Predicting Patient Arrival Rates in a Multi-Specialty Outpatient Department

Hana Ben Mbarek, Safa Bhar Layeb, Najla Omrane Aissaoui, Amel Jaoua
LR-OASIS, National Engineering School of Tunis,
University of Tunis El Manar, Tunis, Tunisia
hana.benmbarek@enit.utm.tn, safa.layeb@enit.utm.tn, najla.aissaoui@enicarthage.rnu.tn, amel.jaoua@enit.utm.tn

Atidel B. Hadj-Alouane
Department of Industrial Engineering and Engineering Management, College of Engineering,
University of Sharjah, Sharjah, UAE
aalouane@sharjah.ac.ae

Abstract

Although the medical interventions provided by Outpatient Departments (OPDs) are scheduled, the process of patient arrival is often unpredictable. This can lead to a variety of issues such as overcrowding, long waiting times, and patient dissatisfaction. Therefore, predicting the patient arrival rate is necessary to allow healthcare managers to be more proactive and make more efficient and realistic decisions. In this context, this real-life study aims to predict the hourly rate of patient arrival in a multi-specialty outpatient department using Machine Learning (ML) algorithms. The considered OPD regroups three specialties: pulmonology, allergology, and cardiology. One-year patient arrival data for all specialties were extracted and preprocessed. Thereafter, the exhaustive feature selection was applied to identify the best subset of features. Next, seven ML algorithms were investigated and assessed to determine the best predictive model. The conducted experimental study revealed that support vector regression shows the best performance for the prediction of pulmonary patient admissions, while random forest was identified as the best predictive model for the other specialties.

Keywords
Prediction, Machine Learning, Outpatient Department, Patient Arrival Rate.

1. Introduction
Healthcare planning in Outpatient Departments (OPDs) is highly challenging (Vermeulen et al. 2009). Indeed, the medical structures in these departments are constantly affected by their environments and are subject to a high degree of uncertainty due to the complexity of the clinical pathway and the fluctuation of patient demand, as well as the unpredictability of their arrival (Aissaoui et al. 2022).It is well known that outpatient arrival processes, which decide independently when they arrive, differ from arrival processes generated by appointment schedules (Kim et al. 2018). In fact, the appointment generated arrival processes theoretically have a quasi-periodic structure defined by time slots of appointments. Nevertheless, it is commonly reported in the literature that actual arrival processes can vary significantly due to several factors that lead to schedule disruption (Cayirli and Veral 2003). In fact, several studies have shown that patient non-punctuality causes extensive trouble for health care managers and reveal that patients often arrive earlier rather than later (Klassen and Rohleder 1996; Lehaney et al. 1999), resulting in overcrowding (Welch and Bailey 1952). Similarly, the patient no-shows are a worldwide concern that increase healthcare costs (Incze et al. 2021) and affect the performance of appointment systems (Ho and Lau 1992). Extant studies have suggested that the observed probabilities of missed outpatient appointments vary by department specialty (Aebi et al. 2007; Mohamed et al. 2016). In addition, the walk-ins, or patients who come to OPD without an appointment, constitute another patient arrival pattern. Patients who require urgent medical care and can potentially preempt the current patient visit are a special type of the walk-ins. Like the no-shows, the probability of walk-ins varies by specialty (Field 1980) and by hospital type and location (Munavalli et al. 2021). Furthermore, Swartzman (1970) have emphasized that walk-ins and emergency patients' arrival rates vary significantly during the

day. It is noteworthy that patient demand in the OPDs often differs across the day and between days and may have seasonal patterns.

Given the challenges posed by patients with no-shows, unpunctuality, walk-ins, and fluctuating demand, predicting patient arrival in public health services is critical to enhance managerial decision making. Machine Learning (ML) methods represent one of the best tools for predicting target variables in several fields (Ouerghi et al. 2019; Mejri et al. 2021) and precisely in healthcare to predict patient arrival (Nas and Koyuncu 2019). These methods allow healthcare entities to gain insights into their future workload in order to make better decisions (Pianykh et al. 2020).In the same vein, this study was conducted in a multi-specialty OPD of a Tunisian public hospital specialized in pneumo-phthisiology, in order to predict patient arrivals in this department for better decision-making regarding care planning. The considered OPD covers three specialties, namely pulmonology, allergology and cardiology.

