Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Forecasting Daily Emergency Department Arrivals Using High-Dimensional Multivariate Data: A Feature Selection Approach

Jalmari Tuominen (jalmari.tuominen@tuni.fi)
Tampere University
Francesco Lomio
Tampere University
Niku Oksala
Tampere University
Ari Palomäki
Kanta-Häme Central Hospital
Jaakko Peltonen
Tampere University
Heikki Huttunen
Tampere University
Antti Roine
Tampere University

Research Article

Keywords: Emergency department, Crowding, Feature selection, Machine learning, Time series forecasting, Statistical learning

Posted Date: September 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-907966/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Forecasting Daily Emergency Department Arrivals Using High-Dimensional Multivariate Data: A Feature Selection Approach

Jalmari Tuominen[1,a*], Francesco Lomio[2,a], Niku Oksala[1,4], Ari Palomäki[1,3], Jaakko Peltonen[2], Heikki Huttunen[2], Antti Roine[1],

[1]*Tampere University, Faculty of Medicine and Health Technology, Arvo Ylpön katu 34, 33520 Tampere, Finland*

[2]*Tampere University, Faculty of Information Technology and Communication Sciences*

[3]*Kanta-Häme Central Hospital, Emergency Department, Ahvenistontie 20, 13530 Hämeenlinna, Finland*

[4]*Vascular Centre, Tampere University Hospital, Tampere, Finland, Elämänaukio, Kuntokatu 2, 33520, Tampere, Finland*

[*] corresponding author, jalmari.tuominen@tuni.fi
[a] authors contributed equally

Abstract

Background and Objective Emergency Department (ED) overcrowding is a chronic international issue that is associated with adverse treatment outcomes. Accurate forecasts of future service demand would enable intelligent resource allocation that could alleviate the problem. There has been continued academic interest in ED forecasting but the number of used explanatory variables has been low, limited mainly to calendar and weather variables. In this study we investigate whether predictive accuracy of next day arrivals could be enhanced using high number of potentially relevant explanatory variables and document two feature selection processes that aim to identify which subset of variables is associated with number of next day arrivals.

Methods We extracted numbers of total daily arrivals from Tampere University Hospital ED between the time period of June 1, 2015 and June 19, 2019. 158 potential explanatory variables were collected from multiple data sources consisting not only of weather and calendar variables but also an extensive list of local public events, numbers of website visits to two hospital domains, numbers of available hospital beds in 33 local hospitals or health centres and Google trends searches for the ED. We used two feature selection processes: Simulated Annealing (SA) and Floating Search (FS) with Recursive Least Squares (RLS) and Least Mean Squares (LMS). Performance of these approaches was compared against autoregressive integrated moving average (ARIMA), regression with ARIMA errors (ARIMAX) and Random Forest (RF). Mean Absolute Percentage Error (MAPE) was used as the main error metric.

Results Calendar variables, load of secondary care facilities and local public events were dominant in the identified predictive features. RLS-SA and RLS-FA provided slightly better accuracy compared ARIMA. ARIMAX was the most accurate model but the difference between RLS-SA and RLS-FA was not statistically significant.

Conclusions Our study provides new insight into potential underlying factors associated with number of next day presentations. It also suggests that predictive accuracy of next day arrivals can be increased using high-dimensional feature selection approach when compared to both univariate and nonfiltered high-dimensional approach. However, outperforming ARIMAX remains a challenge when working with daily data. Future work should focus on enhancing the feature selection mechanism, investigating its applicability to other domains and in identifying other potentially relevant explanatory variables.

Keywords: Emergency department, Crowding, Feature selection, Machine learning, Time series forecasting, Statistical learning

1. Introduction

Emergency Departments (ED) worldwide serve a crucial purpose, providing immediate care to patients presenting with health conditions that vary from minor to life-threatening. In this setting, the ability to provide timely and high-quality care is of utmost importance. Unfortunately, ED's all over the world suffer from regular overcrowding which has been repeatedly associated with suboptimal care leading to both increased morbidity (1) and increased 10-day mortality (2–4). The ability to successfully forecast future overcrowding would enable better resource allocation that could alleviate the problem or even eliminate it altogether.

Following this rationale, there has been a continued academic interest in ED forecasting (5) but much of the previous work has focused on investigating applicability of different algorithms (6–9) or the predictive value of a singular independent variable such as website visits (10), road traffic flow (11) or aggregated acuity of admitted patients (12). Due to extremely interdependent nature of ED's the number of potential input features is high and testing each of them one by one is a painstaking process. Moreover, since these input features likely demonstrate significant multicollinearity, testing them one by one can provide a misleading picture of their relative importance. Despite these issues, there has been little to no emphasis on the number and quality of the used independent variables and, most importantly, on their aggregated value when used in conjunction with one another.

Reluctancy towards high-dimensional multivariate input is understandable from both computational and practical standpoint. From computational perspective the amount of added noise is usually proportional to number of input dimensions which often leads to

loss of predictive accuracy. Moreover, ED forecasting is almost always performed using statistical time series forecasting algorithms (5) most of which are strictly univariate by design, with the notable exception of regression with ARIMA errors (ARIMAX). It is thus not a coincidence that ARIMAX with very limited and arbitrarily selected calendar and weather variables seems to outperform other statistical models (12,13). We hypothesise, that if this kind of arbitrary feature selection works as well as it does, it should be possible to completely automate the feature selection process, which would make it significantly faster to identify useful input features and potentially enhance model accuracy.

