Omega 127 (2024) 103098

Contents lists available at ScienceDirect

Omega

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

On-line strategy selection for reducing overcrowding in an Emergency Department[☆]

Cristiano Fabbri [a,c,1], Michele Lombardi [b,1], Enrico Malaguti [c,1,*], Michele Monaci [c,1]

[a] Enterprise information systems for integrated care and research data management, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy
[b] Department of Computer Science and Engineering - DISI, University of Bologna, Bologna, Italy
[c] Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" - DEI, University of Bologna, Bologna, Italy

ARTICLE INFO

Keywords:
Decision Support System
Deep Neural Network
Natural Language Processing
Simulation
Emergency Department

ABSTRACT

Overcrowding is a well-known major issue affecting the behavior of an Emergency Department (ED), as it is responsible for patients' dissatisfaction and has a negative impact on the quality of workers' performance. Dealing with overcrowding in an ED is complicated by lack of its precise definition and by exogenous and stochastic nature of requests to be served. In this paper, we present a Decision Support System (DSS) based on the integration of a Deep Neural Network for dealing with the sources of uncertainty and a simulation tool to evaluate how specific management policies affect the ED behavior. The DSS is designed to be run on-line, dynamically suggesting the most suitable policy to be implemented in the ED. We evaluate the performance of the DSS on a specific major ED located in northern Italy. Numerical results show that overcrowding can be considerably reduced by allowing a dynamic selection among a limited set of simple policies for queue management.

1. Introduction

The Emergency Department (ED) is a medical facility dedicated to receiving and treating patients with unexpected illness and injury within a short period of time. It works 24 h a day, for a variable number of week days, usually 5 or 7, depending on the number of accesses. Due to its own nature, activities performed within an ED are unprojectable. Patients arrive either by their own or with an ambulance, and claim different types of treatment for a wide variety of diseases.

Due to the complexity of this environment, the admission of patients is handled according to a priority-based policy [1]. As a consequence, the first activity performed during an ED pathway is to determine the patient's priority with a process called *triage*. ED triage is aimed at defining the urgency of treatment while taking into account scarcity of the resources [2]. Triage activities are coded at the regional or national level based on a scale, and assign each patient an urgency of treatment according to a specific scale, the most common ones being the Australasian Triage Scale, the Canadian Triage and Acuity Scale, the Manchester Triage System, and the Emergency Severity Index (see, [3]). These activities are usually performed by a nurse, who takes notes about the patient's health and personal condition and assigns a priority code. In addition to structured data (e.g., sex, age, oxygen saturation, etc.), the nurse may also fill-in a diary with more detailed information, such as the circumstances of accident or others elements that can be used by the physician during decision-making process. Data collected during the triage process are useful for the whole ED pathway; as shown in [4], triage nurses are capable of assessing the patient's complexity in a reliable and valid way. In addition, analytics or AI-based techniques can be used to effectively support triage decision making, see, e.g., [5].

Even in case the triage correctly assigns the level of care to each patient, the performance of an ED may be affected by overcrowding, arising when the demand for ED services exceeds the available resources. Overcrowding may have a negative impact on different operational aspects, such as waiting times, length of stay (LoS), increasing number of patients leaving without being seen (LWBS), which can increase medical errors and decrease efficiency [6,7]. Overcrowding is a complex phenomenon for which there exists no universally accepted definition and measure. The most common way to quantify ED overcrowding is the so-called National ED OverCrowding Study (NEDOCS) indicator, proposed in [8]. This is a one-dimensional indicator that,

☆ Area: Production Management, Scheduling and Logistics. This manuscript was processed by Associate Editor Otto.

* Corresponding author.
E-mail addresses: cristiano.fabbri@aosp.bo.it (C. Fabbri), michele.lombardi2@unibo.it (M. Lombardi), enrico.malaguti@unibo.it (E. Malaguti), michele.monaci@unibo.it (M. Monaci).

[1] Equally contributed to this paper.

https://doi.org/10.1016/j.omega.2024.103098
Received 13 July 2023; Accepted 9 April 2024
Available online 19 April 2024

based on the available resources (e.g., ED beds, Hospital beds) and on the ED state (e.g., total patients simultaneously present in the ED), returns the ED overcrowding score (from "not busy" to "Dangerously Overcrowded"). Although NEDOCS is not suitable in some cases [9], it may be a useful indicator for detecting areas where efforts have to be put for addressing congestion (e.g., total admits in the ED, total patients simultaneously present in the ED, number of respirators, longest admit time, etc.). More generally, different Operations Research and Operations Management approaches have been proposed to deal with overcrowding. A first class of actions affects the ED intake process [10], including techniques for allocating ambulances within a network of EDs [11,12], or for rerouting ambulances to others hospitals in periods of crowding (Ambulance Diversion, see [13]). A second possibility is to focus on the internal ED patient flow, in order to use available resources efficiently. Although many attempts for creating decision support system tools have been proposed in recent years [14], their application within an ED environment is challenging. Indeed, forecasting patient pathways can be hard [15] for different reasons, as large number of pathways variants or missing information. Nevertheless, mining the patients' pathway is a key issue for improving the internal flow of patients in the ED, by correctly identifying bottlenecks and waiting times. The mining problem is typically addressed by either using Process Mining or by means of Machine Learning. These two approaches exploit information from structured data, usually avoiding non-structured ones. Process Mining exploits data in order to provide a pathway representation [16]. However, this technique tends to be ineffective (creation of very complex models) with high variety processes (so-called Spaghetti processes), and this is the case of ED pathways. Nevertheless, in the literature there are different attempts to avoid this problem. For example, the authors of [17] propose an innovative ad hoc process mining approach to discover patients' pathways, that tries to solve the problem through an initial clustering of patients.

Conversely, Machine learning techniques, and artificial neural networks in particular, have the capability of predicting the future pathway of a patient (see, e.g., [18]). This approach allows to rely on extensive information about a patient (represented via a set of *attributes*) to achieve higher accuracy predictions of their needs within the ED. The use of those techniques in predicting patients' needs in terms of resources is clearly not restricted to the ED. For example, [19] presents a machine learning model applied to the master surgical scheduling problem, with the aim of predicting the impact of surgical patients in terms of occupied beds on other areas of the hospital (e.g., intensive care unit).

Once patients' pathways are predicted, the next step is to use this information to improve ED performance and avoid overcrowding. A commonly used approach for addressing this task makes use of simulation [20–22], allowing the creation of what-if scenarios and the selection of the best resource allocation policies to improve the patients' flow. Traditional approaches use discrete-event simulation in an off-line configuration [23], to assess how a specific policy performs. This kind of approach has been successfully used for taking decisions also in other hospital departments. For example, [24] presents a simulation-based optimization method to obtain optimal decisions on patients discharge.

Recently, alternative approaches based on agent based simulation have been introduced [25]. A first attempt of taking operational decisions in an ED based on real-time prediction is proposed in [26], where an ED simulator is used to evaluate the performance of a (fixed) pre-selected policy. The implementation of this policy requires a real-time prediction of patients' pathways, which is obtained by means of a process-mining discovery model exploiting structured data. The (fixed) policy to be implemented in the real ED is then determined by evaluating a portfolio of possible policies according to some performance indicators.

Approaches based on off-line simulation and decision making are in general not fully satisfactory within a highly dynamical system such as an ED, where an on-line approach trying to solve problems before they happen could be preferable. The aim of this paper is to propose a new Decision Support System (DSS) based on the integration of a Deep Neural Network and a simulation tool to take decision on-line. Our approach is original in two aspects: first, the neural network is used to predict patients' clinical pathways by exploiting *all information*, i.e., both structured and not-structured data, collected during the triage process. To the best of our knowledge, there exists no previous attempt reported in the related literature to exploit unstructured data; for example, even the very recent paper [26] makes use of only structured data for prediction. Second, predicted pathways are used within a discrete-event simulator aimed at *on-line* testing different simple policies and *dynamically* selecting the most appropriate one, so as to decrease overcrowding. In other words, the tool is designed to react immediately to any undesired behavior of the system by switching management policy when needed. To the best of our knowledge, this is another unique feature of our approach, while even very recent contributions in the literature only perform an evaluation of fixed policies, without deciding when to switch to a different one (see again [26]). We will focus on "normal" operating conditions, though the approach can be re-trained and re-calibrated to handle exceptional circumstances (such as an ongoing pandemic).

The proposed DSS has been implemented, validated and tested on a real case study derived from one of the biggest EDs located in a major city in northern Italy. The main contribution of our work is the design and the implementation of the DSS. In particular, this requires a nontrivial integration of the different components aimed at predicting and simulating the behavior of the real system. A key task for providing a *useful* tool is to define a proper level of detail in the representation of the ED and the associated data collection. Indeed, while a too detailed representation would result in a DSS with a limited applicability in an ED that is not the one of our case study, a too simplified representation of the operations and resources would yield an oversimplified tool, unable in providing useful guidance. As we discuss in the following, we adopt a representation of the ED where the specific processes and the associated resources are simplified enough to obtain a tool which can be adapted to different settings with a manageable effort. At the same time, the numerical analysis of our case study shows that the resulting DSS is highly effective in reducing overcrowding: that is, we are able to simplify the representation of the system without losing relevant information.

The paper is organized as follow: Section 2 describes the context and data of the problem, with particular attention to the uncertainties that are addressed within the ED. Section 3 presents the predictive models developed to deal with uncertainties, while Section 4 shows the integration of simulation and optimization. Finally, Section 5 reports the results obtained with the DSS in our case study, while Section 6 concludes the paper.

2. Problem context and data

All EDs include the same kind of human actors and resources such as doctors, nurses, clinical staff, technicians, devices, stretchers and beds, all of them interacting within similar processes. Therefore, in this section we present a *generic model* for describing an ED pathway, the interaction with the aforementioned resources, and the information needed to build an effective DSS upon this model. Although the case study under examination comes from an ED located in Italy, the model itself is general enough to allow an adaptation of the proposed DSS to basically any other ED working with the same mechanism.

2.1. The ED pathway

Each patient within an ED follows a pathway that can be summarized as in Fig. 1.

