predicting daily patient volumes at three A&E sites in the USA. Alvarez-Chaves et al. [12] compared various machine learning algorithms to time series algorithms for forecasting A&E occupancy for one site in Spain, finding that time series methods performed better in general but that machine learning techniques performed similarly for shorter term (7 day) forecast horizons. Vollmer et al. [13] analysed a range of machine learning techniques versus more traditional time series methods for two A&E departments in the UK, finding similar performance across the techniques over a forecast horizon of 1 to 7 days. Gafni-Pappas et al. [14] also compared machine learning approaches with univariate time series models for one A&E department in the USA, finding that a random forest model was superior at predicting patient volumes for a 2.5 month forecast horizon. And for a child asthma cohort, AlSaad et al. [15] found a multinomial logistic regression could be outperformed by deep machine learning methods for predicting attendances to A&E.

Beyond A&E itself, Abraham et al. [23] used time series methods to forecast the admission of A&E patients to bedded hospital wards for one site in Australia. Smith et al. [24] derived a metric based on the number of patients with acute respiratory illness staying in A&E for more than four hours, and subsequently trained a classification algorithm to predict wider pressure for two sites in the USA. While these reviewed studies support the value of predictive modelling, it is important to note that the developed approaches are bespoke to the urgent care metric considered. To the best of the authors' knowledge, no published paper concerns the development of a solution which can be scaled across the range of demand, utilisation and performance metrics whose prediction could be valuable to healthcare managers.

In this paper, we review the development and use of a scalable Random Forest (RF) algorithm for obtaining near real-time predictions of urgent care performance metrics at three hospitals in South West England. We demonstrate that the RF outperforms a baseline model and an ARIMA in the local system, and discuss the implementation of the RF for forecasting demand over a 24-hour period.

2. Methods

2.1. Setting

The setting of this study was the Bristol, North Somerset and South Gloucestershire (BNSSG) Integrated Care System (ICS) – a local NHS healthcare system in South West England. The BNSSG ICS serves a one million resident population across a mixture of urban, rural and coastal communities, with an overall age profile similar to that of England. Within BNSSG are three hospitals with A&E departments: Bristol Royal Infirmary (BRI, 669 beds); Southmead Hospital (SH, 800 beds); and Weston General Hospital (WGH, 261 beds). WGH has restricted A&E opening hours from 08h00 to 22h00.

BRI and SH are located in Bristol, the largest metropolitan area in BNSSG. BRI is located in central Bristol, whereas SH is in North Bristol. Weston General Hospital is located in Weston-Super-Mare, a coastal town in North Somerset, 20 miles south-west of Bristol. The population of Bristol is younger (13% over 65) and more ethnically diverse (16% non-white) [25] than North Somerset (24% over 65, 3% non-white) [26].

2.2. Data

Our data contained 11 urgent care performance metrics for BRI, SH and WGH, recorded from 3 November 2021 to 16 October 2022, at a resolution of 15 min. Metrics included the number of ambulance handover delays over 15 min in the last hour; the number of patients in A&E; the number of patients awaiting admission to the hospital from A&E; and total A&E discharges in the last hour (Table 1). Although each metric is available at 15-minute increments, to minimise errors due to fluctuations in the data we forecast each metric at hourly intervals.

2.3. Models

To forecast the 11 urgent care performance metrics, we implemented a RF: a machine learning (ML) algorithm composed of an ensemble of decision trees. During training, a RF will select a random sample of the training data with replacement and fit a decision tree to the sample. This process is repeated many times, the exact number being a parameter of the algorithm, to create an ensemble (or forest) of decision trees [27]. We trained 33 independent RFs to predict the value of each metric, at each hospital, using metric values over the past 24 h plus variables representing *hour of day*, *day of week* and *bank holiday*. Each RF consisted of 100 decision trees, with a minimum leaf size of 5.

For each metric, at each hospital, performance of RF was compared to two alternative models: a naïve baseline (NB) and ARIMA. The NB, currently used in BNSSG, estimates future values of a metric by taking an average of past values at the same time and day, over the previous six weeks. ARIMA is an established linear time series model: it is commonly used for forecasting A&E arrivals, using past values of a time series to predict future values. To capture daily patterns, ARIMA was fitted using the Hyndman–Khandakar algorithm [28], with a 24-hour periodicity.

All models were implemented in R, version 4.2.2 [29]. The *forecast* package [30] was used for ARIMA and *randomForest* [31] for RF. The data was initially split into a development set (3 November 2021 to 15 July 2022) and a test set (16 July 2022 to 16 October 2022). The development set was only used for training the models. The test set was used for comparing model performance and was divided into time-ordered chunks of overlapping 24-hour periods. For each of these 24-hour periods, RF was trained on the development set, plus any preceding whole weeks of data within the test set. Due to computational expense, for each 24-hour period ARIMA was trained on the preceding 6 weeks of data only. As an average of past values, NB did not require any training.

Using the trained models, the 11 variables, at each site, were forecast over the next 24-hour period and model performance was assessed using metrics outlined in Section 2.4. This process was repeated for each 24-hour period in the test set until all data in the test set had been appended to the development set and no test data remained. This resulted in 2,232 sets of 24-hour forecasts for each model.

2.4. Statistical analysis

Individual model performance for forecasting 24-hours ahead was evaluated and compared using three established measures: Mean Absolute Error (MAE); Root Mean Square Error (RMSE); Symmetric Mean Absolute Percentage Error (SMAPE).