Feature selection processes have conventionally been utilized in pre-processing of imaging and biomedical signals as well as in genetic studies. In addition to eliminating noise and increasing computational speed, they can provide new understanding on the factors behind the phenomenon of interest (14) which could ultimately inform wider health care policies. To our knowledge there is only one publication by *Jiang et al* that has documented a feature selection process specifically in the ED forecasting context. However, even then the selection is done out of a very limited set of weather and holiday variables, which questions the necessity and performance of their approach (15).

In this study we demonstrate a feature selection process to identify predictors of ED crowding using a dataset from a large Nordic ED along with a largest-to-date collection of predictor candidates. Using this data, we test two feature selection mechanisms: simulated annealing and floating search and benchmark our results against current gold standard.

2. Materials and Methods

2.1 Data

Tampere University Hospital is an academic hospital located in Tampere, Finland serving a population of 535,000 in Pirkanmaa hospital district and as a tertiary hospital an additional population of 365,700 and providing level 1 trauma center equivalent capabilities. The hospital ED *"Acuta"* is a combined ED with total capacity of 111-118 patients with 70 beds (and additional 7 beds as a reserve) and 41 seats for walk-in patients. Approximately 100,000 patients are treated annually. For this study, the daily numbers of all registered ED visits were obtained from hospital database created during the sample period from June 1, 2015 to June 19, 2019 resulting in 386 579 individual visits. The number of next day total arrivals (DTA) was used as the target variable.

Based on previous literature and intuition, explanatory variables were collected from different data sources as listed in Table 1. Historical weather data was acquired in hourly resolution from the nearest observation station (16). Timestamps of Finnish holidays were provided by University Almanac Office (17). Calendar variables were encoded according to their status as national holidays and working days. Additionally, we included each national holiday as a categorical variable since their impact on ED service demand likely differs significantly due to different levels of social activity. Weekdays and months were also included as can be expected.

Timestamps of local public events were provided by Tampere city officials. The provided log contained an event name, date of organisation and event size. Two feature

sets were engineered using this data. First, we computed a timeseries of the total number of ongoing events each day within the Tampere area, with the hypothesis that increased activity (and often increased substance consumption) might have an impact in ED service demand. The total number of events was further divided by event size into the number of minor and major public events. Additionally, we identified 73 recurring events that are organized each year. These events were included as individual binary vectors, since, analogous to different holidays, different events likely have different impact on service demand.

A timeseries containing the number of available beds in 34 inpatient facilities in Pirkanmaa Hospital district catchment area was provided by Unitary Healthcare Ltd which provides a logistics software for patient transfers. The rationale of including these features into the dataset resides in the hypothesis that the availability of hospital beds is inversely correlated with ED arrivals. More precisely, if a primary care physician is unable to find a bed for a patient in need, they are often forced to send the patient to the ED merely to organise the bed that the patient requires. In addition to including the capacity of each individual hospital and health care centre we also included both the mean and sum of all the available beds on any given day. Temporal availability of hospital beds in included facilities is visualised in Figure 1.

Figure 1. Temporal availability of beds in 33 catchment area hospitals or health centres as extracted from Uoma© which is a software developed by Unitary Healthcare Ltd. used to facilitate easier patient transfers. Negative availability is drawn as 0 for clarity. White space represents missing data, caused mainly by sequential introduction of the software. There are interesting differences between facilities, some demonstrating constant overload which likely significantly contributes to catchment area access block.

The numbers of website visits to two domains (www.tays.fi and www.tays.fi/acuta) were acquired from Tampere University Hospital Information Management. The former of these was available in hourly resolution and the latter in daily resolution. Daily sums of visits to both domains were included. Additionally, we summed the visits

between 18pm and midnight in the identical manner as was suggested and justified by *Ekström et al* and named this feature as "*Ekström's visits*" (10). Moreover, a stationary version of this variable was included by dividing the evening visits by earlier visits during the day. This variable is referred to as "*Ekström's ratio*". The number of daily Google searches for word "Acuta" was also used as an input (18).

Website visits, Google searches and available hospital beds were lagged by one day whereas weather variables were not, assuming that weather can be forecasted with satisfying precision one day ahead. All explanatory variables are collected and presented in Table 1.

Table 1. List of potential explanatory variables. N = number, Int = integer, float = floating point, N Columns = number of columns

Variable Name	N Columns	Type	Lag (days)
N of available hospital beds	33	Int	-1
N of available hospital beds	1	Float	-1
N of available hospital beds$_{\Sigma}$	1	Float	-1
Weekday	7	Binary	0
Month	12	Binary	0
Specific holiday	18	Binary	0
Lagged holiday	3	Binary	0
Working day	1	Binary	0
Cloud count	1	Int	0
Air pressure	1	Float	0
Relative humidity	1	Float	0
Rain intensity	1	Float	0
Snow depth	1	Float	0
Air temperature	1	Float	0
Dew point temperature	1	Float	0
Visibility	1	Int	0
Air temperature min	1	Float	0
Air temperature max	1	Float	0
Website Visits$_{tays.fi}$	1	Int	-1
Website Visits$_{tays.fi/acuta}$	1	Int	-1
Ekstöm's visits$_{tays.fi}$	1	Int	-1
Ekström's ratio$_{tays.fi}$	1	Int	-1
Google Trends$_{"Acuta"}$	1	Int	-1
N of minor public events	1	Int	0
N of major public events	1	Int	0
N of all public events	1	Int	0
Specific public event	65	Binary	0
	158		