The patient enters the ED either autonomously or with an ambulance. In both cases, the triage phase starts as soon as possible, and a

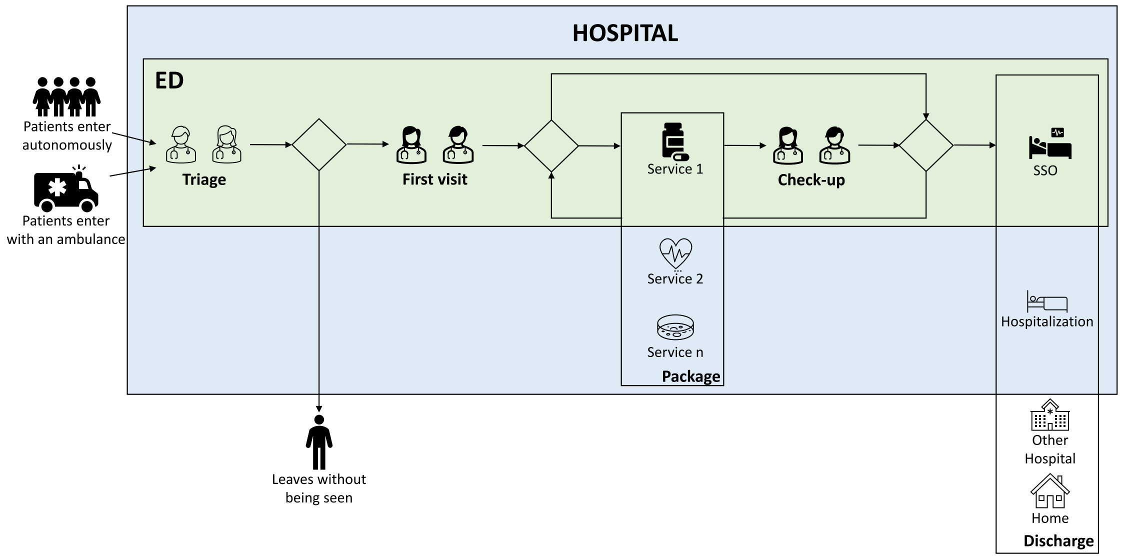


Fig. 1. Generic ED pathway.

staff member, usually a nurse, assesses the patient and registers their personal information (name, age, sex, etc.) and clinical observations (oxygen saturation, blood pressure, etc.). In addition to the structured information, nurses often fill in a text box (nursing diary) with more detailed information about the patient, such as the type of injury or other diseases affecting the patient. This unstructured textual information is sometimes more relevant than the structured one. For example, in the ED we considered, the nursing diary reported annotations for all patients registered in the period from January 2018 to October 2019; in the same period, the field "Main Problem" recorded a very generic information ("Other injuries") in more than 50% of the cases. This is a common situation which applies to our ED and to many others. After receiving a priority during the triage phase, the patient is possibly placed in a waiting room. In case the waiting time extends too much, as it may happen in overcrowded EDs, some patients may leave without being seen. After the waiting time, a patient is checked for the first time by a physician (First Visit).

At the end of the visit, the physician either decides to discharge the patient or requests a further set of services, including, for example, X-ray exams, ultrasound, specialist visit, laboratory exams, therapies etc. In the following we will denote by "package" a set of services which are prescribed together for the same patient with no specific order. A check-up is performed after each package is completed; during this check, the patient is re-evaluated and possibly discharged, or an additional package with further services can be requested. This loop can be repeated several times, until the physician has a diagnosis. Once patients receive the diagnosis, they may have different destinations: discharged, hospitalized, transferred to a Short-Stay Observation (SSO), or transferred to another hospital. In the last three cases, the bed availability has a major impact on the patients' LoS.

2.2. ED components and definition of services

Patients and services are the key elements of our model of the ED. Patients enter the ED because they need to access services, which are associated with scarce resources; therefore, patients may have to queue and the ED may experience overcrowding. Our analysis starts right after triage, reason why we only discuss operations and resources involved after this phase.

Resources are either internal to the ED, i.e., owned and managed by the ED itself, or external to the ED, i.e., owned and managed by different entities while providing services to the ED. The first type of resources includes, among others, ED physicians, ED nurses, ED areas (in terms of number of medical stretchers), while the second type includes, among others, physicians who provide a specific consultation (e.g., orthopedic examination), and the laboratory.

The distinction between internal and external resources induces a different level of detail in their representation within a simulation model. Indeed, a fine granularity must be used for the former resources, whose dynamic heavily impacts on the time spent within the ED by the patients. On the other hand, a too detailed representation would result in an unnecessarily complicated and computationally challenging model. Accordingly, a proper representation of the internal resources is as follows:

- For what concerns internal areas of patient care, the main attributes are the type of treatable patients (e.g., low-urgency patients or high-urgency patients), the schedule of availability (at night or during public holidays some areas may be closed) and number of simultaneously treatable patients (this figure depending on the availability schedule);
- Physicians can be either assigned to a specific area or shared among different areas. In addition, the number of available physicians may vary during the day and physicians may have different skills, which induce different policies for what concerns preemption in case of life-threatening arrivals;
- Care staff (e.g., nurses) and support staff (e.g., porters) are not explicitly modeled, as their contribution can be merged with that of the previous two resources. Indeed, despite their relevance for running the system, their contribution is transversal to the whole process and can hardly be framed in well-defined phases. This is often compounded by a lack of information (e.g., registration) on the tasks performed by this group.

The internal resources mentioned above are needed for performing first visits, checks-up and a subset of services (e.g., therapies), as shown in Fig. 1. For example, for the first visit to begin, a place in the appropriate area and a physician must be available.

As already mentioned, a coarser granularity is sufficient in representing external resources, as only the (potentially) blocking elements for ED patients need to be highlighted. External resources are characterized by the schedule of availability (typically, there are specific slots reserved to the ED) and the number of simultaneously treatable patients (this figure depending on the availability schedule).

All resources mentioned above require the physical presence of the patient for providing the service. In addition, we consider the laboratory, which is a relevant (external) resource not requiring the physical presence of the patient. In major hospitals, laboratory services are provided 24/7 and typically have the capability of simultaneously performing multiple tests for different patients. Laboratory activities are therefore performed in parallel to any other service. From a simulation point of view, we model the laboratory as a service without a queue but having a duration (sampled on historic data) that can stretch the overall duration of the package including it.

2.3. Uncertainty model

In order to timely detect and react to overcrowding, one has to know how the ED will evolve in the near future. We describe the evolution of the ED system by observing its *state*, characterized by a set of measurable variables at a given time, and by forecasting the future value of those variables. As the evolution of those figures depends both on internal actions and on exogenous stochastic events, one has to deal with different sources of uncertainty, namely:

1. uncertainty of arrivals;
2. uncertainty of pathways (temporal sequence of first visit, service packages and check-ups until the patients' discharge);
3. uncertainty of duration of the first visit, check-ups, and services that make up the packages;
4. uncertainty of the effect of internal actions performed on the system.

Indeed, the evolution of the ED state is strongly influenced by the temporal distribution of the future patients' arrivals, as well as by the patients' pathways and by the duration of each specific activity of the pathway.

Formally, let us assume we are interested in modeling uncertainty over a set of n patients, for an Emergency Department operating with m possible packages. We will proceed by introducing, for each patient i, multiple random variables and in particular:

- a random variable T_i with support $\mathbb{R}^+$, representing the arrival time for the patient;
- a random variable X_i representing the information collected on the patient at triage time. This variable is a vector of values associated with a fixed set of attributes, thus its support depends on the type of information that are collected;
- a sequence of random variables $\{Y_{ij}\}_{j=1..n_i}$ each one with support $M = \{1,\dots,m\} \cup \{\perp\}$, representing the sequence of n_i (say) packages for patient i (i.e., their pathway); the value $\perp$ denotes a special package signaling sequence termination, so that $Y_{i,n_i} = \perp$ by construction;
- a random variable D_{ijk} with support $\mathbb{R}^+$, representing the time for the kth service, in the jth package, for the ith patient.

Part of our analysis (see Section 3) will be devoted to determine reasonable distributions and correlations for these variables.

We note that all sources of uncertainty are exogenous, i.e., they are not affected by sequencing decisions. On the other hand, the overall behavior of the ED (including performance indicators such as the number of patients waiting for a visit) depends on the complex interplay between the uncertain factors and the operated choices. For this reason, improving the performance of an ED requires to forecast these sources of uncertainty, to assess their impact, and to define how to search for an optimal policy.

2.4. DSS architecture

In the reminder of the paper, we will describe the building blocks of a Decision Support System (DSS) aimed at optimizing the ED performance. In the next section, we introduce the modules that are used for dealing with the sources of uncertainty, namely:

- an arrival time generator, based on statistical approaches, for the first source of uncertainty;
- a Deep Neural Network to predict patients' pathway;
- a service duration generator, based on statistical approaches, for the visit/check-ups/services duration.

Then, in Section 4 we detail a discrete-event simulator to evaluate how specific management policies affect the ED behavior, and to perform an on-line selection of the best policy.

All data-driven approaches (i.e., the statistical models and the neural network) are meant to be calibrated over available historical data.

3. Predictive model

In this section we present the techniques adopted to address uncertainty about arrivals, pathways and activities duration. The analysis can be applied to any ED, provided the necessary information is available.

There is a major distinction in terms of prediction between patients that have already entered the ED (for whom triage information has already been collected) and patient that might arrive in the future (for whom no information has been observed yet). We discuss the two cases separately.

3.1. Predicting pathways for patients within the ED

The expected pathway is the most relevant aspect to be predicted for patients, as it affects activity queues, patients' LoS and ED global overcrowding. Predicting pathways is a very challenging task, mainly for two reasons. First, since an ED typically offers many services, the number of potential pathways for a patient is very large, as the number of possible combinations of service packages grows exponentially with the pathway length. Second, many pathways are similar, since they include many common services (possibly, in a different order), thus increasing uncertainty in the prediction task. From a formal perspective, this implies that the probability distribution for the package sequences (i.e., $\{Y_{ij}\}_j$) has a very large support and complex correlations.