Defining the true value of each metric at hour t as y_t, and the predicted value as y'_t, these measures are defined according to:

$$MAE = \frac{1}{24}\sum_{t=1}^{24}|y_t - y'_t|$$

$$RMSE = \sqrt{\frac{1}{24}\sum_{t=1}^{24}(y_t - y'_t)^2}$$

Table 1
Summary of urgent care performance metrics in the test set (16 July to 16 October 2022).

Abbreviation	Metric	BRI		SH		WGH	
		Mean (SD)	Median [IQR]	Mean (SD)	Median [IQR]	Mean (SD)	Median [IQR]
ab-h15-lhr	Number of ambulance handovers lasting over 15 min in the last hour	1.78 (1.38)	2 [1, 3]	2.28 (1.65)	2 [1, 3]	0.70 (0.99)	0 [0, 1]
ab-h15-mid	Number of ambulance handovers lasting over 15 min since midnight	28.62 (17.83)	27 [14, 43]	31.30 (21.63)	27 [13, 49]	7.49 (7.82)	5 [0, 13]
ab-h30-lhr	Number of ambulance handovers lasting over 30 min in the last hour	1.33 (1.27)	1 [0, 2]	1.28 (1.28)	1 [0, 2]	0.43 (0.80)	0 [0, 1]
ab-h30-mid	Number of ambulance handovers lasting over 30 min since midnight	21.79 (14.90)	19 [10, 32]	17.91 (13.92)	14 [7, 26]	4.45 (5.23)	3 [0, 7]
ab-h60-lhr	Number of ambulance handovers lasting over 60 min in the last hour	0.85 (1.10)	0 [0, 1]	0.58 (0.90)	0 [0, 1]	0.26 (0.61)	0 [0, 0]
ab-h60-mid	Number of ambulance handovers lasting over 60 min since midnight	14.49 (11.80)	13 [4, 22]	8.74 (8.93)	5 [2, 13]	2.61 (3.60)	1 [0, 4]
ab-ht-mid	Total ambulance handover time lost since midnight (in minutes)	2607.15 (2350.01)	1995 [700, 4033]	535.82 (782.76)	199 [0, 689]	1647.02 (1961.52)	951 [405, 2145]
ae-avg-tta	Mean time to assessment for patients in A+E (in minutes)	123.33 (50.51)	121 [90, 156]	124.55 (51.39)	125 [91, 159]	62.39 (37.31)	58 [36,80]
ae-dch-lhr	Number of discharges from A+E in the last hour	7.50 (3.40)	7 [5, 10]	9.78 (4.68)	9 [6, 13]	4.06 (2.99)	3 [1, 6]
ae-num-dta	Number of patients with a decision to admit from A+E to hospital	20.49 (8.01)	21 [14, 26]	10.36 (8.35)	8 [4, 15]	19.87 (7.49)	21 [14, 25]
ae-occ	Number of patients occupying A+E	56.88 (14.00)	57 [47, 67]	68.23 (20.11)	67 [53, 83]	42.01 (18.43)	41 [28, 57]

Abbreviations: BRI, Bristol Royal Infirmary; SH, Southmead Hospital; WGH, Weston General Hospital.

$$SMAPE = \frac{100}{24} \sum_{t=1}^{24} \frac{|y_t - y'_t|}{|y_t| + |y'_t|}$$

MAE and RMSE are symmetric, penalising under-prediction and over-prediction equally. All measures have a minimum of 0. MAE and SMAPE are linear in the prediction error (i.e. $y_t - y'_t$) but RMSE is quadratic, meaning for RMSE larger prediction errors are more heavily weighted. Both MAE and RMSE can take any value greater than 0, and the upper bound of each is determined by both the model and the range of true values of each metric. In comparison, SMAPE is bounded between 0 and 100, meaning its upper bound is not influenced by the range of values of each metric, allowing for a more direct comparison across metrics and models.

3. Results

3.1. Data

3.1.1. Urgent care performance metrics

Table 1 details the metrics considered in this study, their abbreviations (used in Figs. 2 and 3), and descriptive statistics in the test set (the period from 16 July to 16 October 2022).

Of the 11 metrics for each hospital, six are counts of ambulance handover delays, two are measures of time (in minutes) and three patient counts. Metrics for WGH are, on average, lower than those for BRI and SH, with the exception of *Total ambulance handover time lost since midnight (in minutes)* and *Number of patients with a decision to admit from A+E to hospital*, reflecting the smaller size of WGH compared to BRI and SH.

3.1.2. Population

Table 2 details demographic information of the patients who attended each A&E department between 16 July and 16 October 2022. During this period, of all A&E attendances in BNSSG, SH had the largest number (93,027 (45%)) and WGH the smallest (43,969 (21%)). The age profile of patients attending each department reflects their geographical location, with patients attending BRI, an inner-city department, being younger (median age 40 [IQR: 26, 62]) than patients attending WGH (median age 50 [IQR: 26, 72]) which serves a more rural community. The inner-city location of BRI is also reflected in the deprivation profile of patients as measured by Index of Multiple Deprivation (IMD): 20,521 (30%) of patients come from regions with IMD 1–2 (most deprived population deciles) compared to only 8095 (12%) with IMD 9–10 (least deprived population deciles).

3.2. Model comparison

Results, aggregated over both the 33 variables and the 2232 sets of 24-hour predictions within the test set, are provided in Fig. 1. The upper-left panel shows that RF outperforms NB and ARIMA in approximately 75% of predictions, when measured by SMAPE (which, as a standardised assessment, can be used when aggregating over variables of different units and scales). There is no clear trend between day of week and accuracy (upper-right panel), although there is some association with hour of day (lower-left panel). Performance of both RF and ARMIA is best for predictions made between 04h00 and 06h00, decreasing over the course of the day to around 24h00. As the forecasting horizon increases, the accuracy of RF and ARIMA decreases (lower-right panel). In contrast, the accuracy of NB is uniform across the 24-hour forecasting window and consistently less than RF.