2.2 Models

2.2.1 Benchmark models

Autoregressive Integrated Moving Average (ARIMA) is a widely used statistical forecasting model the performance of which has been previously extensively documented in ED forecasting (19). It has established a position as one of the most important benchmarks not only in ED forecasting but in time series forecasting in general (12,20). Due to established nature of the model, we refer to Chapter 9 of (21) for the basic concepts. In essence, ARIMA is a combination of three components: autoregression (AR), integration (I) and moving average (MA). Integration step serves to ensure stationarity of the data by differencing whereas AR and MA perform the actual modelling and predicting. Number of required differences and the length of history that is used as an input for AR and MA components constitutes the model order which is referred to as (p, d, q) in which p is the number of time lags for AR, d is number of differencing and q is number of time lags for MA. The order of the model is then determined either manually by dedicated statistical procedures or using an automated approach. When additional independent variables are used in conjunction with the univariate historical signal, the model is referred to as regression with ARIMA errors or ARIMAX. For seasonal data, it is often useful to define time lags as a multiple of the known seasonality and perform seasonal differencing, in which case the model is referred to as Seasonal ARIMA or SARIMA. In this study, model order was defined with Auto-ARIMA as initially described by Hyndman et al (22) using a Python implementation provided by *Smith et al* (23). We provide three ARIMA benchmarks:

one trained with both univariate signal and all 158 explanatory variables (ARIMAX-A), one trained only with univariate historical signal (ARIMA) and one trained with features inspired by work of *Whitt et al* (13) (ARIMAX-W) containing a limited number of weather and calendar variables. ARIMAX trained with features identified by simulated annealing and floating search are referred to as ARIMAX-SA and ARIMAX-FS respectively. The known weekly seasonality of the target variable was provided to the optimizer which automatically defines whether seasonal lags are required for best available fit.

We also include Random Forest as a benchmark, which is one of the most used machine learning models and is particularly beneficial in the case of high dimensional data since it natively uses subsets of the input data. In addition, it can work well with features of different types (binary, numerical, categorical). It is an ensemble technique, meaning that is uses a set of simpler models to solve the assigned task (24). In this case, RF uses an ensemble of decision trees. An arbitrary number of decision trees is generated, each considering a randomly chosen subset of the samples of the original dataset. To reduce the correlation between the individual decision trees, a random subset of the features of the original dataset is selected. Each tree is therefore trained on its subset of the data, and it can give a prediction on new unseen data. The RF regressor uses the results of all these trees and averages them to generate the prediction. Four versions of RF with different inputs were tested: RF-U with only univariate signal, RF-FS with variables identified by FS, RF-SA with variables identified with SA and RF-A with all variables.

Naïve and Seasonal Naïve (SNaive) were also included as benchmark models to establish the ultimate baseline of performance. Naïve model uses the latest observed

value as the prediction, e.g. when predicting arrivals of Wednesday, observed values of Tuesday are used. SNaive uses the latest observed value a season ago as the prediction, e.g. when predicting arrivals of next Wednesday, observed value of last Wednesday is used.

2.2.2 LMS and RLS filters

Due to the nature of the data used, characterized by seasonal variations and high number of input dimensions, we focused our attention on classical signal processing including LMS filters and RLS filters (25). These models have the benefit of being both simple and efficient which is required due to high number of train-test iterations in the feature selection phase. LMS and RLS can be characterized as gradient learning models, as they adjust the model parameters according to the gradient of the prediction error.

LMS filter is a digital Finite Impulse Response filter with time-varying (adaptive) weights. As such the LMS filter is commonly used for adaptive signal processing tasks, where the environment changes dynamically such as echo cancellation (25). As the environment in our study is not necessarily stationary, and all latent factors affecting the dynamics are not measurable, the prediction model needs to be able to adapt to the changes in the input-output relationships and the LMS filter is able to do so.

The LMS filter can be formulated as follows. Denote the prediction target (e.g. ED arrivals) at time step n as y(n), and inputs as $\mathbf{x}(n)$, n=1,2,…,N. The inputs are constructed as a vector, whose elements in our case consist of both endogenous

variables (historical values of arrivals) and explanatory variables. The LMS filter predicts the output $\hat{y}(n)$ as a weighted sum (inner product) of inputs and weights:

$$\hat{y}(n) = \boldsymbol{h}(n)^T \boldsymbol{x}(n)$$

The weight vector **h**(n) is initialized with zeros and adaptively updated. The update computes the prediction error $e(n) = y(n) - \hat{y}(n)$ and applies the gradient update rule:

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \mu e(n) \boldsymbol{x}(n)$$

where $\mu > 0$ is the learning rate.

The Recursive Least Squares (RLS) filter is another adaptive filtering formulation, that has significantly faster convergence compared to LMS. The RLS filter is approximate the theoretical solution for the weight vector **w** minimizing the prediction error:

$$\boldsymbol{w}(\boldsymbol{n}) = \boldsymbol{R}^{-1}(\boldsymbol{n}) \boldsymbol{r}(\boldsymbol{n}),$$

where **R** is the expectation of the autocorrelation matrix of input x, and **r** is the expectation of the cross-correlation of input x and target y:

$$\boldsymbol{R}(\boldsymbol{n}) = \sum_{\boldsymbol{i}=\boldsymbol{0}}^{\boldsymbol{n}} \boldsymbol{\lambda}^{\boldsymbol{n}-\boldsymbol{i}} \boldsymbol{x}(\boldsymbol{i}) \boldsymbol{x}^{\boldsymbol{T}}(\boldsymbol{i}),$$

$$\boldsymbol{r(n)} = \sum_{\boldsymbol{i=0}}^{\boldsymbol{n}} \boldsymbol{\lambda^{n-i}\, y(i) x(i)},$$

Under a nonstationary situation, these correlations must be computed for each time step. In practical implementation, the expectations are replaced by their sample-based estimates which are updated at each time step to minimize a weighted prediction error that downweighs older errors. Moreover, the RLS algorithm directly updates the inverse of the autocorrelation matrix in order to avoid matrix inversion. Similar to the learning rate of the LMS filter, the speed of adaptation of the RLS filter can be controlled by the forgetting factor λ, which determines the weight given to old measurements.