The main objective of this work is to predict the hourly arrival rate of patients of the three specialties of the OPD considering that analysis and prediction of daily rather than hourly data would lead to the loss of significant information related to patient arrival variations. In addition, it should be noted that some studies have proceeded to categorize patients and then eliminate some based on their constraints (Nas and Koyuncu, 2019). Since the total number of hourly arrivals is a key input for decision making tools, following through with this process can lead to inaccurate results. Therefore, all the OPD arrivals for different patient specialties are included in this study. To the best of our knowledge, this is the first study that applies ML algorithms to predict the hourly patient arrival rate in a multi-specialty OPD.

2. Literature Review

In the medical sector, predictive models aim to enhance the quality of patient care and increase logistical efficiency (Obermeyer and Emanuel 2016).In fact, an effective early demand prediction can provide healthcare managers, including those in outpatient departments, with essential information for scheduling and resource allocation (Jiang et al. 2017).Given the importance of forecasting demand and patient arrivals in the healthcare system to maintain its performance, several prediction techniques have been developed. These techniques can be broadly categorized into qualitative and quantitative methods. Qualitative methods predict future events based on the opinions and judgments of experts (Pope et al. 2002). In contrast, quantitative methods refer to mathematical models (Ozcan 2005). Techniques in this category are mainly based on time series analysis, which consists in exploiting historical data to build a mathematical model that can be applied to discover the future behavior of a series of observations. It is worth mentioning that this category of methods is characterized by its simplicity and effectiveness and is best suited for practical applications (Kadri et al. 2014). As noted by Klute et al. (2019), classical time series predictive methods have been commonly used for healthcare demand forecasting. Linear regression (Jones et al. 2008; Boyle et al. 2012; Marcilio et al. 2013; Xu et al. 2013), autoregressive integrated moving average (Sun et al. 2009; Xu et al. 2011; Calegari et al 2016; Carvalho-Silva et al. 2018) and exponential smoothing (Jones et al. 2008; Boyle et al. 2012; Calegari et al. 2016) were used as prediction models for patient arrivals.

However, according to Vollmer et al. (2021), methods for predicting admission rates while understanding the underlying dynamics have not been investigated extensively in the literature. Therefore, ML-based approaches are the next evolution of prediction in healthcare (Klute et al., 2019). Precisely, ML techniques can offer many advantages for time series forecasting, including automatic learning of time dependence and automatic processing of temporal structures (Brownlee 2018). Indeed, ML includes multiple mathematical models and algorithms for the learning process. Increasing attention has recently been given to Machine Learning for predicting patient arrivals at healthcare facilities. Hong et al. (2018) applied several ML algorithms to predict admissions at triage in emergency departments. They trained a set of binary classifiers using Linear Regression (LR), Gradient Boosting Machines (GBM), and Deep Neural Networks (DNN). The authors revealed that GBM and DNN outperformed LR whereas the performance of GBM and DNN was without significant difference. Yousefi et al. (2019) studied the factors affecting daily demand in an emergency department and applied DNN to build a reliable predictive model for horizons up to seven days. The authors claimed that better predictions are viable using the capabilities of ML algorithms. The numerical results showed that the proposed method outperforms statistical alternatives such as multiple linear regressions, Autoregressive Integrated Moving Average (ARIMA), generalized linear models, seasonal ARIMA and ARIMA combined with linear regression. Yucesan et al. (2018) investigated the performance of single and hybrid methods for forecasting patient arrival in an emergency department. The methods include LR, exponential smoothing, ARIMA, Artificial Neural Network (ANN), and hybrid ARIMA models, namely ARIMA-LR and ARIMA-ANN. The authors showed that the hybrid ARIMA-ANN model is the most accurate forecasting method. Klute et al (2019) examined appointment requests in an outpatient department. The authors applied twenty state-of-the-art, ML and hybrid methods to determine the best prediction outcomes. The results show that the GBM model showed the best performance. Recently, Srinivas (2020) predicted patient-specific late arrival risk using ML models. Data from two different ambulatory care centers were extracted, and a comprehensive range of predictor variables was determined. Four ML algorithms (logistic regression, GBM, random forests, and ANN) were tested. Numerical results revealed that ML algorithms can accurately predict patient tardiness. It is worth mentioning that patient arrival issues are generally related daily (Schiavo et al. 2015; Xu et al. 2013), weekly (Carvalho-Silva et al. 2018) or monthly (Carvalho-Silva et al. 2018; Bergs et al. 2014). However, few studies consider hourly arrival rates, which exhibit high variability (Boyle et al. 2012; Nas and Koyuncu 2019). Moreover, no existing research has investigated the prediction of the hourly patient arrival rate in a multi-specialty OPD, using ML algorithms.