For patients that have already entered the ED, all triage information is used to perform more accurate inference, resulting in a contextual prediction problem that can be tackled via Machine Learning. Our Machine Learning-based approach is used to forecast the actual pathways by exploiting the available patients' information, including the unstructured one from the nurse's diary, and predicts the packages of services for a patient, one at a time (see Fig. 1), until the whole pathway is determined. This is coherent with the metric that we adopt for evaluating the performance, in terms of accuracy, of our pathway predictor (see Section 5.3), which considers the accuracy in predicting the *next* package, right after the current one. Indeed, the proposed approach is more accurate that an alternative one in which the entire pathway is forecast in one step, as the number of service packages is much smaller than the number of their combinations into pathways.

Formally, we adopt a factored approximation for the distribution of all possible package sequences. In particular, we approximate the distribution of possible sequences with a product of probabilities:

$$P(\{Y_{ij}\}_{j=1..n_i}) \simeq P(Y_{i,1}) \prod_{j=2..n_i} P(Y_{i,j} \mid \{Y_{i,h}\}_{h=1..j-1}) \tag{1}$$

In practice, when sampling the next package we use as an input to the estimator the sequence of all packages observed or generated so far for the considered patient.

In addition to the pathway taken so far (sequence of service packages), our prediction is based on a number of additional inputs/conditioning factors, i.e., X_i from our probabilistic model. These variables correspond to information that is systematically collected at triage time, namely:

- Age;
- Sex;

- Patient's urgency (triage code);
- Text in the nurse's diary.

In order to use these features within a Deep Neural Network (DNN), a preprocessing step is needed. All categorical attributes (i.e., sex and urgency) are represented using a one-hot or label encoding; in our implementation, we use methods available in the scikit-learn library [27]. The nurse's diary consists of free text and requires preprocessing for being used. To this end, we consider a Natural Language Processing (NLP) approach based on a Bag-of-words model. We have tested three alternatives for processing the text:

1. a combination of the Python Natural Language Toolkit (NLTK) [28] for normalizing the text (e.g., removing stop words) and scikit-learn for creating tokens and word n-grams;
2. the same approach as above, plus a stemming step, performed via the Snowball algorithm for the Italian language [28];
3. an open source tokenizer based on a version of the Bidirectional Encoder Representations from Transformers (BERT) [29] for the Italian language (see [30]), implemented using the PyTorch framework [31].

A comparison of these approaches will be discussed in Section 5.3.

All features are given as input to our Machine Learning model, which is a Feed-forward Deep Network classifier, implemented in PyTorch. In particular, all hidden layers are handled by means of a simple architecture using ReLU neurons. Indeed, ReLu is the most frequently used activation function in feed-forward networks since it has some nice properties, including simplicity and short execution time while preventing vanishing gradient effects (see, [32]). As to the output layer, a SoftMax activation function is used, so as to allow a normalization of the output of the network and, therefore, its use as a mass probability function.

The model is trained to approximate the probability of the next package, conditioned on the information encoded by the model input. Within this setup it is easy to sample the next package at random, with a distribution defined by the model output. As an alternative, a deterministic behavior can be obtained by considering the class (i.e., the package) with the highest estimated probability as the only prediction. We will consider both these operating modes in our experimental evaluation.

3.1.1. Limitations

A known limitation of our factorized approach consists in its inability to account for future packages when making predictions: correlations between such packages may arise due to hidden variables, e.g., the actual patient ailment at arrival time. However, in practice, packages (i.e., groups of services) are assigned based on information that becomes available only as a result of the services themselves. For example, a medical exam may reveal additional information, which is then used to define the next package for the patient. For this reason, we expect our factorization to be a reasonably good approximation. Our preliminary experiments aimed at training a model for the non-factored distribution seem to confirm this conjecture.

Moreover, in our historical dataset, prescribed packages are affected by service availability in addition to the patient condition (e.g., a particular exam or medical specialist may not be available at a specific time). Since we are not providing availability information as input to our model, these effects will act as noise on the training distribution, thus making the learning problem more complex.

3.2. Predicting arrivals and pathways for future patients

Since no observed information is available for patients that might enter the ED in the future, we need to predict both their arrival times and pathways.

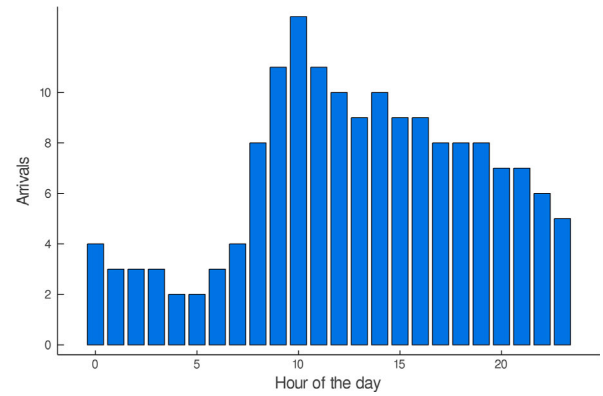


Fig. 2. ED daily arrivals.

In an emergency department, arrivals are not scheduled in advance (non-elective), hence, they are best modeled as a stochastic phenomenon. Nevertheless, the large amount of historical data that is typically available allows to forecast with good approximation the number of patients who will arrive during a specific time period.

Under the reasonable hypothesis that, for a limited time interval (e.g., one hour), arrivals are independent and occur with constant mean rate, the arrival count is well described by a Poisson process (see [33]). An analysis of our dataset confirmed the validity of this assumption, and showed that using time-dependent rates consistently leads to better estimates. In particular, the rate seems to be driven by three factors: the hour of day, the day of week, and the month, the first having by far the largest impact. Fig. 2 shows the average number of arrival λ_h as a function of the hour of the day h.

Accordingly, in our DSS we modeled inter-arrivals as independent events, following an exponential distribution with time-dependent rate,

$$P(T_{i+1} - T_i) = \lambda_h e^{-\lambda_h (T_{i+1} - T_i)}, \tag{2}$$

i.e, the probability of an inter-arrival delay equal to $(T_{i+1} - T_i)$ is controlled by a parameter λ_h which depends on the hour h of the day. By assuming $T_1 = 0$, this is sufficient to characterize the full arrival-time sequence.

As a consequence of our modeling choice, the number of arrivals per hour is a Poisson process, with rate dependent on the hour of the day. As a direct consequence, we can calibrate λ_h values by simply computing means over historical data. More in details, λ_h is computed as the average number of arrivals in time interval $[h, h+1)$.

Moving to the problem of predicting the pathways, we generate those by adopting a simple statistical approach.

- First, we generate urgency code by sampling from a discrete probability distribution, that has been calibrated by computing historical frequencies for all urgency code values.
- Then, we sample packages (i.e., all $\{Y_{ij}\}_j$) recursively according to Eq. (1), using conditional probabilities that have been calibrated via historical frequencies, computed separately for each urgency code.

The alternative would be to build a data generator for the X_i distribution, then use the neural approach. However, this approach introduces an additional source of noise. In principle, we could have introduced additional conditioning factors that are easy to determine even within a simulation, such as the hour of the day, the day of the week and the month. However, in our case study, we did not observe strong influence of these drivers on the prediction.

3.3. Predicting activities duration

The state of the ED also depends on the duration of services, which influences the availability of resources, the patients flow and the waiting time.

We adopted a traditional statistical approach to model and sample service duration. In particular, we use a parameterized LogNormal distribution, with no conditioning factor, as it turned out to be the choice giving the best fitting compared with other distributions (e.g., exponential or normal). We estimate the distribution type (from a predefined set) and the distribution parameters via classical approaches (e.g., sample mean and sample variance computation) over historical data.

3.3.1. Limitations

Given the highly dynamic environment of an ED, it may happen that duration information for some services is not properly registered, thus reducing data quality. Typical issues arise, for example, for services that are not registered or are only partially registered (e.g., the starting time but not the ending time are registered), or are registered in a wrong way. These issues can be handled by means of a preprocessing step, aimed at identifying outliers, activities with incomplete information, etc.

4. Simulation-optimization DSS

In this section we present the logic of our DSS: it includes a simulation tool which integrates the aforementioned predictors and can be used to test a portfolio of alternative policies for managing the ED. By selecting the best policy, the DSS performs an improvement of the ED.

4.1. DSS functional architecture

Each predictor of the DSS addresses a specific source of uncertainty. Those predictors are embedded within a simulator, so as to model the dynamic behavior of the ED, including patients flow, resources utilization, and evolution of queues.

We have developed a discrete-event simulator of the ED, having the level of detail discussed in Section 2.2. The simulator reads the current state of the ED (hour of the day, availability of the resources, length of the associated queues, patients within the ED and their features) and uses the predictors to forecast the system evolution under a specific policy. The simulator functional scheme is shown in Fig. 3. In the figure, the yellow block highlights the state information taken from the ED, the green blocks indicate the predictors that are used, and the blue block represents the simulation system. Finally, the red block represents the expected result. Through the DNN new packages are assigned so as to complete the pathways of the patients within the ED. Meanwhile, new arrivals are generated by the arrival predictor, and the corresponding pathways are defined through the process described in Section 3.2. Each service of the pathway is assigned a duration through the procedure described in Section 3.3. Both types of pathways are passed to the discrete-event simulator and used to evaluate the evolution of the ED system. The simulator was implemented within SimPy [34], a process-based discrete-event simulation framework based on Python.

The DSS is designed with the goal of dynamically identifying the most suitable policy to be implemented in the real system. To this end, it must replicate the ED behavior, while evaluating the impact of alternative policies. Indeed, the DSS allows the decision maker to test different policies and to choose the most appropriate one based on suitable metrics. Fig. 4 shows the integration of the DSS with the ED system. The DSS tool is triggered every Δz time units, when it is fed with the actual state of the real ED. It simulates the behavior of the ED under different policies, for a time interval ΔT, denoted as search depth. Each simulation is repeated ω times, so as to obtain statistically relevant information. The simulation returns, after a limited time, the selected policy to the policy-maker (red arrow), who can either accept or reject the proposal.

Parameters ΔT and ω affect the computing time needed to run the simulation. Clearly, having a fast answer is mandatory in order to be able to implement the suggested policy before the real ED has changed too much. In addition, using a too large ΔT value is counterproductive, as the resulting prediction is mostly based on hypotheses about future arrivals. Concerning the trigger time Δz, the lower the value of this parameter, the better the uncertainties of prediction are addressed. However, if the trigger time is too short, the implemented policy can change too often, thus confusing the decision maker.