2.3 Feature selection

To obtain the most important features in terms of predictive accuracy, we used two different techniques: simulated annealing (SA) and floating search (FS). These algorithms were chosen since they are both fast to deploy and easy to understand. Moreover, both provide a faster execution compared to other greedy feature selection techniques, while still maintaining excellence performance.

SA consists of selecting an arbitrary variable and randomly selecting a neighbor to minimize the internal energy of the system. More specifically: for each variable selected, the algorithm selects a second and checks whether the new "solution" is better (low energy state) or worse than the previous one. If the selected feature improves the overall result, it is kept, otherwise a new variable is tested.

FS feature selection, iteratively adds and removes some of the variables until it reaches a stable subset of features. During the addition phase, the algorithm tests recursively one feature at the time, adding a new feature if this improves the result: this is done until 10 features are added. In the removal phase, it removes one feature at the times from the subsect selected in the previous phase, until the 5 least beneficial features are removed. The FS continues until it doesn't exist a set of 10 features which improves the result when added, nor it exists a set of 5 features which improve when removed.

Both LMS and RLS were used as predictive models in feature selection phase, resulting in four models which are later referred to as LMS-FS, LMS-SA, RLS-FS and RLS-SA.

2.4 Cross validation, error measures and statistical tests

The dataset was divided into training set containing the samples from June 1, 2015 to December 31, 2017 (944 days, 64 %) and test set containing the samples from January 1, 2018 to June 19, 2019 (534 days, 36 %). Out-of-sample accuracies over the test set were calculated using a rolling forecast origin with predictive horizon of one day. Mean Absolute Percentage Error (MAPE) was used as the error metric since it is scale-invariant and because its wide adoption allows comparisons to previous studies (5). The formula for MAPE is defined as follows:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}$$

where n = number of samples, y_i = ground truth, $\hat{y}_i$= prediction.

We used ANOVA and two-tailed Dunnett's post-hoc test to investigate statistical significance between reported MAPE's. Multiple comparisons to both Seasonal Naïve and to the best performing model were performed. Statistical significance was specified as $P < .05$. Statistical analyses were performed using SPSS Statistics version 27.0.1.0.

3. Results

3.1 Model Accuracy

ANOVA showed statistically significant differences between models with $p<0.001$. Model performance and multiple comparisons are presented in Table 2 and predictions are visualized in Figures 2 and 3. ARIMAX-W(2,0,2) provided the best out-of-sample accuracy with MAPE of 6.6 % but did not differ statistically from RLS-FS or RLS-SA. Estimated coefficients of this model are provided in Table 3. RLS was identified as the second-best model with MAPE of 6.9 % when trained with SA features and MAPE of 6.9 % when trained with features identified by FS. Univariate LMS resulted in MAPE of 7.0 %. LMS-U, RLS-SA and RLS-FS outperformed univariate ARIMA(1, 0, 0)x(1, 0, 0)$_7$ which provided an accuracy of 7.1 %.

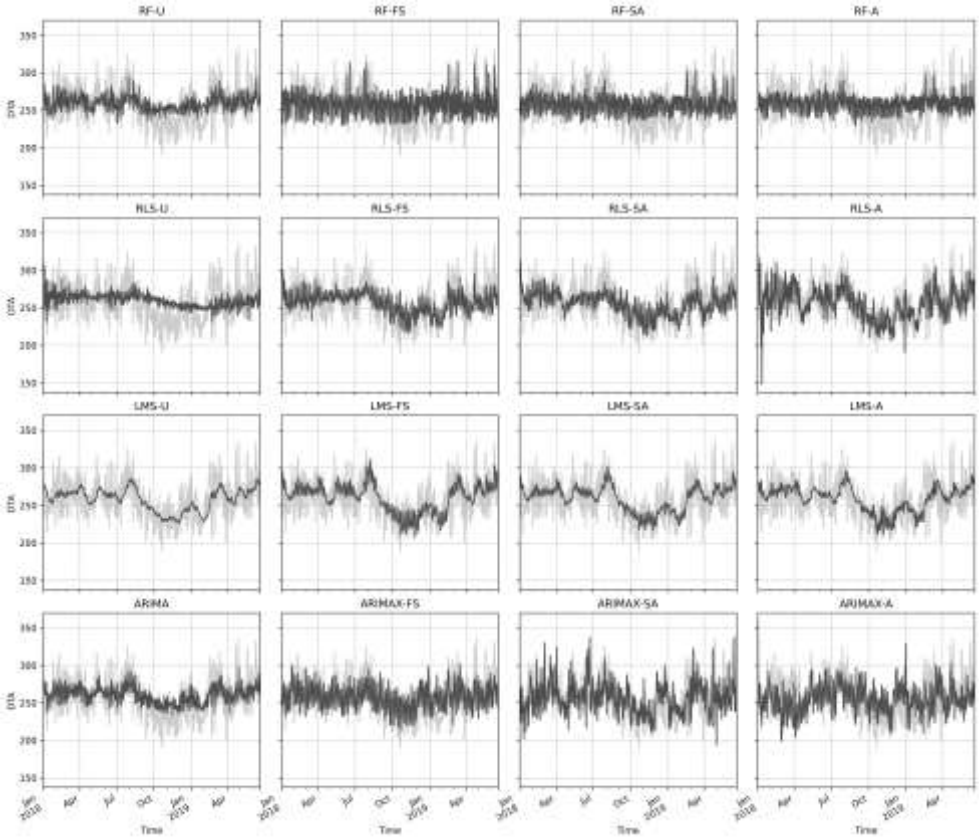

Figure 2. Predictions superimposed with ground truth. Light grey line = ground truth, dark grey line = prediction. RF = random forest, RLS = recursive least squares, LMS = least mean squares, ARIMA = autoregressive integrated moving average, ARIMAX = regression with ARIMA errors, FS = floating search, SA = simulated annealing