3. Methods

In order to predict the hourly patient arrival rate at the considered OPD, four main steps are undertaken as follows: (i) collecting patient arrival data, (ii) pre-processing the collected data, (iii) designing experiments for training the ML models, and (iv) evaluating the predictive models. In the following, we will detail each step of the applied methodology.

Step 1: Data collection

This step consists in gathering all necessary data for forecasting purposes using ML algorithms. These data concern the response variable, which in our case is the rate of the patient arrival at the OPD, and the explanatory variables (independent variables).

Step 2: Data preprocessing

Since the prediction model's ability to learn relies directly on the quality of the data and the meaningful insights that can be extracted from it, data preprocessing is crucial before building any predictive model. Data preprocessing makes raw data more accurate and suitable for predictive model building and training. It refers to the techniques that follow.

- *Data cleaning*. Real-world data is often inconsistent, incomplete, and inaccurate due to several reasons such as program errors and human factors. In this regard, the collected data should be cleaned to get simple, clear, and complete sample sets.

- *Data transformation*. It is basically a feature encoding process that consists in converting data into an appropriate input format so that the computer can derive insights from it without altering its original meaning. The main basis of both traditional predictive models and machine learning based algorithms are mathematical equations. Thus, it is necessary to encode categorical data, which are records of information having specific categories represented by discrete variables, prior to training models. Moreover, feature scaling constitutes another important data pre-processing. It refers to a method of normalizing variables in a dataset to a particular range, avoiding the construction of incorrect prediction models during training and performing data analysis.

- *Feature selection*. It is a common aspect of predictive model optimization. Precisely, it is the process of selecting the best predictors of the target variable, allowing to decrease overfitting and training time and to improve the model accuracy.

Step 3: ML algorithms implementation

ML algorithms enable the learning of a target function f that best relates input variables X to an output variable Y such as $Y = f(X)$. They require splitting the data into learning and testing datasets to avoid data leakage. The training dataset is used to establish the parameters of a given model, the so-called *model training*, and generally holds 80% of data while the remaining 20% is reserved for testing purposes (Incze et al. 2021). The test set includes data not previously considered by the model and is useful to ensure that the findings are meaningful and reproducible. In addition, to overcome the risk of over-fitting or noise learning, a k-fold cross-validation can be applied in the training phase. Accordingly, the training dataset is split into k subsets, with one subset being used for validation and the others for model learning. The procedure is then repeated until each subset is used precisely once for validation (Srinivas 2020). In our study, to ensure a good prediction of patient arrival rates, multiple ML algorithms were applied. Precisely, we focus on the following seven ML techniques:

- *Linear Regression (LR)* is one of the simplest and most widely used ML algorithms (Maulud and Abdulazeez 2020). Basically, the LR is a mathematical approach applied to undertake prediction analysis and used to find the most suitable linear relationship between the independent and the dependent variables.

- *K-Nearest Neighbors (KNN)* is well-known to be one of the basic prediction models that memorizes the training dataset and makes the prediction when the attributes of the test object exactly match one of the examples from the training dataset (Yu et al. 2002). This algorithm can make local and different estimates of complex target concepts for every newly predicted instance. Although the KNN had shown good performance in several fields (Wu et al. 2008), it needs to store all the data, which slows down its performance with large datasets, since all the training data should be browsed. Its accuracy also decreases as noise and irrelevant attributes increase throughout the training dataset.

- *Support Vector Regression (SVR)* is based on the same principle as the support vector machines algorithm, one of the most popular algorithms for large-scale prediction (Yu et al. 2002), but for regression problems. The SVR is like linear regression in that the usual straight line is herein called a hyperplane. Those points of data on either side of the hyperplane that are closest to this line are referred to as support vectors, which are used to draw the boundary line. Unlike other regression models that try to reduce the error between the actual value and the predicted one, the SVR tries to find the best fit line within a threshold value: the distance from the hyperplane to the boundary line.

- *Decision Tree (DT)* consists in subdividing a given dataset into progressively smaller subsets, and then incrementally constructing an associated decision tree (Mahesh 2020). The resulting model forms a tree structure having decision nodes and leaf nodes. A decision node has two or more branches, each representing values for the attribute being tested. The leaf node represents a decision on the value of the target.