Table 2
Demographics of A&E attendances during test set period (16 July to 16 October 2022).

Characteristic	BRI	SH	WGH
Total A&E attendances[a]	70,915 (34%)	93,027 (45%)	43,969 (21%)
Unique patients	68,508	91,611	42,948
Age[b]	40 [26, 62]	44 [27, 67]	50 [26, 72]
Sex[a]			
Male	34,260 (50%)	44,914 (49%)	20,899 (49%)
Female	34,248 (50%)	46,697 (51%)	22,049 (51%)
Index of Multiple Deprivation[a]			
1–2	20,521 (30%)	16,260 (18%)	10,395 (24%)
3–4	16,726 (24%)	18,830 (21%)	7,036 (16%)
5–6	10,805 (16%)	15,252 (17%)	6,640 (15%)
7–8	12,361 (18%)	19,100 (21%)	13,814 (32%)
9–10	8,095 (12%)	22,169 (24%)	5,063 (12%)

Abbreviations: BRI, Bristol Royal Infirmary; SH, Southmead Hospital; WGH, Weston General Hospital.
[a]N (%).
[b]Median [IQR].

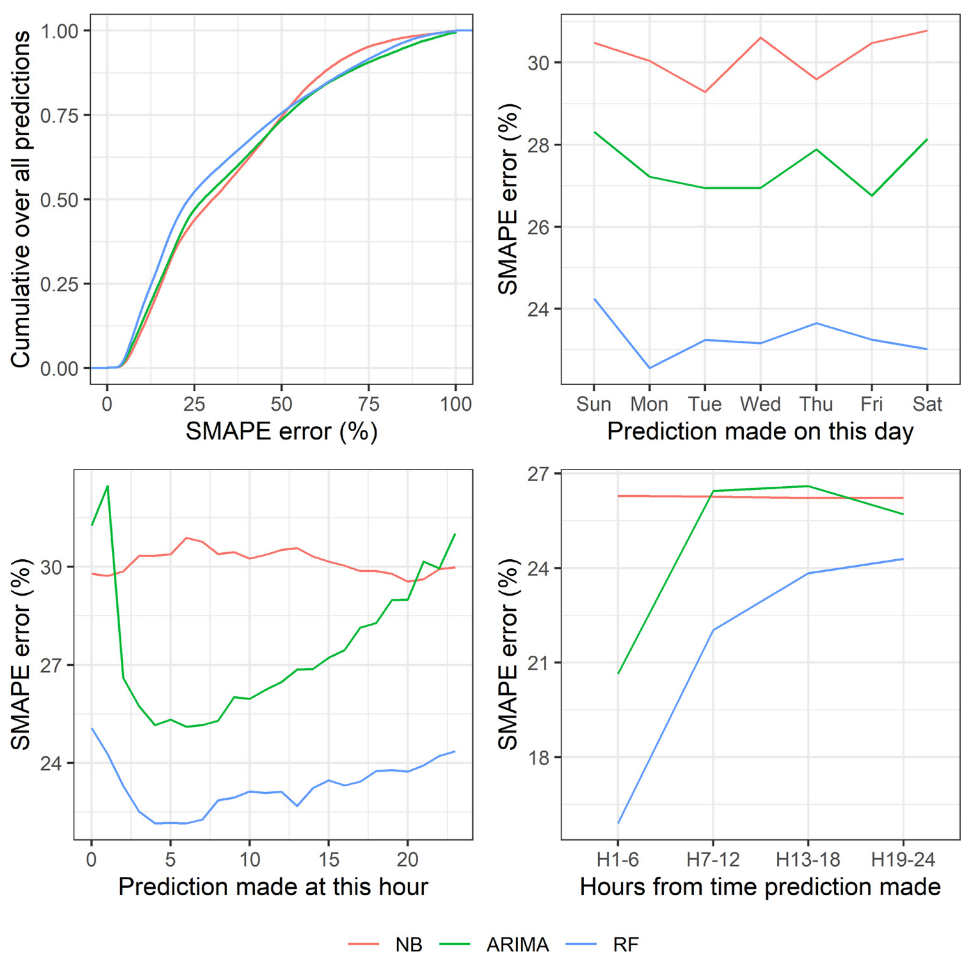

Fig. 1. Accuracy as measured by SMAPE for results aggregated over both the 33 variables (representing the different performance metrics and hospital sites) and the 2232 sets of 24-hour predictions within the test set. Abbreviations: SMAPE, Symmetric Mean Absolute Percentage Error; NB, Naïve Baseline; ARIMA, Auto-Regressive Integrated Moving Average; RF, Random Forest.

The individual performance for each of the 33 metrics is detailed in Fig. 2, including assessments by MAE and RMSE (which can be used when not aggregating variables due to different units of measurement). Comparing models using SMAPE, RF outperformed NB and ARIMA in 76% of cases. For BRI, RF performed better than NB and ARIMA for all 11 metrics. For SH, RF was outperformed in only one metric (*SH_ab-h-60-lhr*) when measured by SMAPE, and one metric (*SH_ae-avg-tta*) when measures by MAE and RMSE. While RF performs best for WGH according to MAE and RMSE, it was outperformed in 7 of 11 metrics when measured by SMAPE. Summarising model performance by aggregating over all metrics, RF performed best (SMAPE 23.0 [13.5, 47.7]), followed by ARIMA (SMAPE 27.3 [14.4, 51.6]) and NB (SMAPE 28.4 [16.1, 48.4]). The full variable-level accuracy profiles are provided in Fig. 3.