4.2. Alternative policies

Since overcrowding is the major issue we want to approach by means of our DSS, when selecting the best policy to implement within the ED we consider a KPI which is a proxy of overcrowding, as detailed in the next section. Our framework is easily adaptable in case the decision maker is interested in analyzing different KPIs.

We consider alternative policies which are related to the way in which queues are handled, and dynamically select the one which performs at best with respect to our KPI. In detail, each patient can be prescribed one or more services within the current package. Services typically require the access to some scarce resource, and are associated with queues where patients wait for the resource to be available. All considered policies treat in the same way patients requiring multiple services within the same package, and assign the patient to the queue of the service having the shortest (expected) waiting time. The expected waiting time depends on the characteristics of other patients in the same queue, their forecast service times, and on the queue handling policy. In our DSS, we implemented three simple policies for selecting the *next* patient of a given queue to be served. Although the prediction of patients' pathway is a complex task and is handled by means of sophisticated statistical methods, the actual policy used for operating the ED is defined by a combination of these simple policies. We believe that simplicity of those policies is important for two reasons: first, the implemented policy can change each time the DSS is invoked, and thus considering more involved rules would probably have a negligible impact on the system performance. Second, simplicity is a plus when policies have to be understood, accepted, and implemented by human decision makers. The same argument suggests the adoption of a small set of policies: increasing the size of the policy pool or the degrees of freedom of their combination would considerably add complexity to the system, making its behavior less transparent and acceptable. Conversely, the prediction of future events, which directly implies the most suitable policy to be implemented given the current state of the ED, is outside the decision maker's control, and is handled with sophisticated techniques. Finally, the dynamic combination of those simple policies allows to obtain an overall flexible policy whose a priori static definition would be far from trivial.

In all policies, priority is given according to the urgency code. In order to avoid patients starvation, the ED implements a mechanism to push forward patients who are experiencing a too long waiting time. This is obtained by increasing the urgency of a patient who has spent a certain amount of time in a queue. Ties in priority are then broken as follows:

Policy 1 selects the patient with smallest expected time for the not yet performed services of the current package, breaking ties by smallest expected time for future packages. This policy gives priority to patients who are likely to complete their current service package shortly, and can be useful in case of overcrowding.

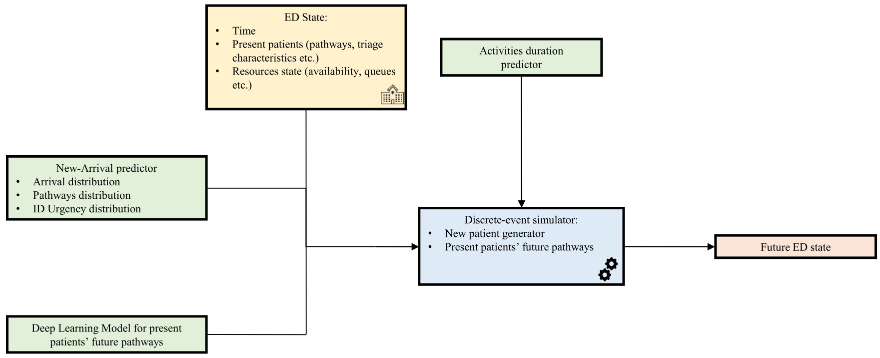


Fig. 3. DSS functional scheme. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

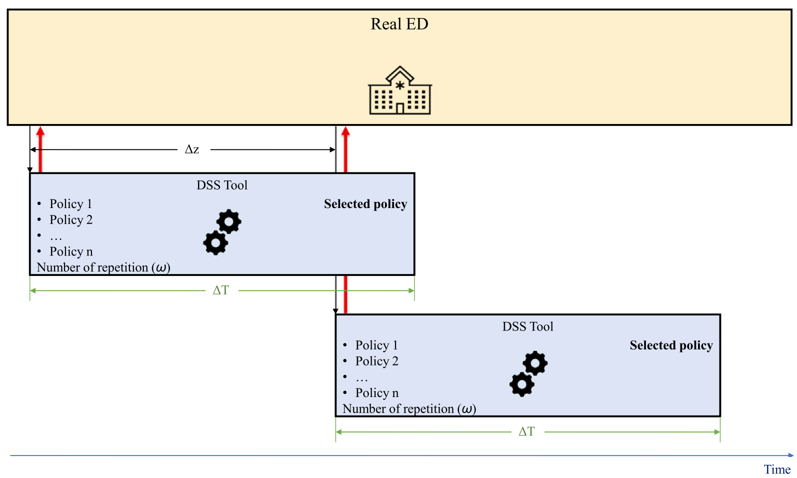


Fig. 4. Integration between ED and DSS tool. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Policy 2 selects the patient with largest expected time for the not yet performed services of the current package, breaking ties by largest expected time for future packages. This policy gives priority to patients with long expected service times and can be useful to process resource-demanding patients when the ED is not under pressure.

Policy 3 selects the next patient according to a FIFO rule. This policy is commonly used in EDs, including the one in our case study.

Regardless of the selected policy, a patient cannot change the queue to which they are assigned and move to a different queue (before being served). This constraint follows from the design of hospitals, where ambulatories for different services can be far from each other, and moving a patient can take time. In our analysis of the real case, we will also consider an additional policy that violates this constraint, i.e., where we assume that a patient can change their queue (in no time). Although this additional possibility is in general infeasible, its evaluation provides an optimistic estimate of the ED performance. It should be kept in mind that an implementation of such a policy could however require a redefinition of the hospital layout and logistics. We will refer to this policy as fixed-queue relaxation (FQR) policy in the rest of the paper.

5. Empirical evaluation and validation

In this section we present a quantitative study of the introduced DSS and provide a detailed analysis of its capabilities, based on a real-world use case.

5.1. Validation on a digital twin of the real ED

The integration of the developed DSS with the real ED requires a preliminary phase of validation and fine tuning, and a non-negligible investment in terms of time, human and financial resources. In order to assess the capability of the system before the integration is performed, and to perform such a tuning, we validate the tool and study its effect on the system through *a second simulation model*, which in our experiments plays the role of the real system in Fig. 4. This digital twin (DT) is implemented as a discrete-event simulator, and it is intended to be a tight approximation of the real ED: for this reason, it is fed with historical observations of the past ED behavior, i.e., real arrivals, patients' features (urgency, triage information etc.), clinical pathways, duration of the activities, and resource availability information.

The digital twin can be configured either with the same policy used within the real ED, so as to assess its capability in replicating the latter, or can follow the prescription of the DSS, so as to evaluate the impact of the DSS on the system.

The experiments performed on our case study showed that the digital twin, in the former configuration, provides a tight approximation of the real ED: the mean LoS recorded within the DT and the real one, for the whole available period (from January 2018 to October 2019), differ by 2% (Student's t-tests revealed a non-significant difference between the two statistics), and they have a similar resource occupation (global mean error equal to 5%, with a non-significant difference between the two statistics).

Once we have verified the performance of the digital twin, we can switch to the second configuration, which is used to replace the real ED in our tuning and experiments, so as to estimate the impact of the proposed DSS.

Table 1
Structure of the dataset providing information on ED accesses.

Case_ID	Arrival time	First visit time	Discharge time	Fast-track	Age	Sex	Urgency code	Diary	Symptoms
numerical	numerical	numerical	numerical	binary	numerical	binary	numerical	string	string
Random key identifying the single ED access of a specific patient	Timestamp of the ED access, i.e., after the triage	Timestamp of the first visit	Timestamp when the patient is discharged	Whether patient follows a fast-track or not	Age of the patient	Sex of the patient	Code denoting the urgency of treatment	Nurse's diary	Main symptoms registered at triage according to a regional encoding
100% filled in	100% filled in	100% filled in but for patients who died before arrival and LWBS patients	100% filled in but for patients who died before arrival and LWBS patients	100% filled in	99.9% filled in	99.9% filled in	100% filled in	99% filled in	99% filled in

5.2. The case study

The proposed DSS has been tested on a data set derived from real observations from one of the largest metropolitan EDs of a northern Italian region. The considered ED is classified as a second-level emergency acceptance department (DEA 2), the highest complexity level within the Italian classification (see, [35]). Such an ED provides a number of highly specialized services, and a 24/7 radiology.

5.2.1. Internal organization

After arrival, patients undergo the triage process, which is performed by two nurses working in parallel. Low-urgency patients with specific needs (e.g., oculist problems) are enrolled in so-called fast-track pathways, and sent to a specialist. After the triage, patients undergo the first visit in two different rooms, depending on their needs and urgency, while a shock room is available for caring severely urgent patients. Close to the ED there are dedicated radiology clinics, as well as an MRI and a CT scan, serving the ED and the whole hospital. In addition, the hospital provides laboratory services and 24/7 specialist consultancy.

Once a patient has completed their urgency treatments and has received a diagnosis, they may have different destinations:

- discharged, in case the there is no immediate need for further clinical treatments or assistance;
- hospitalized, in case further complex clinical treatments are needed;
- transferred to a SSO, in case short stay observation and/or low intensity care is needed;
- transferred to another hospital;
- other possible results (including death).

5.2.2. Retrospective data analysis

In this section we analyze the available historical data from the ED at study. We considered anonymized real information registered for patients treated between January 1st 2018 and October 31st 2019, for a total of 109 201 accesses (this figure includes also patients who died before arriving to the ED).[2]

Raw data were accessed through 5 different datasets.

The first dataset provides all general information on the ED access of patients. Information is organized as described in Table 1. Missing values in numerical columns were replaced with median values. As for the binary ones, a random value was generated based on the distribution of the existing values, when needed.

The remaining 4 datasets refer to services provided to patients during the ED pathway, and are related to different medical areas:

- ED dataset: contains information on services delivered directly within the ED (e.g., therapies), whereas it does not include the first visit (found in the previous dataset), and check-ups (see 5.2.3);
- Laboratory dataset: contains information on services provided by the laboratory;
- Radiology dataset: contains information on radiological services;
- Ambulatory specialty dataset: contains information on other types of specialist services (e.g., surgical examination).