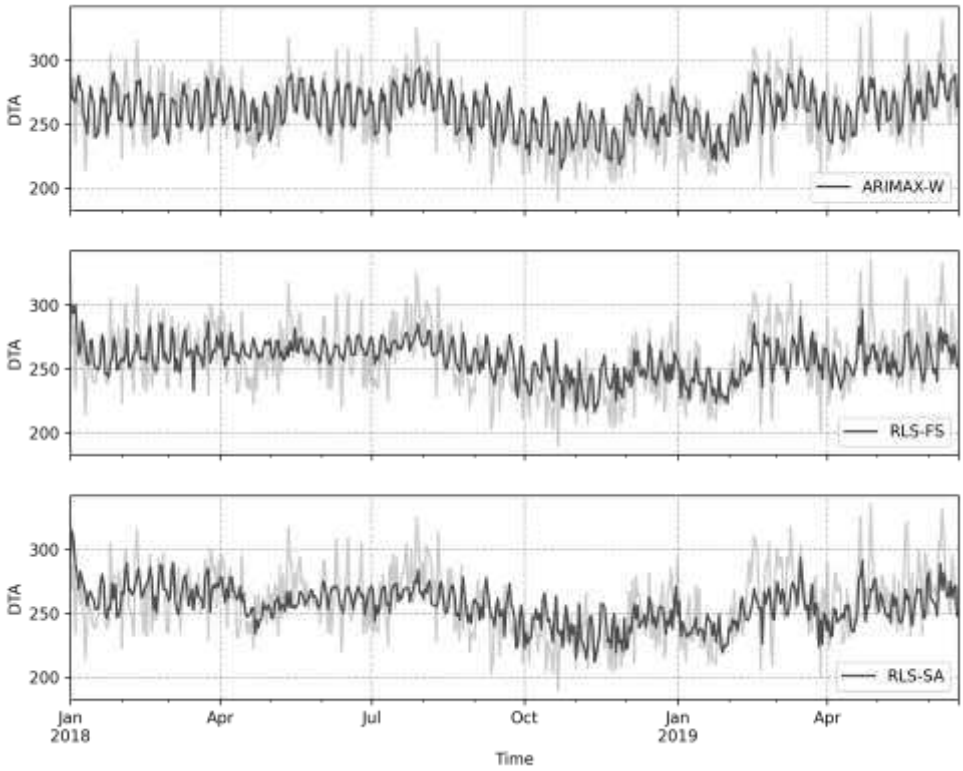

Figure 3. Three best performing models. Light grey line = ground truth, dark grey line = prediction. ARIMAX-W = regression with ARIMA errors using features identified by Whitt et al (13), RLS = recursive least squares, SA = simulated annealing, FS = floating search

Table 2. Model accuracies in terms of absolute percentage errors. ARIMA = autoregressive integrated moving average, ARIMAX = regression with ARIMA errors, RLS = recursive least squares, RF = random forest, LMS = least mean squares, SA = simulated annealing, FS = floating search, SNaive=seasonal naïve, A = all features, U = univariate, W = Whitt's features. Statistical significance is calculated using two-tailed ANOVA with Dunnet's post hoc test for multiple comparisons.

	Mean	Standard deviation	Median	Max	Differs from SN (p)	Worse than best (p)
Naive	8.4	6.4	6.9	36.4	1.00	**< .001**
ARIMAX-A	8.4	6.2	6.9	33.7	1.00	**< .001**
RLS-U	8.3	6.2	7.1	37.7	1.00	**< .001**
SNaive	8.2	6.6	6.6	41.8		**< .001**
ARIMAX-SA	8.0	6.5	6.5	39.0	1.00	**< .001**
RF-FS	8.0	5.9	6.6	33.5	1.00	**.002**
LMS-FS	7.8	5.9	6.5	32.6	.98	**.007**
RF-SA	7.7	5.7	6.5	28.5	.72	**.035**
RF-U	7.5	5.7	6.1	33.2	.42	.10
RF-A	7.4	5.7	6.4	36.6	.22	.22
LMS-A	7.3	5.6	6.3	34.3	.16	.30
ARIMAX-FS	7.3	5.9	5.9	36.2	.12	.37
LMS-SA	7.2	5.5	6.1	31.6	.07	.53
RLS-A	7.2	5.5	6.4	39.3	**.048**	.64
ARIMA	7.1	5.5	5.7	29.5	**.019**	.86
LMS-U	7.0	5.3	5.8	30.7	**.011**	.95
RLS-SA	6.9	5.1	5.9	24.6	**.003**	1.00
RLS-FS	6.9	5.2	5.9	30.1	**.002**	1.00
ARIMAX-W	6.6	5.3	5.3	31.7	**< .001**	

Table 5. Estimated coefficients of the ARIMAX-W(2,0,2) model. ϕ = non-seasonal autoregression, θ = non-seasonal moving average.