3.3. Implementation in BNSSG

RF was implemented in BNSSG on 9 December 2022, during a period of severe pressure on local healthcare services. Accessible through a web browser and updated every hour, it has provided near real-time

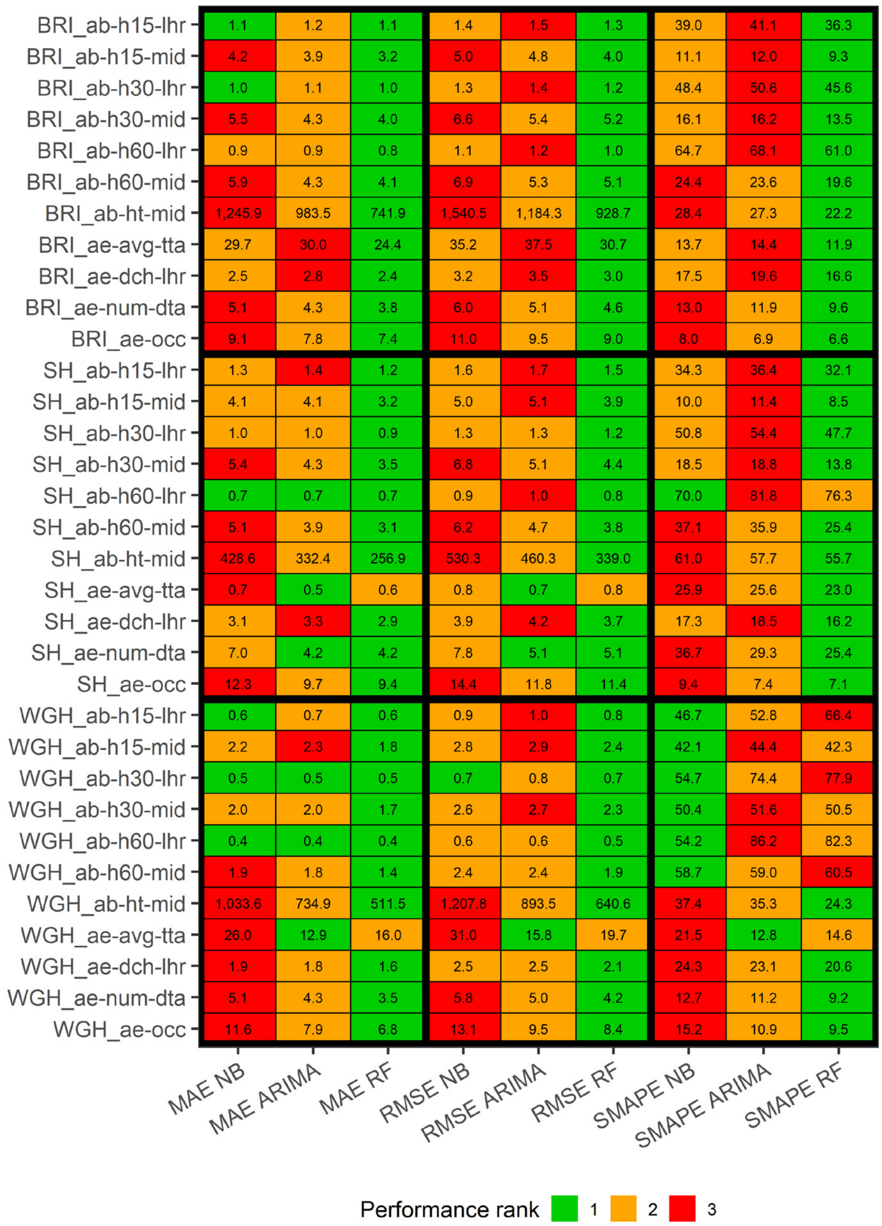

Fig. 2. Accuracy as measured by median MAE, RMSE and SMAPE for the 2232 sets of 24-hour predictions for each variable within the test set. Abbreviations: BRI, Bristol Royal Infirmary; SH, Southmead Hospital; WGH, Weston General Hospital; MAE, Mean Absolute Error; RMSE, Root Mean Square Error; SMAPE, Symmetric Mean Absolute Percentage Error; NB, Naïve Baseline; ARIMA, Auto-Regressive Integrated Moving Average; RF, Random Forest.

predictions for the 11 metrics at each site to clinicians, managers and planners in the healthcare system.

Most recent data is extracted via a bespoke Application Programming Interface (API) every hour. An automated R script applies the pre-trained RF to the data, to produce revised forecasts for the next 24-hour period. The revised forecasts are pushed to an SQL table, which is queried whenever a user opens or refreshes the web application. The application was built using R Shiny [32] and is hosted on a local server. It displays the 24-hour forecasts of each metric at each site (see Fig. 4 for an example).

The models are retrained every week using all available data. Both hourly predictions and weekly re-training of the RF are automated.

In Fig. 4, each plot is divided into three 24-hour periods by the two vertical dashed lines. The right hand panel represents the next 24 h and the centre panel represents the last 24 h. In each plot, predicted values are shown in turquoise and true values are shown in red. The predictions in the centre panel of each plot were made 24 h prior to the current time, i.e. the current time in Fig. 4 is 9 December 16h30 and the predictions in the centre panel were made on 8 December 16h30.