All these 4 datasets share the same structure, which is described in Table 2. In this table, missing values in the request timestamp were filled in with the timestamp of the start of the service. Missing data concerning the timestamp of the start of the service were filled in by subtracting the median duration of the service from the ending timestamp (this figure being always available). Concerning missing data in the Resource field, in some cases the team that provided the service could be inferred with good approximation from similar services performed in the same time slot; when this was not possible (less than 1% of the cases) a generic resource was assigned to the service, with 8 AM–8 PM availability in the simulation.

The data was then cleaned, under clinical manager guidance, by removing compilation errors, i.e., life-threatening patients who left without being seen or patients with a LoS greater than 48 h. This way, a sample of 109 176 accesses was obtained.

Table 3 shows the number and percentage share of patients for each of the 5 urgency codes, from the most to the least urgent.

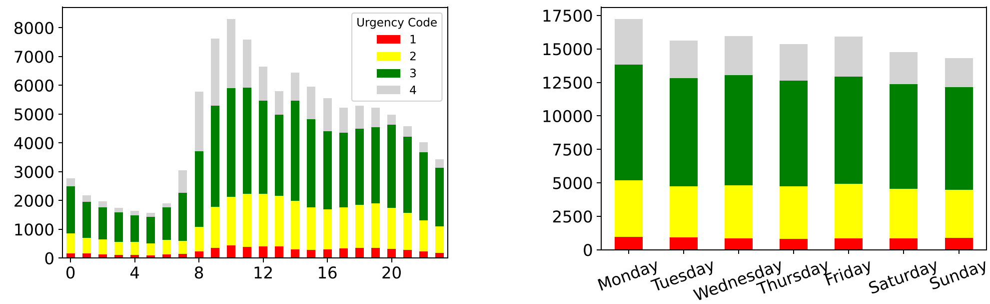
Table 4 reports the different kind of pathways of each patient who entered the ED, disaggregated by urgency code. Accordingly, we do not report figures for patients who died before arrival.

The table shows that most of the patients follow a pathway within the ED, with the exceptions of those following the fast-track pathway or leaving the ED without being seen (LWBS). It is well-known (see, e.g., [36–38]) that the number of LWBS is strongly correlated with overcrowding and waiting time for the first visit. However, unfortunately, the time when people leave an ED without being seen is typically not registered, the only timestamp available being the one when an ED operator notices the patient be missing. This time is probably a coarse upper bound of the true timestamp when the patient left, especially during periods of overcrowding. Thus, the time spent in the ED by LWBS patients is unknown and cannot be forecast. For this reason, these patients are not further considered in our analysis, also because they do not use the ED resources and leave it after the triage.

Accordingly, we focus on the remaining 81 771 accesses (49% Female and 51% Male), with an average age of 59.5 ± 22.9, which include patients who left before end-of-treatment (LBET); for the considered patients, we registered an average of 122.4 daily accesses, with a median of 123, a minimum of 86 and a maximum of 153.

Fig. 5 reports the daily 5(a) and weekly 5(b) trend of arrivals for each urgency code. The figure shows that, while the trend of arrivals for each urgency code is almost independent on the day of the week, it

[2] All data were treated according to the European GDPR 2016/679. All structured and unstructured data were cleansed of information that might make patients recognizable to themselves or others.



(a) Daily trend of arrivals

(b) Week trend of arrivals

Fig. 5. Trend of arrivals.

Table 2
Structure of the dataset providing information on services.

Case_ID	Activity	Request timestamp	Starting timestamp	Ending timestamp	Resource
numerical	string	timestamp	timestamp	timestamp	string
Random key identifying the single ED access of a specific patient	Description of the service provided	Service request timestamp	Start of service timestamp	End of service timestamp	Team that provided the service
100% filled in	100% filled in	70% filled in	92% filled in	100% filled in	98% filled in

Table 3
Urgency codes: description, number and percentage share.

Urgency code	Type of patients	#	%
1	Life-threatening patients	6 193	5.7
2	Patients who need an urgent visit	27 249	25.0
3	Patients who need a not-urgent visit	56 381	51.6
4	Patients with minor injuries	19 301	17.6
5	Patients who died before arrival	52	0.1

Table 4
Type of treatment based on urgency code.

Treatment urgency code	ED pathways	LBET	Fast-Track	LWBS
1	6 193	0	0	0
2	26 879	174	0	196
3	45 767	282	94	10 238
4	2 419	57	12 011	4 814
Sum	81 258	513	12 105	15 248

Table 5
Frequency of main symptoms.

Symptom	Frequency
Other injuries	50.5
Trauma	17.2
Abdominal pain	8.7
Dyspnea	5.1
Chest pain	4.0
Stroke	3.1
Different from the previous	11.4

is strongly affected by the hour of the day, with a peak in late morning, and 68% of arrivals between 8 AM and 8 PM.

The main symptoms registered for the patients at triage are shown in Table 5. In the majority of the cases, the nurse selected a generic "Other injuries" entry, and provided more unstructured information in the nurse diary (compiled in about 99% of cases). This observation witnesses the importance of the nurse diary as a source of valuable information for forecasting the patients' pathways.

Table 6 reports the average and median waiting time for the first visit, and length-of-stay within the ED, disaggregated by urgency code. This table will be the basis for evaluating the effect of our DSS on the ED; in particular, it shows that critical aspects refer to low-urgency code patients, who may experience long waiting time and LoS.

Table 6
Patients' waiting time and LoS.

Urgency code	Waiting time [min]		LoS [min]	
	Mean	Median	Mean	Median
1	11	8	154	124
2	34	19	219	191
3	210	183	361	334
4	276	230	373	326
Total	139	77	298	263

5.2.3. *Assumptions*

In this section we describe some assumptions introduced in our analysis, which derive from data availability.

- As to the first visit, only the timestamp associated with the start of the visit is registered. In order to estimate the duration of the visit, we assume its ending time be the instant in which the last request for a service in the first package is registered;
- Similarly, the duration of the check-up visit taking place right after each package of services is not registered. Based on an empirical observation, the duration of each check-up is set to 15 min;
- For what concerns the services, timestamps are registered both at the beginning and at the end, which allows to compute their duration. If this information is missing, the median value for other occurrences of the same service is used;
- Services are grouped into packages according to their request timestamp. We assume that requests registered within 15 min belong to the same package;
- We only consider services that are performed during the stay of the patient within the ED, i.e., we ignore services that are prescribed in the ED but that are performed *after* the patient is discharged (for example, services requested by the ED physician but performed during hospitalization or SSO);
- We exclude non-traceable events that were not digitally registered. These events represent a marginal fraction of the total, and typically include non-critical treatments or prescriptions (e.g., dietary prescriptions);

- Although we only consider the patients' pathways within the ED, the availability of resources *beyond* the ED may have an impact on the LoS *within* the ED itself. A relevant case is bed-blocking, which can force a patient to remain within the ED after their pathway is terminated, because no immediate hospitalization is possible. Our DSS has no effect on activities beyond the ED. However, in order to have a fair comparison with the real behavior of the system, in the following analysis all figures are obtained by assuming that, after the last registered activity, each patient may have an extra waiting time in the ED depending on their pathway and final destination;
- A package requiring the physical presence of a patient can only start upon completion of the previous package. In addition, we assume that any physical movement of the patients occurs in zero time.

5.3. *Performance of the pathway predictor*

We now evaluate the performance, in terms of accuracy, of the two pathway predictors embedded within the DSS. For training and testing purposes, we considered all historical data but those from 01/10/2019 to 15/10/2019, which were reserved for evaluating the DSS.

5.3.1. *Performance of the predictor for patients within the ED*

Concerning patients within the ED, our historical data include more than 195 000 packages, combined into 3704 distinct pathways. Among them, the 130 most frequent pathways cover 90% of the occurrences; these pathways are composed by 46 different packages of services. We restrict our attention to those packages. Accordingly, the Deep Learning Classifier predicts the next package of services within these 46 variants only, the "End" of the pathway being represented as a single package. The complete list of the 46 packages, together with their frequencies in the selected pathways, is reported in Table 7. Within the selected pathways, the number of packages always ranges between 1 and 4, with an average value equal to 2.7 and a standard deviation equal to 0.7. Pathways with a single package correspond to patients who are discharged right after the first visit, in which case the package corresponds to the "End". Some packages have a relatively large frequency compared with other packages which appear quite rarely; however, we decided not to use re-sample techniques within our DNN, which is intended to be used as a generator of the patient's subsequent packages and thus must learn the true distribution of the data (prediction based on the knowledge domain).

In order to train, validate and test the classifier, the dataset of packages was randomly divided into three parts with a ratio 80-10-10. As mentioned, prediction of the next package is based on both structured and textual information, as well as the packages prescribed up to now. Concerning the text processing, we used n-grams with length up to two, so as to represent semantic concepts such as "*not-something*". At the end of processing, we obtained three different vocabularies (one for each preprocessing approach described in Section 3.1) with a different number of terms. For each considered vocabulary, we tested different configurations of the DNN in terms of number of layers, their size, and training parameters. In our study, we did not observe significant improvements by using more than 4 layers. Hyperparameter tuning was conducted manually, as preliminary experiments showed the system be robust with respect to alternative configurations. This also allowed us to avoid overfitting over the validation data.

Table 8 reports, for each vocabulary, the best network setup and the associated accuracy result over the test set, computed by using the predictor in a deterministic fashion (i.e., by returning the class with the largest estimated probability as output). The results show that all approaches are capable of correctly predicting the next package among the 46 existing variants in more than 50% of the cases. The best accuracy level is obtained by Approach 3, which makes the right prediction in more than 60% of the cases. Indeed, these results are satisfactory, given that many packages are very similar each other in terms of composing services and frequency, and quite often a wrong prediction is indeed partially correct. In other words, in our application, some "different" packages are not "that different", and predictions that are evaluated as wrong in a binary classification, are not completely wrong if one performs a more refined analysis. For example, predicting a package that includes "Laboratory, Therapy, X-ray, CT scan" while the real package is "Laboratory, X-ray, CT scan" is an error, though not as severe as if the real package were "Neurological examination". Being our case a multiclass classification where many classes have a high degree of similarity, a better evaluation of the classifier can be obtained with a different approach. We computed the similarity of the packages through the Jaccard index [39] applied to their component services, and represented it through a diagonalized heatmap (Fig. 6). The color distribution of the figure shows that several very similar packages exist, confirming the hardness of performing a perfect prediction. By reporting, within each cell in the figure, the actual frequencies of the packages shown on the y-axis, against the packages predicted by the network on the x-axis, we are able to obtain a finer information about the behavior of the network when the prediction is not "perfect". Indeed, the figure shows a good accuracy on the packages with the highest frequency (e.g., "End") and, at the same time, a tendency for errors to spread to packages with the highest similarity. In other words,

Table 7
Package frequencies.