- *Random Forest (RF)* is a ML algorithm based on a set with many trees. In RF, the performance of several weak learners, DT in this case, is strengthened by a voting system. The main properties of RF are random feature selection, bootstrap sampling, out-of-bag error estimation, and deep decision tree growth (Jiang et al. 2009).

- *Gradient Boosting Machines (GBM)*. The term "Gradient boosting" refers to the predictive model being trained iteratively from a set of weaker models, DT in this case, powered by a gradient descent function. Boosting is performed by training each weak model and then evaluating its performance, with inputs failed to be predicted correctly by previous models weighted more heavily and successfully predicted inputs weighted less heavily (Incze et al. 2021). The sample distribution is modified in this process at each iteration, allowing subsequent models to focus on poorly predicted data.

- *Multilayer Perceptron (MLP)* is a class of artificial neural networks, which are mathematical models that imitate the human brain's learning process in collecting and processing data. Artificial neural networks could find complex input-output relationships without the need for assumptions of linearity, normality, or independent variables (Golmohammadi 2016).

It is worthy to note that tree set-based techniques, including GBM and RF, have often outperformed other ML algorithms in the existing literature (Wainer 2016).

Step 4: ML algorithms evaluation

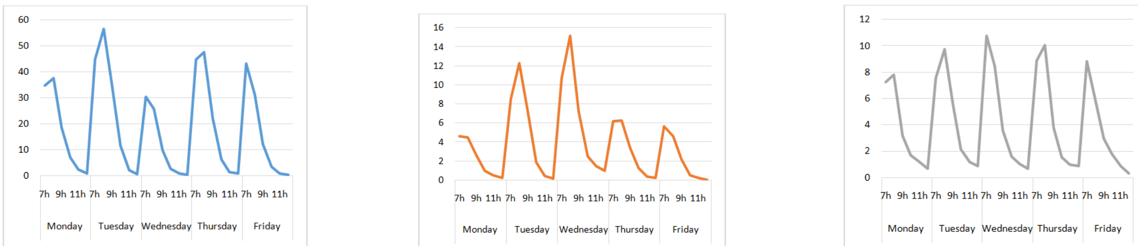
Following the training phase, the established ML algorithms should be assessed using the same evaluation metric. In this study, we used the Mean Absolute Error (MAE), which is a useful and widely adopted measure in ML models evaluation (Chai and Draxler 2014). Precisely, having a set of n predicted values $(p_1, p_2, \dots, p_n)$ and the associated set of real values $(x_1, x_2, \dots, x_n)$, the MAE is computed as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} abs\ (x_i - p_i). \qquad (1)$$

The predictive performance of all evaluated algorithms should then be compared and the best ML model for predicting patient arrival rates should be selected.

5. Results and Discussion

It is important to highlight the fact that the patient arrival rate at the OPD is characterized by a high degree of uncertainty and varies depending on the hour and the day of the week, as well as over the year. Indeed, Figure 1 shows that most patients for the three specialties arrive between 7am and 9am, being the most crowded time. Furthermore, the least crowded time is between 11am and 1pm.



(a) (b) (c)

Figure 1. Patient arrival patterns by hour for (a) pulmonology, (b) allergology and (c) cardiology

Furthermore, Figure 2 displays that almost 50% of the patients visit the OPD on Tuesdays and Thursdays. The patient arrival rate by month is shown in Figure 3. Not surprisingly, the number of arrivals of allergology patients increases in the winter. However, it decreases by almost 36% in the summer. For pulmonology patients, the number of arrivals is high in spring and autumn seasons and decreases by almost 10% in the summer compared to that in autumn. For the cardiology specialty, there is an increase in the patient arrival rate in winter and spring and a decrease in summer and autumn.