The predicted values (turquoise line) are the mean of the 100 predictions by the 100 decision trees in each RF. The shaded areas represent the 80% (dark grey) and 95% (light grey) confidence range, calculated from the values predicted by the individual decision trees in each RF.

4. Discussion

This study addresses a deficit in the literature of near real-time solutions for forecasting urgent care hospital performance metrics. We have demonstrated that RF outperforms both NB (a simple baseline approach) and ARIMA in 76% of metrics according to SMAPE, and presented its implementation in the local system. Our results provide an example of how ML models, trained on historical data, can be implemented in local healthcare systems to provide operational insight to service planners.

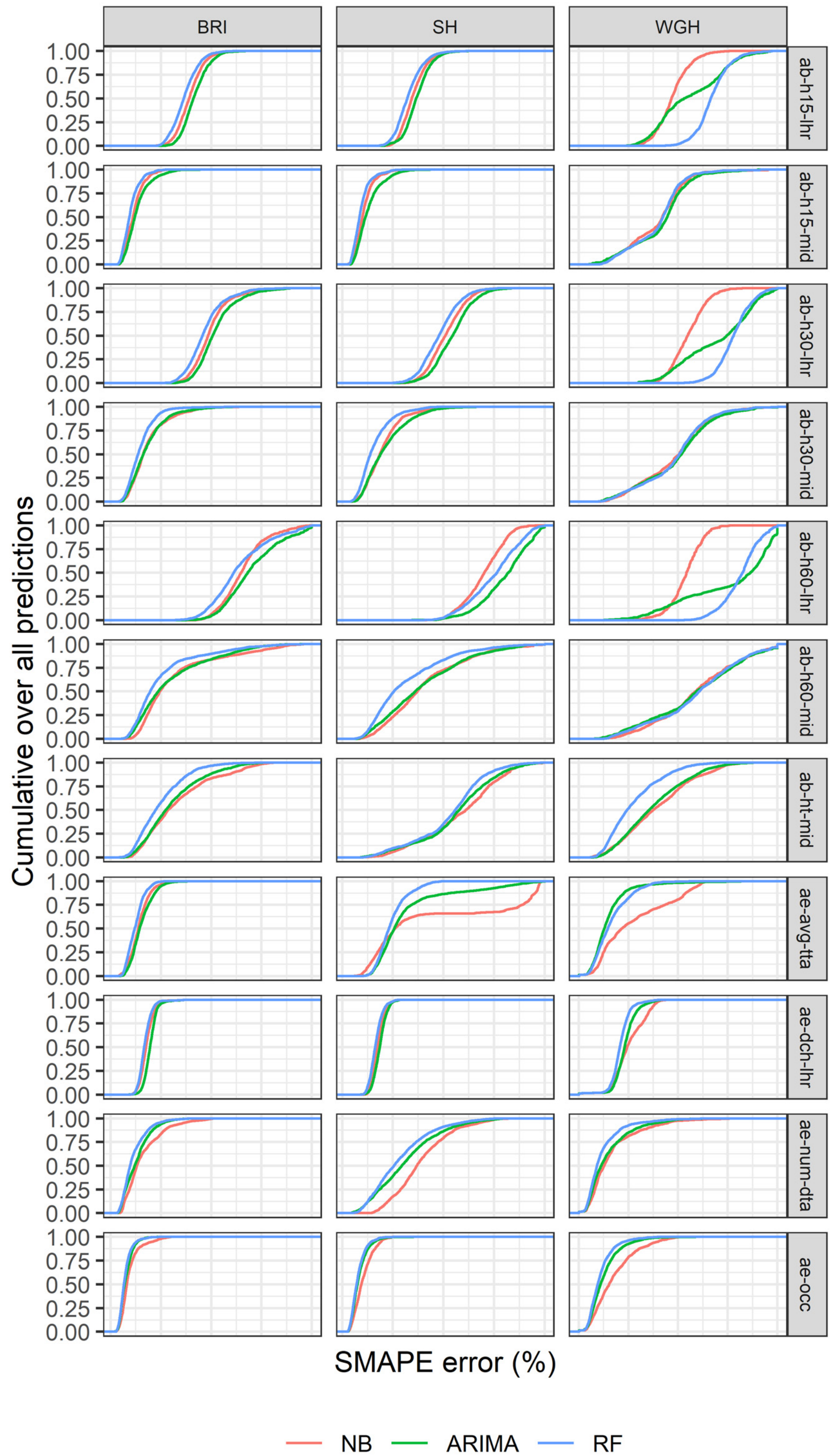

Fig. 3. Distribution function of accuracy as measured by SMAPE for the 2232 sets of 24-hour predictions for each variable within the three-month assessment period to 16 October 2022.

We used RF as it is known to be robust, can handle non-linear relationships between independent and dependent variables, and has lower variance compared to a single decision tree. While performance of RF, when averaged across all metrics, is better than NB and ARIMA, we have not carried out an exhaustive comparison of existing models that could be implemented. Previous work has demonstrated that Neural Networks are superior for predicting short-term ED attendances in certain settings [17,33,34], time series models can outperform ML approaches on the same task [11–13,35], and an ensemble of time-series and ML approaches can offer a performance improvement [36]. As RF outperformed ARIMA, our results suggest that, in the BNSSG ICS application, an ML approach is superior to time series models. Although a Neural Network may improve performance compared to RF, this is at the cost of interpretability: RF, while not interpretable, is explainable to service planners familiar with decision trees. Future work will assess whether other decision tree based ensembles, such as