Packages	Frequency
End	37.00
Laboratory	13.24
X-ray	10.63
Laboratory, X-ray	8.63
CT scan	5.49
Laboratory, CT scan	3.33
Ultrasound, Laboratory	2.33
Ultrasound	2.15
Ent examination	2.11
Therapy	1.91
Laboratory, X-ray, CT scan	1.64
Ultrasound, Laboratory, X-ray	1.44
Ultrasound, X-ray	1.44
X-ray, CT scan	1.27
Laboratory, CT scan, Neurological examination	0.68
Neurological examination	0.68
Therapy, X-ray	0.65
Laboratory, Therapy	0.63
Orthopedic examination	0.62
Laboratory, Therapy, X-ray	0.62
Ultrasound, Laboratory, X-ray, CT scan	0.46
Laboratory, Ent examination	0.46
CT scan, Neurological examination	0.28
Eye examination	0.27
Ultrasound, X-ray, CT scan	0.26
Laboratory, Neurological examination	0.19
X-ray, Ent examination	0.16
Ultrasound, Laboratory, Therapy	0.16
Dermatological examination	0.15
Urological examination	0.13
Therapy, CT scan	0.12
Laboratory, Therapy, CT scan	0.10
Ultrasound, Laboratory, Therapy, X-ray	0.09
Ultrasound, Therapy	0.08
Ultrasound, Therapy, X-ray	0.07
Laboratory, X-ray, Ent examination	0.07
Surgical examination	0.07
Laboratory, Therapy, X-ray, CT scan	0.05
X-ray, Orthopedic examination	0.05
Therapy, X-ray, CT scan	0.04
Other treatments	0.04
Laboratory, X-ray, CT scan, Neurological examination	0.04
Ultrasound, CT scan	0.04
Therapy, Ent examination	0.04
Laboratory, CT scan, Ent examination	0.04
Ultrasound, Laboratory, CT scan	0.04

Table 8
Predictor results based on the NLP approach.

Approach parameter	1	2	3
Vocabulary size	42 034	34 932	8919
Layer size	4004, 2000, 500, 46	4004, 6500, 2500, 46	6004, 3500, 1500, 46
Opt. alg.	Adam	Adam	Adam
Batch size	32	32	32
Epochs	10	3	1
Learning rate	0.005	0.005	0.005
Test set accuracy	55%	54%	62%

Fig. 6. Packages distribution of prediction, grouped by Jaccard similarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

when the prediction is not perfect, still the network has the capability of selecting packages which are very similar to the one that is observed in the real data.

In the next section, we will show that the obtained accuracy prediction performances are appropriate and allow to obtain a reliable DSS built on top of the predictor.

5.3.2. Performance of the predictor for future patients

Concerning the future arrivals, we only have statistical information based on historical data. Figs. 5(a) and 5(b) show that the hour of the day strongly affects the number of arrivals, which is only marginally affected by the day of the week.

For a given hour of the day, the distribution of arrivals is well approximated by a Poisson distribution, as witnessed by Fig. 7, showing the tight fitting of that distribution with observed arrivals at given hours of the day (2 AM, 11 AM, 2 PM and 8 PM).

In order to evaluate the tightness of the distributions, the dataset was split in training set and test set, with a ratio 80-20. We computed the parameters of each hourly distribution on the training set, and measured the Mean Absolute Error (MAE) on both subsets, obtaining values equal to 1.64 and 1.68, respectively, thus confirming that the Poisson distribution has a good fit with observed data.

5.4. DSS performance

In this section we study the capability of the DSS of performing tasks of increasing complexity.

- First, we assess its accuracy in predicting the ED future state under the current policy. The DSS is fed with the initial state of the ED, and its capability of correctly forecasting the ED evolution over time is tested;
- Second, we assess its accuracy in identifying the best policy over a set of possible policies. The DSS is fed with the initial state of the ED and, in turn, a fixed policy. Then, the DSS capability of correctly forecasting the ED evolution over time under the fixed policy is evaluated. This allows the decision maker to identify the best policy for the initial state;
- Finally, we assess its capability of improving the ED performance. In this experiments, the DSS dynamically selects the best policy to be implemented in the ED in the next time period.

In the following, all experiments were carried out with a number of simulation repetitions $\omega = 50$. Although smaller than what is reported in other discrete-event simulation studies, this number allows us to

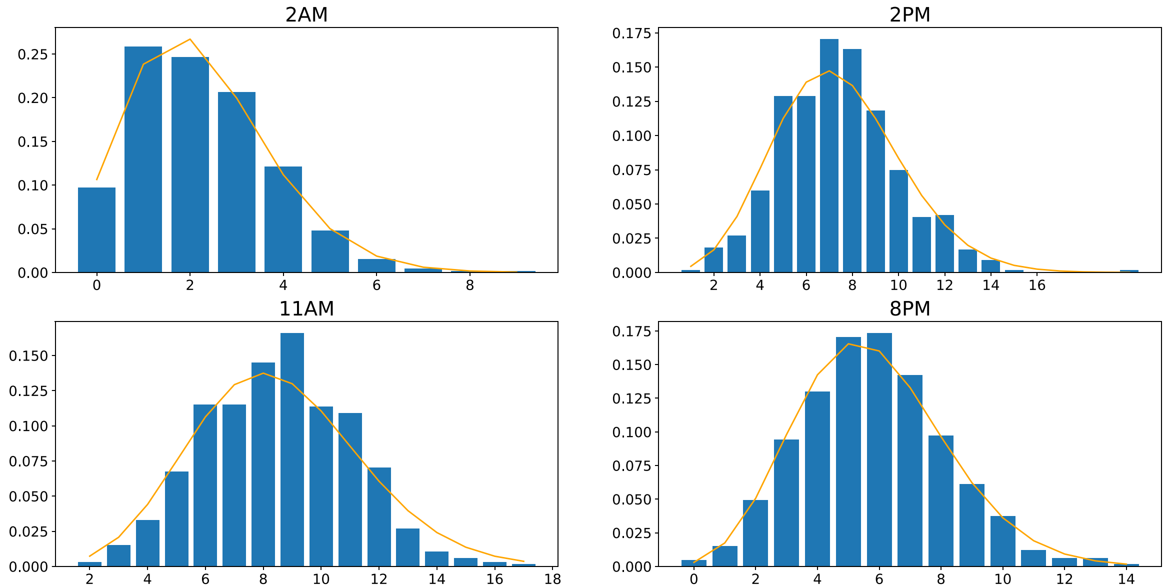


Fig. 7. Poisson distribution fitting with arrivals at different hours of the day.

derive a fast DSS, which is capable to provide indication to the decision maker in a real time. In addition, as the period to be simulated is very short, any error can be corrected at the next run, and the overall number of simulation runs is much larger than 50. A preliminary tuning process showed that this value provides good results while requiring a short execution time (less than 8 min for all 50 runs), while only marginal improvements are obtained by largely increasing this figure.

5.4.1. *ED future state prediction*

In order to evaluate the accuracy in predicting the ED future state, we randomly selected 100 different days and took snapshots of the ED state at 2 AM (the time when ED has minimum overcrowding), 11 AM (one hour after the arrival peak, when decisions have a critical effect on overcrowding), 2 PM (typically, the hour of maximum overcrowding), and 8 PM (a second critical time in the day, due to the staff change of shift). These snapshots were used to initialize the ED state, and to predict the number of patients in the ED after 1, 2, and 3 h by following the policy currently used in our ED, i.e., the third one described in Section 4.2. The obtained predictions were compared with the real historical values.

Table 9 gives the number of patients in the ED (mean value and standard deviation) and reports data on the accuracy in predicting the number of patients in the ED, measured by the MAE between the predicted number and the observed figure. Data are reported separately for each of the three approaches introduced in Section 5.3.1, both in their deterministic (considering the package with the highest estimated probability) and stochastic versions (extracting a random package based on the estimated probability). The MAE of the best configuration is reported in bold.

The results show that our approaches are capable of producing tight predictions, as MAE values are smaller then the corresponding standard deviations (std). The best accuracy is obtained with Approach 1 in the deterministic mode, and with Approaches 2 and 3 in the stochastic mode. In all cases, the performance worsens by increasing the search depth ΔT, thus showing that 3 h is probably a too long time period for being simulated.

5.4.2. *Best policy prediction for improving the ED performance*

The ED performance can be evaluated under different metrics whose relevance varies for different stakeholders. A first performance indicator we consider is overcrowding, for which a good proxy is given by the average number of patients within the ED. Overcrowding affects

Table 9
ED future state prediction results: historical data and Mean Absolute Error of the considered prediction approaches.

Hour	ΔT	# patients		MAE Appr. 1		MAE Appr. 2		MAE Appr. 3	
		Mean	std	Det.	Stoc.	Det.	Stoc.	Det.	Stoc.
2 AM	1	18.3	5.5	**1.6**	1.7	1.9	**1.8**	**1.8**	**1.8**
	2	16.6	5.7	**2.7**	2.9	3.0	**2.8**	3.0	**2.9**
	3	15.2	5.9	**2.8**	3.0	3.3	**3.0**	3.2	**3.0**
11 AM	1	30.2	8.2	**3.2**	3.4	3.9	**3.5**	3.7	**3.0**
	2	31.5	7.7	**4.5**	4.6	5.9	**5.0**	**5.7**	5.8
	3	30.3	7.8	**5.4**	5.5	7.2	**6.8**	**7.1**	**7.1**
2 PM	1	32.2	7.1	**2.4**	**2.4**	3.2	**2.5**	3.2	**2.6**
	2	32.6	7.4	**3.9**	4.0	5.3	**3.9**	5.2	**4.0**
	3	31.3	6.4	**4.0**	4.3	5.7	**4.2**	5.6	**4.3**
8 PM	1	26.1	6.1	2.7	**2.6**	3.2	**2.7**	3.1	**2.7**
	2	26.3	6.6	**3.6**	**3.6**	**3.9**	**3.9**	**3.7**	3.8
	3	25.0	6.2	**4.0**	4.4	**4.4**	**4.4**	**4.3**	4.4

the stress level of the operators and hence the quality of the service provided within the ED. From the patient's perspective, a very relevant indicator is instead represented by the overall length of stay. Clearly, under the reasonable assumption that arrivals are independent of the ED state, these indicators are two faces of the same phenomenon: the smaller the average length of stay of patients, the smaller the average number of patients in the ED. However, while the number of patients can be measured for each instant, thus being a meaningful figure even for small time intervals, the average length of stay is typically some hours; thus, its value is meaningful only when a sufficiently large time interval is considered. As the simulated interval (search depth) ΔT is much smaller than the average length of stay, within the DSS we use the average number of patients in the simulated interval, computed as

$$c_{avg} = \frac{\int_{\Delta T} c(t)dt}{\Delta T},$$

where $c(t)$ is number of patients in the ED at time t. Our DSS is aimed at minimizing this internal KPI through policy selection.