	Estimate	Standard error	p
January	112.93	3.68	<0.001
February	111.17	3.30	<0.001
March	101.35	3.80	<0.001
April	90.24	3.70	<0.001
May	83.41	4.70	<0.001
June	84.78	3.49	<0.001
July	81.19	4.08	<0.001
August	78.43	4.39	<0.001
September	86.49	3.69	<0.001
October	88.64	3.46	<0.001
November	94.97	3.09	<0.001
December	109.51	3.16	<0.001
Monday	170.97	2.00	<0.001
Tuesday	148.29	1.94	<0.001
Wednesday	147.47	1.97	<0.001
Thursday	145.46	2.23	<0.001
Friday	164.24	2.04	<0.001
Saturday	176.05	2.10	<0.001
Sunday	170.63	2.05	<0.001
Min temp	0.45	0.21	0.03
Max temp	0.89	0.23	<0.001
Holiday+1	5.68	3.35	0.09
Holiday+0	-8.57	2.99	<0.001
Holiday-1	19.12	2.66	<0.001
φ_1	-0.11	0.14	0.44
φ_1	0.69	0.10	<0.001
θ_1	0.28	0.14	0.05
θ_2	-0.58	0.10	<0.001
σ^2	352.37	16.26	<0.001

3.2 Identified Features

For the sake of brevity, only features identified by better performing RLS are presented here. RLS-SA identified a total of 62 features, out of which 30 were individual public events, 11 were available beds vectors from wards, and 8 were holiday variables. *Ekström's visits* were included as were the numbers of major and all public events. All weekdays were included except Saturday. *December*, *September,* and *March* were identified as impactful. Out of weather variables all but *snow depth* were excluded. Please see Table 3 for details.

RLS-FS identified a total of 55 features, out of which 29 were individual public events and 7 were individual holidays. Website visits to both domains were included. Out of weather variables all but *cloud count* were excluded. All weekdays were included, but out of months only *March*, *February* and *December* were considered significant. Please see Table 4 for details.

Table 3. Most important explanatory variables for next day arrivals identified by simulated annealing and recursive least squares

Feature Family	Feature
Website visits	Ekströms visits
Holiday name	Independence Day Eve
Holiday name	Easter Day
Holiday name	Shrove Sunday
Holiday name	All Saint's Day
Holiday name	May Day
Holiday name	Ascension Day
Holiday	Holiday_{t+0}
Holiday	Holiday_{t+1}
Available hospital beds	Regional hospital A, Ward 9
Available hospital beds	Health centre 10
Available hospital beds	Regional hospital A, Ward 8
Available hospital beds	Health centre 12
Available hospital beds	Regional hospital A, Ward 5
Available hospital beds	Health centre 11, Ward 3
Available hospital beds	Health centre 2
Available hospital beds	Health centre 11, Ward 2
Available hospital beds	Regional hospital B, Ward 1
Available hospital beds	University hospital, ED ward
Available hospital beds	Health centre 11, Ward 1
Month	December
Month	September
Month	March
Public event	30 individual public events*
Public event	Number of major daily public events
Public event	Number of total daily public events
Weather	Snow depth
Weekday	Sunday
Weekday	Monday
Weekday	Wednesday
Weekday	Friday
Weekday	Thursday
Weekday	Tuesday

* Individual public events are not shown here due to their high number

Table 4. Most important explanatory variables for next day arrivals identified by floating search and recursive least squares

Feature Family	Feature
Holiday name	Shrove Sunday
Holiday name	Easter Day
Holiday name	Midsummer
Holiday name	Christmas Eve
Holiday name	All Saint's Day
Holiday name	Independence Day Eve
Holiday name	Ascension Day
Holiday	Holiday_{t-1}
Available hospital beds	Health centre 2
Available hospital beds	Health centre 11, Ward 1
Available hospital beds	University hospital, ED ward
Calendar variable	Working day
Month	March
Month	February
Month	December
Public event	29 individual public events*
Public event	Number of major public events
Weather	Cloud count
Website visits	$\text{Website visits}_{\text{tays.fi/acuta}}$
Website visits	$\text{Website visits}_{\text{tays.fi}}$
Weekday	Thursday
Weekday	Saturday
Weekday	Friday
Weekday	Wednesday
Weekday	Tuesday
Weekday	Sunday
Weekday	Monday

* Individual public events are not shown here due to their high number

4. Discussion

To the best of our knowledge, this was the first study to investigate feature selection in truly high-dimensional multivariate ED forecasting. We demonstrated that using high-dimensional multivariate input in conjunction with appropriate feature selection slightly enhances predictive accuracy when compared to using complete feature set or a univariate model. Calendar variables, load of secondary care facilities and local public events were dominant in the identified predictive features.

Both feature selection methods resulted in a somewhat similar collection of features and in almost identical predictive accuracies. A high number of local public events was included in both feature sets, some of which are intuitively unlikely to have marked impact on ED service demand mostly due to their small size. It is possible that some public events end up in the final feature set not because they are especially important but simply because of their abundance. For example, in the case of FS, a high number of features increases their likelihood to appear in the addition phase which might risk an increase in false positives. It is also difficult to differentiate the impact of the weekly seasonality from the impact of the public events since most of the public events are naturally organized in the weekend. It is possible that the weekly seasonality "leaks" into the public event variables due to multicollinearity with calendar variables.

Capacity of many secondary care facilities was prominent among explanatory variables identified by SA. If any underlying causality can be assumed, it serves to highlight the interdependent nature of the ED and importance of access block as an important contributor to overcrowding as previously suggested by (26) and as hypothesised above.