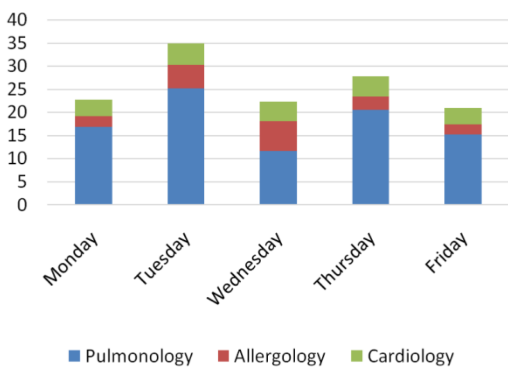


Figure 2. Average hourly arrival rate of patients at the OPD per day

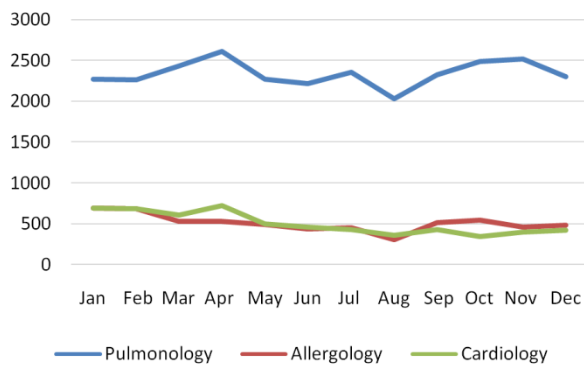


Figure 3. Patient arrival rate per month

Once the variation in demand for the three OPD specialties are broadly scanned, the 4 steps of the methodology described in Section 2 are followed for predicting the patient arrival rate. All subsequent processing was carried out using Anaconda Python.

Step 1: Data collection. Initially, the required data was collected. Precisely, the historical data for this research was obtained from the hospital's information system. For each specialty, a full year of patient arrival data is extracted between January 1, 2019, and December 31, 2019. During this period, the hospital had a total of 40,716 patient visits. The primary response variable in this study is the hourly patient arrival rate. We investigated the arrival date for every study hour by month, season, weekday, week of the year, and hour of the day, constituting the explanatory variables.

Step 2: Data preprocessing. The following data preprocessing were performed:
- Data cleaning. Approximately 5.32% of the records had invalid values and are completely removed. This resulted in a final count of 38,550 observations, of which 72.39% were pulmonology patients, 15.25% were allergology patients, and 12.36% were cardiology patients.

- Data transformation. In the set of data mentioned above, there are three categorical variables (month, season and weekday) that are encoded in numerical variables. Thereafter, the min-max scaling technique was used to shift and rescale the values in the dataset so that they range from 0 and 1.

- Feature selection. For the most accurate predictions, we had to identify the best subset of explanatory variables. The correlation coefficients between all variables were calculated. The results showed that the variables that are the most correlated with the number of patients for the three specialties are the weekday and the hour of the day. Nevertheless, as Nas and Koyuncu (2019) point out, the combination of the explanatory variables can lead to very different findings. Thus, to determine the optimal subset of the features, the exhaustive feature selection method was used. This method selects the optimal subset that reduces the loss function using any ML technique and analyzes all combinations of the features. In the present study, we applied the Random Forest method to assess the contribution of the explanatory variables and used the mean absolute error as the loss function. We determined that the best subset for pulmonology contains weekday and hour. For allergology and cardiology, month, weekday, hour, and week of the year appear to be the best predictors.

Step 3:ML algorithms implementation. First, the preprocessed data is split into two subsets: a training set containing 80% of the data and a test set containing 20% of the data. For the implementation of the 7 ML models described is Section 4, we started with the training phase of the algorithms using the training data. Their parameters were optimized using the grid search method with cross validation. Accordingly, the ML algorithm tests several possible combinations of a finite number of parameter values using cross-validation (Srinivas and Ravindran2018). The best performing parameter value, relative to the MAE, is used to predict the patient arrival rates.

Step 4:ML algorithms evaluation. An evaluation of the prediction performance of the 7 ML algorithms is carried out using the test data. The obtained results are displayed in Table 1.

Table 1. MAE values of the 7 ML algorithms for the 3OPD specialties

ML algorithm	**OPD Specialty**		
	Pulmonology	**Allergology**	**Cardiology**
LR	10.3595	2.6182	2.4871
KNN	4.7914	1.4786	1.7765
SVR	*4.6414*	1.4291	1.6981
DT	4.7024	1.4587	1.6395
RF	4.7040	*1.3994*	*1.6386*
GBM	4.6962	1.4726	1.6596
MLP	4.7576	1.5197	1.8543

As presented in Table 1, the best predictive models are the SVR for the pulmonology specialty and the RF for the specialties of allergology and cardiology. Not surprisingly, the LR lags far behind all the other algorithms. However, the findings clearly show that there is no ML algorithm that performs better than the others for all the specialties. At this point, it is worth noting that many factors can affect the performance of predictive models, including the structure and size of the dataset (Nas and Koyuncu 2019). This ties in with the interest of trying several ML-based prediction techniques, while using a test data set to compare the performance of the models and identify the best one.