In order to evaluate the accuracy in selecting the best policy in the portfolio of available strategies, we designed a second set of experiments using the same setting (100 days and four snapshots) as that in the previous section. In particular, for each day and snapshot, we first run the DSS for each policy and determine the best one according to the internal KPI; then, we "run" the real ED (i.e., we run its digital twin) in the same setting, and determine the best policy for the real system. In

Table 10
Accuracy of each approach in identifying the best policy.

ΔT	Approach 1		Approach 2		Approach 3	
	% Det.	% Stoc.	% Det.	% Stoc.	% Det.	% Stoc.
1	73.00	70.00	72.50	76.00	75.00	82.00
2	73.50	71.00	75.50	71.50	77.50	82.00
3	62.00	67.50	69.50	62.50	68.50	73.50

Table 11
Δz vs. ΔT, Det. configuration.

ΔT / Δz	Real	30 min	1 h	2 h	3 h	4 h
30 min	44.36	38.58	37.85	**37.24**	41.23	42.40
1 h	44.36	37.14	37.17	36.20	35.46	<u>**34.40**</u>
2 h	44.36	37.30	38.71	37.15	**35.34**	35.80
3 h	44.36	37.26	38.77	35.80	**34.58**	35.59
4 h	44.36	41.44	42.52	41.68	40.82	**38.79**

Table 12
Δz vs. ΔT, Stoc. configuration.

ΔT / Δz	Real	30 min	1 h	2 h	3 h	4 h
30 min	44.36	38.20	39.51	**37.61**	43.86	42.06
1 h	44.36	37.44	<u>**34.60**</u>	37.50	35.35	35.63
2 h	44.36	38.37	37.74	**35.40**	36.44	35.93
3 h	44.36	37.73	38.31	36.63	**35.23**	35.44
4 h	44.36	41.47	42.68	42.52	46.35	**38.23**

our tests, we assume that the policy maker of our DT always accepts and follows the policy suggested by the DSS. Finally, we count the number of times in which the resulting strategies coincide, meaning that the DSS was able to identify the best policy for the real ED. Table 10 reports the results obtained with different values of $\Delta T = \{1, 2, 3\}$ hours for both the deterministic and the stochastic versions of the DSS. As may be expected, performances get worse when increasing the value of ΔT. In addition, the results confirm that Approach 3 in the stochastic operating mode provides the best policy prediction, with more than 80% of success for 1 and 2 h.

5.4.3. Improving performance

In this section we present the experiments performed in order to evaluate the capability of the DSS to improve the ED performance. For this aim, we set-up an experimental environment replicating the configuration depicted in Fig. 4, where we replaced the real ED by its DT. The system was populated with data of the patients that arrived between 01/10/2019 and 15/10/2019, an interval which was excluded from the previous experiments. We tested the DSS for the whole period of 15 days, thus evaluating the potential cumulative effect of decision. Following the indication provided by the previous experiments, the DSS embedded Approach 3 for the predictor.

Our first order of business is to determine the best setting for parameters ΔT and Δz. To this aim, we tested the DSS in both the deterministic and stochastic configurations, for different values of ΔT and Δz. Tables 11 and 12 show the results in terms of mean number of patients within the ED during the 15 days. The best value for each Δz is shown in bold, while the best overall is underlined.

The tables show that, in both configurations, the best results are obtained when $\Delta T \geq \Delta z$, although in all cases the DSS is able to reduce the mean number of persons compared to the real ED.

Fig. 8 plots the number of patients in the real ED and the same figure obtained through the DSS, in the best Deterministic configuration (ΔT = 4h, Δz = 1h), and in the best Stochastic configuration (ΔT = 1h, Δz = 1h).

Summarized statistics, also reporting the average Length of Stay and Waiting Time for the first visit (WT), can be found in Table 13. The results confirm that, although internally optimizing the mean number of patients, the DSS also considerably improves over the real ED for what concerns the LoS of patients. Although to a lesser extent, an improvement can also be achieved on the patients' WT. In Table 14 the mean and median LoS of patients are disaggregated by urgency code, showing that introducing the DSS has a major impact on the low priority patients, without affecting the (already short) LoS of urgent ones. The same statistic with reference to WT is shown in the Table 15 and again shows a considerable improvement for the less urgent codes.

In order to assess the quality of our results, we evaluate the average number of patients within the ED obtained by applying the (ideal) policy which can move patients through different queues, as many times as needed and with null transfer time (FQR). By using this ideal policy, figure C_{avg} reduces from 34.40 to 33.02, showing that only a marginal improvement could be obtained by a redefinition of the hospital layout and logistics.

Finally, we run the DSS for each day of the period individually, for both the Deterministic and Stochastic configurations in their best settings. Table 16 shows the mean number of patients within the ED, highlighting the best figure each day. In all but two cases the DSS provides a considerable improvement; in the remaining two days, the DSS results are only slightly worse than the historical ones, thus confirming the robustness of the proposed approach.

Table 13
15 days analysis results.

Indicator	Real	Det.	Stoc.	Det. gain %	Stoc. gain %
C_{avg}	44.36	34.40	34.60	−22.45	−22.00
LoS [min]	322	263	268	−18.29	−16.80
WT [min]	146	128	139	−12.68	−4.86

Table 14
LoS performance in minutes on the basis of the patients' urgency.

Urgency code	Real LoS [min]		Det. LoS [min]		Stoc. LoS [min]	
	Mean	Median	Mean	Median	Mean	Median
1	173	144	174	146	176	158
2	261	224	242	203	245	206
3	385	356	290	276	295	288
4	379	280	273	263	290	262

Table 15
Waiting time performance in minutes on the basis of the patients' urgency.

Urgency code	Real W. T. [min]		Det. W.T. [min]		SM W.T. [min]	
	Mean	Median	Mean	Median	Mean	Median
1	9	6	2	1	3	1
2	35	33	22	21	30	28
3	223	197	201	191	218	195
4	312	176	289	200	291	201

Table 16
Daily analysis results.

Day	Real	Det.	Stoc.	Det. gain %	Stoc. gain %
2019-10-01	50.28	**47.34**	47.55	−5.85	−5.43
2019-10-02	47.78	**43.12**	43.40	−9.75	−9.17
2019-10-03	49.78	**43.97**	45.45	−11.67	−8.70
2019-10-04	43.24	**39.82**	41.87	−7.91	−3.17
2019-10-05	38.25	34.17	**33.70**	−10.67	−11.90
2019-10-06	48.24	**46.88**	48.94	−2.82	1.45
2019-10-07	**62.54**	62.88	62.99	0.54	0.72
2019-10-08	32.13	**27.98**	28.77	−12.92	−10.46
2019-10-09	26.28	**21.25**	22.16	−19.14	−15.68
2019-10-10	33.03	**28.17**	29.47	−14.71	−10.78
2019-10-11	39.69	30.51	**30.47**	−23.13	−23.23
2019-10-12	42.58	33.68	**33.67**	−20.90	−20.93
2019-10-13	52.15	**46.25**	49.02	−11.31	−6.00
2019-10-14	**63.44**	63.84	63.60	0.63	0.25
2019-10-15	36.15	**33.24**	33.63	−8.05	−6.97

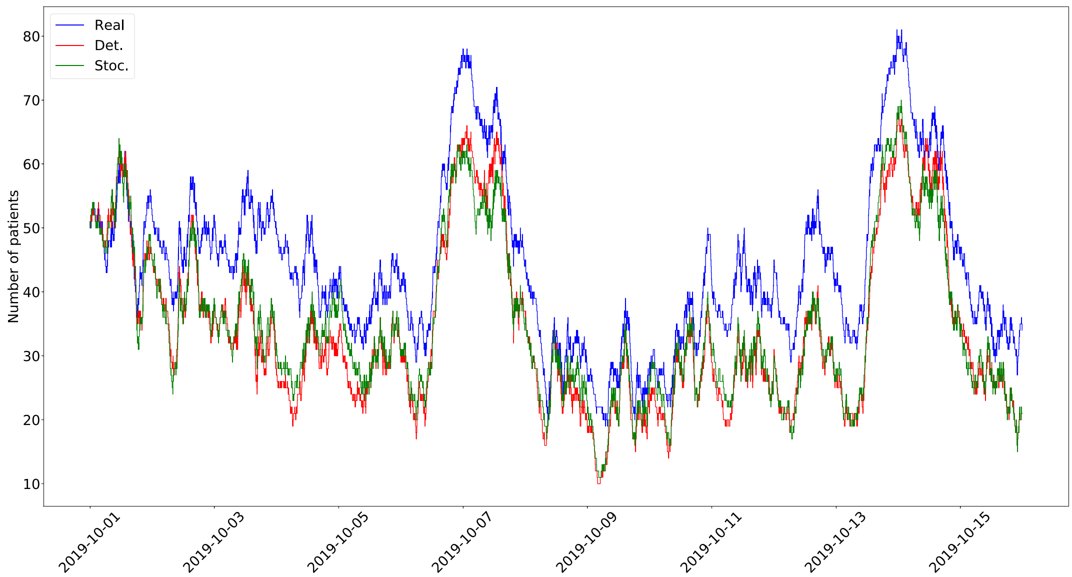


Fig. 8. Comparison between the real trend of the patients present and those obtained with DSS.