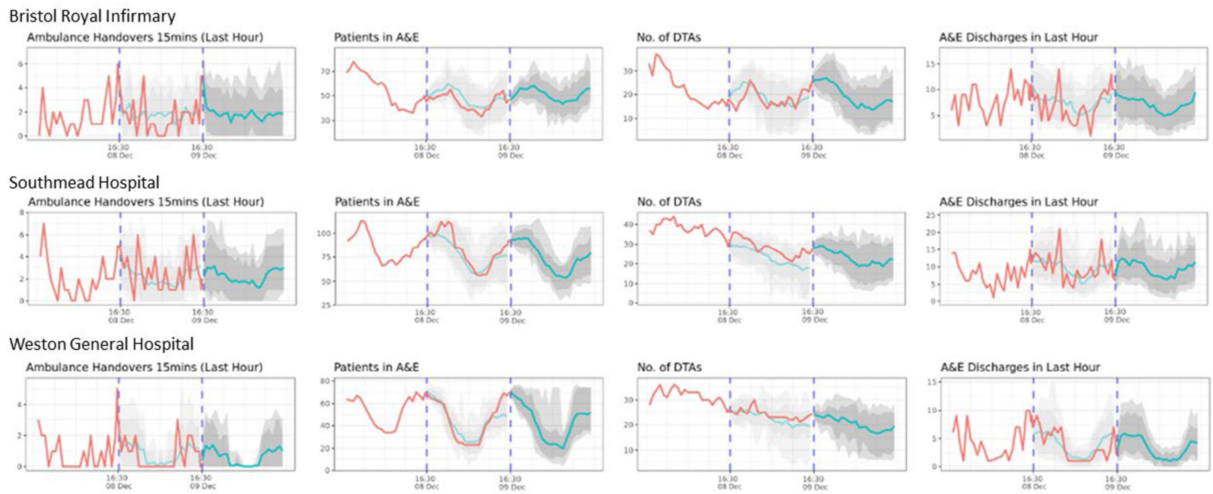

Fig. 4. Example of predictions for four key metrics on the 'summary' page first visible to users on navigating to the web application. This example concerns the set of predictions made at 16h30 on the first day of operation, 9 December 2022. Abbreviations: DTA, decision to admit.

Gradient Boosted Trees [37,38] or Extra Trees [39], which have shown promise in predicting healthcare costs [40,41], offer a performance improvement over RF in the local system.

While we have demonstrated that in our system RF outperforms NB and ARIMA, this result is not necessarily generalisable to other healthcare systems. When comparing model performance using SMAPE, NB outperformed RF in 55% (6/11) of metrics for WGH (Fig. 2), highlighting the importance of model comparison. As a hospital, WGH is the smallest in BNSSG. It has restricted opening hours and its A&E department accounts for only 21% of regional attendances (Table 2). This suggests that, for healthcare systems containing smaller hospitals with lower demand, a simple average of past values may be more appropriate for forecasting urgent care metrics than more complex ML models.

An advantage of RF over ARIMA and NB is that additional variables, extrinsic to the time series, can be incorporated into the model. The implemented RFs included additional variables (*hour of day*, *day of week*, *bank holiday*) which have previously been shown to improve the performance of models predicting A&E attendances [13,17–19,33]. The addition of calendar variables, such as *hour of day* and *day of week*, is necessary for RF to capture daily and weekly patterns in the data, which ARIMA and NB capture intrinsically. Previous work has found that the incorporation of meteorological variables leads to an improvement in model performance [17,19,33,42]. In the future, performance of RF may be improved by incorporating variables relating to historical and/or predicted weather.

To compare model performance across all metrics we used SMAPE as it is bounded between 0 and 100: unlike MAE and RMSE it is independent of the range of values of each metric. However, as a percentage error, SMAPE will penalise small errors in low count data more heavily than in high count data. This is demonstrated when comparing model performance over a 24-hour period (Fig. 1, lower left panel). The accuracy of both RF and ARIMA, as measured by SMAPE, is lowest between approximately 22h00 and 02h00, when demand for urgent care is less. This highlights the importance of considering multiple measures when comparing model performance: no single measure will be informative in all situations.

In terms of practical implications, the RFs takes no more than 30 s of computational time to predict each metric. When parallelised across the 36-core i9 9980XE processors used by the healthcare system, the computational expense of predicting 33 metrics each hour is minimal. In contrast, training the RFs with new data each week takes approximately four hours. To optimise the training process and reduce computational time, further work is needed in the local system to determine how frequently the RFs need to be re-trained, and the optimal set of hyperparameters, for each RF, to minimise training time without affecting model performance.

At the time of writing, the tool has only recently been deployed, but there are already emerging uses of the forecasts. Mainly these account for areas of improving system-wide patient flow and site-level coordination. For instance, setting up diverts for hospital sites that are projected to breach safe occupancy levels, thus transferring demand to other hospitals more able to accommodate the demand. Staffing is also a key area where forecasts can be useful, in reallocating hospital-wide nursing and medical workforce to the A&E when pressures are expected. For the ambulance service, awareness of handover delays can help determine service availability for responding to emergency calls, and any mitigatory actions that may need to be put in place (e.g., conveying fewer patients and attempting to deal with more emergency calls over the phone and without an ambulance response). In the future we plan to extend the application, building on similar work to Pagel et al. [43], by including forecasts of wider hospital flow metrics, such as the number of patients being discharged each hour. Extending the current application with forecasts of additional, more general, metrics has the potential to increase its user base and benefit the operational efficiency of the local system as a whole.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data

Acknowledgements

The authors are grateful to the contributions to this project from Alice Beeching, Chris Davies, Deborah El-Sayed, Scott Parker, Keith Robertson, and Andras Varady. The authors are especially grateful to John Scott for his work on server hosting and automation. The authors must acknowledge the helpful and constructive comments from the editor and two anonymous reviewers, which have improved the quality and legibility of this article.