6. Conclusion

In this study, machine learning tools were applied to predict the hourly patient arrival rates. A one-year patient arrival data was investigated and processed for three specialties of an outpatient department of a Tunisian hospital. To obtain the most accurate predictions, data extracted from the hospital information system regarding patient admissions to the concerned OPD was processed. Then, using the exhaustive feature selection method, the best subset of the explanatory variables that gives the most reliable predictions was determined. Seven ML algorithms were developed to predict patient arrival rates. Support vector regression was found to be the best performing for the pulmonology specialty, while random forest was identified as the best predictor for the specialties of allergology and cardiology.

Future research will focus on integrating the predictions of the best performing ML models into a demand-driven resource allocation approach that would help healthcare managers to be proactive and more effective in their decision-making.

References

Aebi, A., Lehmann, D., Olivet, M. B., and Stalder, H., Missed appointments at a Swiss university outpatient clinic, *public health*, vol. 121, no. 10, pp. 790-799, 2007.

Aissaoui, N. O., Ben Mbarek, H., Layeb, S. B., and Hadj-Alouane A. B., A BPMN-VSM based process analysis to improve the efficiency of multidisciplinary outpatient clinics, *Production Planning & Control*, forthcoming, 2022.

Bergs, J., Heerinckx, P., and Verelst, S., Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis, *international emergency nursing*, vol. 22, no. 2, pp. 112-115, 2014.

Boyle, J., Jessup, M., Crilly, J., Green, D., Lind, J., Wallis, M., Miller, P., and Fitzgerald, G., Predicting emergency department admissions, *Emergency Medicine Journal*, vol. 29, no. 5, pp. 358-365, 2012.

Brownlee, J., Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python,*Machine Learning Mastery*, 2018.

Calegari, R., Fogliatto, F. S., Lucini, F. R., Neyeloff, J., Kuchenbecker, R. S., and Schaan, B. D., Forecasting daily volume and acuity of patients in the emergency department. *Computational and mathematical methods in medicine*, 2016.

Carvalho-Silva, M., Monteiro, M. T. T., de Sá-Soares, F., and Dória-Nóbrega, S., Assessment of forecasting models for patientsarrival at emergency department, *Operations Research for Health Care*, vol. 18, pp. 112-118, 2018.

Cayirli, T., and Veral, E., Outpatient scheduling in health care: a review of literature, *Production and operations management*, vol. 12, no. 4, pp. 519-549, 2003.

Chai, T., and Draxler, R. R., Root mean square error (RMSE) or mean absolute error (MAE), *Geoscientific Model Development Discussions*, vol. 7, no. 1, pp. 1525-1534, 2014.

Field, J., Problems of urgent consultations within an appointment system, *The Journal of the Royal College of General Practitioners*, vol. 30, no. 212, pp. 173-177, 1980.

Golmohammadi, D., Predicting hospital admissions to reduce emergency department boarding,*International Journal of Production Economics*, vol. 182, pp.535-544, 2016.

Han, J., Pei, J., and Kamber, M., Data mining: concepts and techniques, Elsevier, 2011.

Ho, C. J., and Lau, H. S., Minimizing total cost in scheduling outpatient appointments, *Management science*, vol. 38, no. 12, pp. 1750-1764, 1992.

Hong, W. S., Haimovich, A. D., and Taylor, R. A., Predicting hospital admission at emergency department triage using machine learning, *PloS one*, vol. 13, no. 7, pp. e0201016, 2018.

Incze, E., Holborn, P., Higgs, G., and Ware, A., Using machine learning tools to investigate factors associated with trends in 'no-shows' in outpatient appointments, *Health & Place*, vol. 67, pp. 102496, 2021.

Jiang, R., Tang, W., Wu, X., and Fu, W., A random forest approach to the detection of epistatic interactions in case-control studies, *BMC bioinformatics*, vol. 10, no. 1, pp. 1-12, 2009.

Jiang, S., Chin, K. S., Wang, L., Qu, G., and Tsui, K. L., Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department, *Expert systems with applications*, vol. 82, pp. 216-230, 2017.