6. Conclusions

In this paper we described a decision support system (DSS) to improve the performance of an ED by addressing the serious problem of overcrowding. This complex task involves a non-univocal definition of the metrics to be considered, and the management of stochastic events.

The DSS includes 4 main elements, namely a predictor for patients arrivals, a predictor for the patients pathways, a predictor for activities duration, and a discrete-event simulator. The DSS includes different policies that are dynamically tested, and identifies the one that provides the best expected performance for the actual ED state. In addition to structured information collected at triage, the predictor for the patients pathways also exploits unstructured information from the nurse's diary, processed through a Natural Language Processing module.

An experimental application of the DSS to a digital twin of a major real ED in northern Italy has demonstrated that a dynamic selection of the best policy, among a limited set of simple alternatives, allows a relevant reduction of the number of patients within the ED, as well as a noticeable reduction of the Length of Stay of low priority patients.

The presented DSS is ready for being used in the ED of our case study. In addition, as the prediction-simulation modules implement a quite general framework, we expect the adaptation of the DSS to other EDs to require limited effort.

A first line for future development concerns the evaluation of the results obtained by assuming a partial rejection of the tool "suggestions" by the decision-maker. This would make it possible to assess the *rejection rate* under which the ED performance could still be improved by the DSS.

The study on the selection rate of each policy is another interesting area for improvement. This indicator and the organizational patterns in which a given strategy is selected would make it possible to reduce the set of strategies to be tested by excluding those that historically showed less effective in that condition.

Finally, expanding the set of available policies, including, e.g., the *First Consultation Priority Rule* or the *Second Consultation Priority Rule* [40,41], appears to be an other promising direction of future research, still reminding that explicability and acceptance by human decision-makers is a relevant issue. One may wonder if our DSS would still be able to identify the best policy in case of a larger policy pool. Since our DSS is based on predicting arrivals, service packages, and service duration and then inferring the performance of each policy via simulation, we do not expect the introduction of additional policies to significantly affect the system ability to make recommendations.

In a more challenging perspective, instead of expanding the policy pool with predefined policies, it is even possible to enable the system itself to discover policies through learning.

CRediT authorship contribution statement

Cristiano Fabbri: Writing – review & editing, Writing – original draft, Software, Conceptualization. **Michele Lombardi:** Writing – review & editing, Writing – original draft, Software, Conceptualization. **Enrico Malaguti:** Writing – review & editing, Writing – original draft, Software, Conceptualization. **Michele Monaci:** Writing – review & editing, Writing – original draft, Software, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Cristiano Fabbri reports a relationship with IRCCS Azienda Ospedaliero-Universitaria di Bologna that includes: employment.

Data availability

The data that has been used is confidential.

Acknowledgments

Enrico Malaguti and Michele Monaci were supported by the PNRR National Recovery and Resilience Plan: National Research Center in High Performance Computing, Big Data and Quantum Computing. All authors are grateful to two anonymous reviewers, for their constructive comments and remarks which considerably improved the presentation and the content of the paper.

References

[1] Leo G, Lodi A, Tubertini P, Di Martino M. Emergency department management in Lazio, Italy. Omega 2016;58:128–38.

[2] FitzGerald G, Jelinek GA, Scott D, Gerdtz MF. Emergency Department triage revisited. Emerg Med J : EMJ 2010;27(2):86–92.

[3] Christ M, Grossmann F, Winter D, Bingisser R, Platz E. Modern triage in the emergency department. Dtsch Ärzteblatt Int 2010;107(50):892.

[4] Vance J, Sprivulis P. Triage nurses validly and reliably estimate emergency department patient complexity. Emerg Med Australas 2005;17(4):382–6.

[5] Bartenschlager CC, Grieger M, Erber J, Neidel T, Borgmann S, Vehreschild JJ, Steinbrecher M, Rieg S, Stecher M, Dhillon C, et al. Covid-19 triage in the emergency department 2.0: how analytics and ai transform a human-made algorithm for the prediction of clinical pathways. Health Care Manage Sci 2023;26(3):412–29.

[6] Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA. A conceptual model of Emergency Department crowding. Ann Emerg Med 2003;42(2):173–80.

[7] Higginson I. Emergency department crowding. Emerg Med J : EMJ 2012;29(6):437.

[8] Weiss S, Derlet R, Arndahl J, Ernst A, Richards J, Fernandez-Frackelton M, Schwab R, Stair T, Vicellio P, Levy D, Brautigan M, Johnson A, Nick T, Fernández-Frankelton M. Estimating the degree of emergency department overcrowding in academic medical centers: Results of the national ED overcrowding study (NEDOCS). Acad Emerg Med : Off J Soc Acad Emerg Med 2004;11:38–50.

[9] Wang H, Robinson RD, Bunch K, Huggins CA, Watson K, Jayswal RD, White NC, Banks B, Zenarosa NR. The inaccuracy of determining overcrowding status by using the National ED Overcrowding Study Tool. Am J Emerg Med 2014;32(10):1230–6.

[10] Welch S, Savitz L. Exploring strategies to improve emergency department intake. J Emerg Med 2012;43(1):149–58.

[11] Aringhieri R, Bruni M, Khodaparasti S, van Essen J. Emergency medical services and beyond: Addressing new challenges through a wide literature review. Comput Oper Res 2017;78:349–68.

[12] Grot M. Decision support framework for tactical emergency medical service location planning. Omega 2024;125:103036.

[13] Lagoe RJ, Jastremski MS. Relieving overcrowded emergency departments through ambulance diversion. Hosp Top 1990;68(3):23–7.

[14] Aboueljinane L, Sahin E, Jemai Z. A review on simulation models applied to emergency medical service operations. Comput Ind Eng 2013;66(4):734–50.

[15] Rebuge A, Ferreira DR. Business process analysis in healthcare environments: A methodology based on process mining. Inf Syst 2012;37(2):99–116, management and Engineering of Process-Aware Information Systems.

[16] Mans R, van der Aalst W, Vanwersch R. Process mining in healthcare: evaluating and exploiting operational healthcare processes. SpringerBriefs in business process management, Germany: Springer; 2015.

[17] Duma D, Aringhieri R. An ad hoc process mining approach to discover patient paths of an Emergency Department. Flex Serv Manuf J 2020;32:6–34.

[18] Simeunovic N, Kamenko I, Bugarski V, Jovanovic M, Lalic B. Improving workforce scheduling using artificial neural networks model. Adv Prod Eng Manage 2017;12(4):337–52.

[19] Schiele J, Koperna T, Brunner JO. Predicting intensive care unit bed occupancy for integrated operating room scheduling via neural networks. Naval Res Logist 2021;68(1):65–88.

[20] Davies R, Davies H. Modelling patient flows and resource provision in health systems. Omega 1994;22(2):123–31.

[21] Günal MM, Pidd M. Understanding target-driven action in Emergency Department performance using simulation. Emerg Med J : EMJ 2009;26(10):724–7.

[22] Paul SA, Reddy MC, DeFlitch CJ. A systematic review of simulation studies investigating emergency department overcrowding. Simulation 2010;86(8–9):559–71.

[23] Jun JB, Jacobson SH, Swisher JR. Application of discrete-event simulation in health care clinics: A survey. J Oper Res Soc 1999;50(2):109–23.

[24] Azcarate C, Esparza L, Mallor F. The problem of the last bed: Contextualization and a new simulation framework for analyzing physician decisions. Omega 2020;96:102120.

[25] Liu Z, Rexachs D, Epelde F, Luque E. An agent-based model for quantitatively analyzing and predicting the complex behavior of emergency departments. J Comput Sci 2017;21:11–23.

[26] Duma D, Aringhieri R. Real-time resource allocation in the emergency department: A case study. Omega 2023;117:102844.

[27] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.

[28] Bird S, Klein E, Loper E. Natural language processing with python: analyzing text with the natural language toolkit. O'Reilly Media, Inc; 2009.

[29] Devlin J, Chang M-W, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018, http://dx.doi.org/10.48550/ARXIV.1810.04805.

[30] Schweter S. Italian BERT and ELECTRA models. 2020, http://dx.doi.org/10.5281/zenodo.4263142, version: 1.0.1.

[31] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. Pytorch: An imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché Buc F, Fox E, Garnett R, editors. In: Advances in neural information processing systems, Vol. 32, Curran Associates, Inc.; 2019, p. 8024–35.

[32] Szandała T. Review and comparison of commonly used activation functions for deep neural networks. Bio-inspired Neurocomput 2021;203–24.

[33] Bell L, Wagner R. Modeling emergency room arrivals using the Poisson process. College Math J 2019;50(5):343–50.

[34] SimPy Community. Simpy: Discrete event simulation for Python. 2002, https://simpy.readthedocs.io.

[35] Italian Ministry of Health. Emergency department and DEA classification. 2008, https://www.salute.gov.it/portale/prontoSoccorso/dettaglioContenutiProntoSoccorso.jsp?lingua=italiano&id=1190&area=118%20Pronto%20Soccorso&menu=vuoto, accessed: 02.05.2022.

[36] Gilligan P, Winder S, Singh I, Gupta V, Kelly P, Hegarty D. The boarders in the emergency department (bed) study. Emerg Med J 2008;25(5):265–9.

[37] Mohsin M, Forero R, Ieraci S, Bauman AE, Young L, Santiano N. A population follow-up study of patients who left an emergency department without being seen by a medical officer. Emerg Med J 2007;24(3):175–9.

[38] Weiss SJ, Ernst AA, Derlet R, King R, Bair A, Nick TG. Relationship between the national ed overcrowding scale and the number of patients who leave without being seen in an academic ed. Am J Emerg Med 2005;23(3):288–94.

[39] Jaccard P. Étude comparative de la distribution florale dans une portion des alpes et des jura. Bull Soc Vaudoise Sci Nat 1901;37:547–79.

[40] Cildoz M, Mallor F, Ibarra A. Analysing the ED patient flow management problem by using accumulating priority queues and simulation-based optimization. In: 2018 winter simulation conference. WSC, IEEE; 2018, p. 2107–18.

[41] Cildoz M, Ibarra A, Mallor F. Accumulating priority queues versus pure priority queues for managing patients in emergency departments. Oper Res Health Care 2019;23:100224.