Funding

CJ is funded by NIHR Bristol BRC, United Kingdom (BRC_1215_20011) and NIHR Research Capability Funding, United Kingdom (RCF 21/22 – 4.2). Otherwise, this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] N.C. Howlett, R.M. Wood, Modeling the recovery of elective waiting lists following covid-19: Scenario projections for england, Value in Health 25 (11) (2022) 1805–1813, http://dx.doi.org/10.1016/j.jval.2022.06.016.
[2] Nuffield Trust, A+E waiting times (last updated 31 2022), 2022, https://www.nuffieldtrust.org.uk/resource/a-e-waiting-times.
[3] British Medical Association, BMA covid review 3: delivery of healthcare during the pandemic, 2022, https://www.bma.org.uk/media/5816/bma-covid-review-report-3-june-2022.pdf.
[4] S. Allan, D. Roland, G. Malisauskaite, K. Jones, K. Baxter, K. Gridley, Y. Birks, The influence of home care supply on delayed discharges from hospital in England, BMC Health Serv. Res. 21 (1) (2021) http://dx.doi.org/10.1186/s12913-021-07206-5.
[5] Royal College of Emergency Medicine, Ambulance handover delays: Options appraisal to support good decision making, 2022, https://rcem.ac.uk/wp-content/uploads/2022/03/Ambulance_Handover_Delays_Joint_Statement_March_2022.pdf.
[6] NHS England, Advanced forecasting techniques: how to use advanced forecasting techniques for estimating demand of NHS services, 2020, https://www.england.nhs.uk/wp-content/uploads/2020/01/advanced-forecasting-techniques.pdf.
[7] P.C. Milner, Forecasting the demand on accident and emergency departments in health districts in the trent region, Stat. Med. 7 (10) (1988) 1061–1072, http://dx.doi.org/10.1002/sim.4780071007.
[8] R. Champion, L.D. Kinsman, G.A. Lee, K.A. Masman, E.A. May, T.M. Mills, M.D. Taylor, P.R. Thomas, R.J. Williams, Forecasting emergency department presentations, Australian Health Rev. 31 (1) (2007) 83, http://dx.doi.org/10.1071/ah070083.
[9] J. Boyle, M. Jessup, J. Crilly, D. Green, J. Lind, M. Wallis, P. Miller, G. Fitzgerald, Predicting emergency department admissions, Emerg. Med. J. 29 (5) (2011) 358–365, http://dx.doi.org/10.1136/emj.2010.103531.
[10] A. Napoli, J. Baird, T. Lin, R. Smith-Shain, 340 The accuracy of predictive analytics in forecasting emergency department volume pre- and post-COVID pandemic, Ann. Emerg. Med. 80 (4) (2022) http://dx.doi.org/10.1016/j.annemergmed.2022.08.367.
[11] S.S. Jones, A. Thomas, R.S. Evans, S.J. Welch, P.J. Haug, G.L. Snow, Forecasting daily patient volumes in the emergency department, Acad. Emerg. Med. 15 (2) (2008) 159–170, http://dx.doi.org/10.1111/j.1553-2712.2007.00032.x.
[12] 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, 2022, http://dx.doi.org/10.21203/rs.3.rs-2322292/v1.
[13] M.A.C. Vollmer, B. Glampson, T. Mellan, S. Mishra, L. Mercuri, C. Costello, R. Klaber, G. Cooke, S. Flaxman, S. Bhatt, A unified machine learning approach to time series forecasting applied to demand at emergency departments, BMC Emerg. Med. 21 (1) (2021) http://dx.doi.org/10.1186/s12873-020-00395-y.
[14] G. Gafni-Pappas, M. Khan, Predicting daily emergency department visits using machine learning could increase accuracy, Am. J. Emerg. Med. 65 (2023) 5–11, http://dx.doi.org/10.1016/j.ajem.2022.12.019.
[15] R. AlSaad, Q. Malluhi, I. Janahi, S. Boughorbel, Predicting emergency department utilization among children with asthma using deep learning models, Healthc. Analytics 2 (2022) 100050, http://dx.doi.org/10.1016/j.health.2022.100050.
[16] Q. Cheng, N.T. Argon, C.S. Evans, Y. Liu, T.F. Platts-Mills, S. Ziya, Forecasting emergency department hourly occupancy using time series analysis, Am. J. Emerg. Med. 48 (2021) 177–182, http://dx.doi.org/10.1016/j.ajem.2021.04.075.
[17] Y. Zhang, J. Zhang, M. Tao, J. Shu, D. Zhu, Forecasting patient arrivals at emergency department using calendar and meteorological information, Appl. Intell. 52 (10) (2022) 11232–11243, http://dx.doi.org/10.1007/s10489-021-03085-9.
[18] M. Ordu, E. Demir, C. Tofallis, A comprehensive modelling framework to forecast the demand for all hospital services, Int. J. Health Plann. Manag. 34 (2) (2019) http://dx.doi.org/10.1002/hpm.2771.
[19] S. Tideman, M. Santillana, J. Bickel, B. Reis, Internet search query data improve forecasts of daily emergency department volume, J. Am. Med. Inform. Assoc. 26 (12) (2019) 1574–1583, http://dx.doi.org/10.1093/jamia/ocz154.
[20] T. Jilani, G. Housley, G. Figueredo, P.-S. Tang, J. Hatton, D. Shaw, Short and long term predictions of hospital emergency department attendances, Int. J. Med. Inform. 129 (2019) 167–174, http://dx.doi.org/10.1016/j.ijmedinf.2019.05.011.