Jones, S. S., Thomas, A., Evans, R. S., Welch, S. J., Haug, P. J., and Snow, G. L., Forecasting daily patient volumes in the emergency department, *Academic Emergency Medicine*, vol. 15, no. 2, pp. 159-170, 2008.

Kadri, F., Harrou, F., Chaabane, S., and Tahon, C., Time series modelling and forecasting of emergency department overcrowding, *Journal of medical systems*, vol. 38, no. 9, pp. 1-20, 2014.

Kim, S. H., Whitt, W., and Cha, W. C., A data-driven model of an appointment-generated arrival process at an outpatient clinic, *INFORMS Journal on Computing*, vol. 30, no. 1, pp. 181-199,2018.

Klassen, K. J., and Rohleder, T. R., Scheduling outpatient appointments in a dynamic environment,*Journal of operations Management*, vol. 14, no. 2, pp. 83-101, 1996.

Klute, B., Homb, A., Chen, W., and Stelpflug, A., Predicting outpatient appointment demand using machine learning and traditional methods, *Journal of medical systems*, vol. 43, no. 9, pp. 1-10, 2019.

Lehaney, B., Clarke, S. A., and Paul, R. J., A case of an intervention in an outpatient's department, *Journal of the Operational Research Society*, vol. 50, no. 9, pp. 877-891, 1999.

Mahesh, B., Machine learning algorithms-a review, *International Journal of Science and Research*, vol. 9, pp. 381-386, 2020.

Marcilio, I., Hajat, S., and Gouveia, N., Forecasting daily emergency department visits using calendar variables and ambient temperature readings, *Academic emergency medicine*, vol. 20, no. 8, pp.769-777, 2013.

Maulud, D., and Abdulazeez, A. M., A review on linear regression comprehensive in machine learning, *Journal of Applied Science and Technology Trends*, vol. 1, no. 4, pp. 140-147,2020.

Mejri, I., Bouzid, A., Bacha, S., and Layeb, S. B., Forecasting Demand Using ARIMA Model and LSTM Neural Network: a Case of Detergent Manufacturing Industry. *In 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies*, pp. 346-353, IEEE, September 2021.

Mohamed, K., Mustafa, A., Tahtamouni, S., Taha, E., and Hassan, R., A quality improvement project to reduce the 'No Show'rate in a paediatric neurology clinic, *BMJ Open Quality*, vol. 5, no. 1, pp. u209266-w3789, 2016.

Munavalli, J. R., Boersma, H. J., Rao, S. V., and Van Merode, G. G., Real-time capacity management and patient flow optimization in hospitals using AI methods, *In Artificial intelligence and Data mining in healthcare*, pp. 55-69, Springer, Cham, 2021.

Nas, S., and Koyuncu, M., Emergency department capacity planning: A recurrent neural network and simulation approach, *Computational and mathematical methods in medicine,* 2019.

Obermeyer, Z., and Emanuel, E. J., Predicting the Future-Big Data. Machine Learning, and Clinical Medicine, The *New England Journal of Medecine*, vol. 375, no. 13, pp. 1216-1219, 2016.

Ouerghi, A., Hasni, M., Jaidi, Z., and Layeb, S. B., A new combined linear-artificial neural network-based model for accurate inflation forecasting in Tunisia. *International Journal of Decision Sciences, Risk and Management*, vol. 8, no. 4, pp. 220-233, 2019.

Ozcan, Y. A., Quantitative methods in health care management: techniques and applications,*John Wiley & Sons*, 2005.

Pianykh, O. S., Guitron, S., Parke, D., Zhang, C., Pandharipande, P., Brink, J., and Rosenthal, D., Improving healthcare operations management with machine learning, *Nature Machine Intelligence*, vol. 2, no. 5, pp. 266-273, 2020.

Pope, C., Van Royen, P., and Baker, R., Qualitative methods in research on healthcare quality, *BMJ Quality & Safety*, vol. 11, no. 2, pp. 148-152, 2002.

Schiavo, M. L., Prinari, B., Gronski, J. A., and Serio, A. V., An artificial neural network approach for modeling the ward atmosphere in a medical unit, *Mathematics and Computers in Simulation*, vol. 116, pp. 44-58, 2015.

Srinivas, S., A machine learning-based approach for predicting patient punctuality in ambulatory care centers, *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, pp. 3703,2020.