RLS-FS provided better accuracy than the 8.4 % that was documented by Whitt et al using a ARIMAX model (13). However, reproducing the approach of *Whitt el al* on our data (ARIMAX-W) produced the best accuracy with 6.6 % suggesting that MAPE errors are not directly comparable over different facilities despite the desired scale-invariance of the metric. *Ekström et al* documented one day ahead accuracy of 6.1 % in two ED's with similar size as ours using a General Linear Model (GLM) with website visits and calendar variables as inputs (10). Interestingly both of our feature selection algorithms included website visits in the final feature set supporting findings of *Ekström et al* but, the resulting accuracy was slightly worse than they documented. We believe this is at least in part due to relatively short validation set of 3 months used by *Ektsröm et al*, in which the inability of a GLM to adjust to changes in the time series does not become evident in the manner that can be seen with RF in our study (Figure 2) which leads to overly optimistic interpretation of model performance.

To the best of our knowledge, as previously stated, there is only one article that has previously investigated feature selection processes specifically in the context of ED forecasting by Jiang et al (15). They documented an approach in which a Genetic Algorithm was used for feature selection prior to fitting a Deep Neural Network (DNN). However, their initial feature space contained mere 22 dimensions consisting completely of calendar and weather variables and it begs the question of whether performing dimensionality reduction in their setting makes sense in the first place. This question will remain unanswered, since they don't document the performance of DNN with the complete feature set. Moreover, Jiang et al divided their test set of 128 days into 6 folds and report aggregated accuracies for different forecasting horizons. For

these reasons it is impossible to make meaningful comparisons between their and our results.

In broader context, feature selection in multivariate time series forecasting is a relatively under-examined subject and readily available software solutions do not exist. For this reason, it would be interesting to see how our approach generalises into other domains such as industrial, commercial, or econometric forecasting in which high-dimensional multivariate time series are abundant but manual feature selection is either impractical or impossible. In retail, for example, the number of target variables of interest are often counted in tens of thousands, and costs of performing any manual model engineering for each target independently greatly surpasses the benefits of potential aggregated accuracy increase. However, computational extraction of relevant features as suggested in this study could result in significant accuracy increase with marginal labour cost.

Neural networks (NN) are readily applied in fields such as machine vision in which number of input dimensions is inherently extremely high, but their use specifically in time series prediction has been a challenge. Only recently a NN used in conjunction with a statistical model outperformed pure statistical time series tools in the M4 time series forecasting competition (20). Following this result, some potentially performant multivariate NN algorithms for time series forecasting have appeared (27) and documenting their performance in ED forecasting with high number of features would be an interesting subject for a follow-up study.

4.1 Limitations

Despite the carefully performed cross validation and moderate size of the validation set, this was a retrospective cohort study, and its results must be confirmed in a prospective setting. This is mainly due to inherent uncertainty in the accuracy of the older visit statistics. Our study suggests that adding non-conventional exogenous variables such as public events and availability of hospital beds as inputs in a predictive model might increase model performance. However, availability of these inputs in a prospective setup might be a challenge in a hospital with suboptimal IT infrastructure. We observed a significant drop in the DTA from September 3, 2018 onwards due to a reorganization of the ED in which underaged patients were redirected to a newly opened pediatric ED. This most likely has a negative impact in the model performance, and it should be considered when interpreting the results. There was a non-trivial amount of missing data in available hospital beds because the software that was used to monitor capacities was introduced sequentially one hospital at a time during the period of our train set. This might have had a negative impact on model performance. Please see Figure 1 for details.

4.2 Conclusions

Our study provides new insight into potential underlying factors associated with number of next day presentations. It also suggests that predictive accuracy of next day arrivals can be increased using high-dimensional feature selection approach when

compared to both univariate and nonfiltered high-dimensional approach. However, outperforming ARIMAX remains a challenge when working with daily data. Future work should focus on enhancing the feature selection mechanism, investigating its applicability to other domains, and in identifying other potentially relevant explanatory variables.

Additional files

Additional File 1: Target variable data. The table contains all daily total arrivals in a machine-readable format observed in the study period (1/6/2015 – 19/6/2019). (XLS 39 kb)

Abbreviations

ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal ARIMA
ARIMAX	Regression with ARIMA errors
RLS	Recursive Least Squares
LMS	Least Mean Squares
MAPE	Mean Absolute Percentage Error
RF	Random Forest
GLM	General Linear Model
FS	Floating Search
SA	Simulated Annealing
ED	Emergency Department
IT	Information Technology
DTA	Daily Total Arrivals

Declarations

Ethics approval and consent to participate

Since our study was retrospective in nature, an approval from the ethics committee was not required. An institutional approval was acquired prior to data collection with following specifications.

Name: Potilaslogistiikan häiriötekijöiden tunnistaminen ja mallintaminen

Number: PSHP/R19565

Date: June 16, 2019

Consent for publication

Not applicable

Availability of data and materials

Complete time series of daily total arrivals is provided along with this manuscript (Additional File 1). We do not have ownership to and are not in the position to share other data that was used to generate explanatory variables.

Conflict of interest statement

NO is a shareholder of Unitary Healthcare Ltd. which has developed patient logistics system currently used in the study emergency department. JT, FL and AR are shareholders of Aika Analytics Ltd. which is a company specialized in time series forecasting.

Funding

The study was funded by the Ministry of Health and Social Welfare in Finland via the Medical Research Fund of Kanta-Häme Central Hospital, by The Finnish Medical Foundation and the Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital and Pirkanmaa Hospital District Grants 9X040

Authors' contributions

Study design (AR, JT, NO, AP, HH). Data collection (JT, NO). Data-analysis (FL, HH, JT, JP). Manuscript preparation (AR, FL, JT, JP). All authors read and approved the final manuscript.