[21] J. Tuominen, F. Lomio, N. Oksala, A. Palomäki, J. Peltonen, H. Huttunen, A. Roine, Forecasting daily emergency department arrivals using high-dimensional multivariate data: A feature selection approach, 2021, http://dx.doi.org/10.21203/rs.3.rs-907966/v1.
[22] L. Zhou, P. Zhao, D. Wu, C. Cheng, H. Huang, Time series model for forecasting the number of new admission inpatients, BMC Med. Inform. Decis. Mak. 18 (1) (2018) http://dx.doi.org/10.1186/s12911-018-0616-8.
[23] G. Abraham, G.B. Byrnes, C.A. Bain, Short-term forecasting of emergency inpatient flow, IEEE Trans. Inf. Technol. Biomed. 13 (3) (2009) 380–388, http://dx.doi.org/10.1109/TITB.2009.2014565.
[24] A.J. Smith, B.W. Patterson, M.S. Pulia, J. Mayer, R.J. Schwei, R. Nagarajan, et al., Multisite evaluation of prediction models for emergency department crowding before and during the COVID-19 pandemic, J. Am. Med. Inform. Assoc. (2022) http://dx.doi.org/10.1093/jamia/ocac214.
[25] JSNA Data Profiles, Bristol City Council, 2023, (Accessed March 16, 2023). https://www.bristol.gov.uk/council-and-mayor/policies-plans-and-strategies/social-care-and-health/joint-strategic-needs-assessment/jsna-data-profiles.
[26] E. Diakou, Spotlight report: north somerset population, 2023, (Accessed March 16, 2023). https://www.n-somerset.gov.uk/sites/default/files/2022-04/JSNA%20population%20spotlight%20report.pdf.
[27] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, http://dx.doi.org/10.1023/a:1010933404324.
[28] R.J. Hyndman, Y. Khandakar, Automatic time series forecasting: the forecast package for R, J. Stat. Softw. 27 (2008) 1–22, http://dx.doi.org/10.18637/jss.v027.i03.
[29] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2022, https://www.R-project.org/.
[30] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmeen, Forecast: Forecasting functions for time series and linear models, 2022, R package version 8.17.0. https://pkg.robjhyndman.com/forecast/.
[31] A. Liaw, M. Wiener, Classification and regression by randomforest, R News 2 (3) (2002) 18–22.
[32] W. Chang, J. Cheng, J.J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert, B. Borges, shiny: Web Application Framework for R. R package version 1.7.4, 2022, https://CRAN.R-project.org/package=shiny.
[33] V.K. Sudarshan, M. Brabrand, T.M. Range, U.K. Wiil, Performance evaluation of emergency department patient arrivals forecasting models by including meteorological and calendar information: A comparative study, Comput. Biol. Med. 135 (2021) 104541, http://dx.doi.org/10.1016/j.compbiomed.2021.104541.
[34] D. Duarte, C. Walshaw, N. Ramesh, A comparison of time-series predictions for healthcare emergency department indicators and the impact of COVID-19, Appl. Sci. 11 (8) (2021) 3561, http://dx.doi.org/10.3390/app11083561.
[35] A. Choudhury, E. Urena, Forecasting hourly emergency department arrival using time series analysis, Br. J. Healthc. Manag. 26 (1) (2020) 34–43, http://dx.doi.org/10.12968/bjhc.2019.0067.
[36] M. Yucesan, M. Gul, E. Celik, A multi-method patient arrival forecasting outline for hospital emergency departments, Int. J. Healthc. Manag. 13 (Sup1) (2020) 283–295, http://dx.doi.org/10.1080/20479700.2018.1531608.
[37] J.H. Friedman, Greedy function approximation: A gradient boosting machine, Ann. Statist. 29 (5) (2001) http://dx.doi.org/10.1214/aos/1013203451.
[38] J.H. Friedman, Stochastic gradient boosting, Comput. Statist. Data Anal. 38 (4) (2002) 367–378, http://dx.doi.org/10.1016/s0167-9473(01)00065-2.
[39] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42, http://dx.doi.org/10.1007/s10994-006-6226-1.
[40] J.W. Robinson, Regression tree boosting to adjust health care cost predictions for diagnostic mix, Health Serv. Res. 43 (2) (2008) 755–772, http://dx.doi.org/10.1111/j.1475-6773.2007.00761.x.
[41] M.A. Morid, K. Kawamoto, T. Ault, J. Dorius, S. Abdelrahman, Supervised learning methods for predicting healthcare costs: systematic literature review and empirical evaluation, in: AMIA... Annual Symposium proceedings. AMIA Symposium, 2017, 2018, pp. 1312–1321.
[42] E.-E. Etu, L. Monplaisir, S. Masoud, S. Arslanturk, J. Emakhu, I. Tenebe, J.B. Miller, T. Hagerman, D. Jourdan, S. Krupp, A comparison of univariate and multivariate forecasting models predicting emergency department patient arrivals during the COVID-19 pandemic, Healthcare 10 (6) (2022) 1120, http://dx.doi.org/10.3390/healthcare10061120.
[43] C. Pagel, V. Banks, C. Pope, P. Whitmore, K. Brown, A. Goldman, M. Utley, Development, implementation and evaluation of a tool for forecasting short term demand for beds in an intensive care unit, Oper. Res. Health Care 15 (2017) 19–31, http://dx.doi.org/10.1016/j.orhc.2017.08.003.