Srinivas, S., and Ravindran, A. R. Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: a prescriptive analytics framework, *Expert Systems with Applications*, vol. 102, pp. 245-261, 2018.

Sun, Y., Heng, B. H., Seow, Y. T., and Seow, E., Forecasting daily attendances at an emergency department to aid resource planning, *BMC emergency medicine*, vol. 9, no. 1, pp. 1-9, 2009.

Swartzman, G., The patient arrival process in hospitals: statistical analysis, *Health services research*, vol. 5, no. 4, pp. 320, 1970.

Vermeulen, I. B., Bohte, S. M., Elkhuizen, S. G., Lameris, H., Bakker, P. J., and La Poutré, H., Adaptive resource allocation for efficient patient scheduling, *Artificial intelligence in medicine*, vol. 46, no.1, pp. 67-80,2009.

Vollmer, M. A., Glampson, B., Mellan, T., Mishra, S., Mercuri, L., Costello, C., Klaber, R., Cooke, G., Flaxman, S., and Bhatt, S., A unified machine learning approach to time series forecasting applied to demand at emergency departments, *BMC Emergency Medicine*, vol. 21, no. 1, pp. 1-14, 2021.

Wainer, J., Comparison of 14 different families of classification algorithms on 115 binary datasets, arXiv preprint arXiv, pp. 1606.00930,2016.

Welch, J. D., and Bailey, N. J., Appointment systems in hospital outpatient departments, *The lancet*, vol. 259, no. 6718, pp. 1105-1108, 1952.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J.,Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., and Steinberg, D., Top 10 algorithms in data mining, *Knowledge and information systems*, vol. 14, no. 1, pp. 1-37, 2008.

Xu, M., Wong, T. C., and Chin, K. S., Modeling daily patient arrivals at Emergency Department and quantifying the relative importance of contributing variables using artificial neural network, *Decision Support Systems*, vol. 54, no. 3, pp. 1488-1498, 2013.

Xu, M., Wong, T. C., Chin, K. S., Wong, S. Y., and Tsui, K. L., Modeling patient visits to accident and emergency department in hongkong, *In 2011 IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 1730-1734, IEEE, December 2011.

Yousefi, M., Yousefi, M., Fathi, M., and Fogliatto, F. S., Patient visit forecasting in an emergency department using a deep neural network approach, *Kybernetes*, 2019.

Yucesan, M., Gul, M., and Celik, E., A multi-method patient arrival forecasting outline for hospital emergency departments, *International Journal of Healthcare Management*, 2018.

Yu, K., Ji, L., and Zhang, X., Kernel nearest-neighbor algorithm, *Neural Processing Letters*, vol. 15, no. 2, pp. 147-156, 2002.

Biographies

Hana Ben Mbarek is an Industrial Engineer, a research associate and PhD candidate at the OASIS Lab at the National Engineering School of Tunis, Tunisia. Her research interests revolve around industrial engineering, lean healthcare and simulation.

Safa Bhar Layeb is an associate professor of industrial engineering and a member of the OASIS Lab at the National Engineering School of Tunis, Tunisia. She is the founding chair of the African Working Group in Health Systems, affiliated with the African Federation of Operational Research Societies (AFROS). She is particularly interested in industrial engineering approaches and their applications in healthcare organizations.

Najla Omrane Aissaoui received her Industrial Engineering degree, her MS and Ph.D. from the National Engineering School of Tunis (ENIT). She is currently an assistant professor of industrial engineering at the National Engineering School of Carthage and a member of OASIS Lab at the National Engineering School of Tunis, Tunisia. Her area of expertise and interest are production planning and control, supply chain optimization, lean healthcare, health systems and performance measurement systems.

Amel Jaoua is an assistant professor of industrial engineering and a member of the OASIS Lab at the National Engineering School of Tunis, Tunisia. She is particularly interested in simulation and data science.

Atidel B. Hadj-Alouane joined the University of Sharjah in Fall 2021 as Professor of Industrial Engineering and Engineering Management. She has over 25 years of academic experience. At the National Engineering School of Tunis, she founded and directed for 15 years, the Optimisation of Service and Industrial Systems laboratory (OASIS) and cofounded several programs such as the PhD and master's programs in Industrial Engineering and the Professional Masters' program in Business Development and Modernisation. Her research interests include integrated logistics systems and green logistics, Risk management in supply chains, Smart scheduling, and OR application to Telecommunications, Cloud systems and infrastructure, and Health systems.