Acknowledgements

We acknowledge Unitary Healthcare Ltd for providing dataset on available hospital beds, City of Tampere for providing timestamps for public events and Tampere University hospital information management for providing website visit statistic

References

1. McCarthy ML, Zeger SL, Ding R, Levin SR, Desmond JS, Lee J, et al. Crowding Delays Treatment and Lengthens Emergency Department Length of Stay, Even Among High-Acuity Patients. Ann Emerg Med [Internet]. 2009;54(4):492-503.e4. Available from: http://dx.doi.org/10.1016/j.annemergmed.2009.03.006

2. Jo S, Jeong T, Jin YH, Lee JB, Yoon J, Park B. ED crowding is associated with inpatient mortality among critically ill patients admitted via the ED: Post hoc analysis from a retrospective study. Am J Emerg Med [Internet]. 2015;33(12):1725–31. Available from: http://dx.doi.org/10.1016/j.ajem.2015.08.004

3. Berg LM, Ehrenberg A, Florin J, Östergren J, Discacciati A, Göransson KE. Associations Between Crowding and Ten-Day Mortality Among Patients Allocated Lower Triage Acuity Levels Without Need of Acute Hospital Care on Departure From the Emergency Department. Ann Emerg Med [Internet]. 2019;74(3):345–56. Available from: https://doi.org/10.1016/j.annemergmed.2019.04.012

4. Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. Med J Aust. 2006;184(5):213–6.

5. Gul M, Celik E. An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. Heal Syst [Internet]. 2018;00(00):1–22. Available from: https://doi.org/10.1080/20476965.2018.1547348

6. Harrou F, Dairi A, Kadri F, Sun Y. Forecasting emergency department overcrowding: A deep learning framework. Chaos, Solitons and Fractals [Internet]. 2020;139:110247. Available from:

https://doi.org/10.1016/j.chaos.2020.110247

7. Sharafat AR, Bayati M. PatientFlowNet: A Deep Learning Approach to Patient Flow Prediction in Emergency Departments. IEEE Access. 2021;9:45552–61.

8. Zhou L, Zhao P, Wu D, Cheng C, Huang H. Time series model for forecasting the number of new admission inpatients. BMC Med Inform Decis Mak. 2018;18(1):1–11.

9. Huang Y, Xu C, Ji M, Xiang W, He D. Medical service demand forecasting using a hybrid model based on ARIMA and self-adaptive filtering method. BMC Med Inform Decis Mak. 2020;20(1):1–14.

10. Ekström A, Kurland L, Farrokhnia N, Castrén M, Nordberg M. Forecasting emergency department visits using internet data. Ann Emerg Med [Internet]. 2015;65(4):436-442.e1. Available from: http://dx.doi.org/10.1016/j.annemergmed.2014.10.008

11. Rauch J, Hübner U, Denter M, Babitsch B. Improving the prediction of emergency department crowding: A time series analysis including road traffic flow. Stud Health Technol Inform. 2019;260:57–64.

12. Cheng Q, Tanik N, Scott C, Liu Y, Platts-mills TF, Ziya S. American Journal of Emergency Medicine Forecasting emergency department hourly occupancy using time series analysis. Am J Emerg Med [Internet]. 2021;48:177–82. Available from: https://doi.org/10.1016/j.ajem.2021.04.075

13. Whitt W, Zhang X. Forecasting arrivals and occupancy levels in an emergency department. Oper Res Heal Care [Internet]. 2019;21:1–18. Available from: https://doi.org/10.1016/j.orhc.2019.01.002

14. Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications. Comput Biol Med [Internet]. 2019;112(July):103375. Available from: https://doi.org/10.1016/j.compbiomed.2019.103375

15. Jiang S, Chin KS, Tsui KL. A universal deep learning approach for modeling the flow of patients under different severities. Comput Methods Programs

Biomed [Internet]. 2018;154:191–203. Available from: https://doi.org/10.1016/j.cmpb.2017.11.003

16. Finnish Meteorological Institute Open Weather Data [Internet]. [cited 2020 Feb 2]. Available from: https://www.ilmatieteenlaitos.fi/avoin-data

17. University of Helsinki Almanac Office [Internet]. [cited 2020 Jul 20]. Available from: https://almanakka.helsinki.fi/en/

18. Google Trends [Internet]. [cited 2020 Jun 7]. Available from: https://www.google.com/trends

19. Gul M, Celik E. An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments. Heal Syst [Internet]. 2018;00(00):1–22. Available from: https://doi.org/10.1080/20476965.2018.1547348

20. Makridakis S, Spiliotis E, Assimakopoulos V. The M4 Competition: 100,000 time series and 61 forecasting methods. Int J Forecast [Internet]. 2020;36(1):54–74. Available from: https://doi.org/10.1016/j.ijforecast.2019.04.014

21. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice [Internet]. OTexts: Melbourne, Australia; 2021. Available from: https://otexts.com/fpp3/

22. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for {R}. J Stat Softw [Internet]. 2008;26(3):1–22. Available from: https://www.jstatsoft.org/article/view/v027i03

23. Smith TG, others. {pmdarima}: ARIMA estimators for {Python} [Internet]. Available from: http://www.alkaline-ml.com/pmdarima

24. Breiman L. Random Forests. Mach Learn [Internet]. 2001;45(1):5–32. Available from: https://doi.org/10.1023/A:1010933404324

25. Haykin S, Haykin SS. Adaptive Filter Theory [Internet]. Pearson; 2014. Available from: https://books.google.ae/books?id=J4GRKQEACAAJ

26. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. Vol. 13, PLoS ONE. 2018. 1–42 p.

27. Lim B, Arık S, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. arXiv. 2019;(Bryan Lim):1–27.

Supplementary Files

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

- additionalfile1.xlsx