



Effective forecasting of key features in hospital emergency department: Hybrid deep learning-driven methods



Fouzi Harrou ^{a,*}, Abdelkader Dairi ^b, Farid Kadri ^c, Ying Sun ^a

^a King Abdullah University of Science and Technology (KAUST), Computer, Electrical and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal 23955-6900, Saudi Arabia

^b Ecole Nationale Polytechnique Oran, Laboratoire des Technologies de l'Environnement LTE, BP 1523 Al M'naouar ENP Oran, 31000, Oran, Algeria

^c Aeroline Data & CET, Agence 1031, Sopra Steria Group, 31770, Colomiers, France

ARTICLE INFO

Keywords:

Hospital systems
Patient flows
ED visits
Forecasting
Hybrid deep learning methods

ABSTRACT

Forecasting the different types of emergency department (ED) demands (patient flows) in hospital systems much aids ED managers in looking into various options to appropriately allocating the restricted resources available per patient attendance. Deep learning networks have recently gained great success in modeling time-dependent in time series data. Thus, this work advocates the use of deep learning-driven models for patient flows forecasting. Notably, we examine and compare seven deep learning models, Deep Belief Network (DBN), Restricted Boltzmann machines (RBM), Long Short Term Memory (LSTM), Gated recurrent unit (GRU), combined GRU and convolutional neural networks (CNN-GRU), LSTM-CNN, and Generative Adversarial Network based on Recurrent Neural Networks (GAN-RNN), to forecast patient flow in a hospital emergency department. We introduce a forecaster layer as output for each model to enable traffic flow forecasting. Patient flow data from different ED services, including biology, radiology, scanner, and echography, in Lille regional hospital in France, is used as a case study in assessing the considered forecasting models. Four metrics of effectiveness are adopted for evaluating and comparing the forecasting methods. The results show the promising performance of deep learning models for ED patient flow forecasting compared to shallow methods (i.e., ridge regression and support vector regression). In addition, the results highlighted the superior performance of the DBN compared to the other models by achieving an averaged mean absolute percentage error of around 4.097% and R² of 0.973.

1. Introduction

Over the last decades, there has been an expanding demand for emergency department (ED) cares, including medical and surgical treatments worldwide (Aboagye-Sarfo et al., 2015; Boujemaa, Jebali, Hammami, Ruiz, & Bouchriha, 2018; Harrou, Dairi, Kadri and Sun, 2020; Kadri, Harrou, Chaabane and Tahon, 2014). Efficiently managing healthcare systems can significantly improve resources management in emergency departments (EDs) at a hospital where the number of visits is unpredictable (Harrou, Kadri, Sun, & Khadraoui, 2021; Kadri, Chaabane and Tahon, 2014). The successful management of EDs is particularly sensitive because they are expected to provide immediate and often lifesaving of patients. In the US, in the period from 1993 to 2003, the EDs visits considerably raised by an average of 26%; however, the number of EDs lowered by approximately 9% (Kellermann, 2006; Warden et al., 2006). For instance, the demand for EDs services has been doubled within the period 1990 and 2014, and it is still continuously increasing (Kadri, Harrou et al., 2014).

To manage the unexpected patient flows, the EDs require significant resources to prevent placing the medical staff under stress that could considerably complicate their work, but the available resources are limited (Kadri, Harrou et al., 2014). It has been reported in previous studies that patient influx could create stressful circumstances that affect the efficiency of staff in ED (Harrou, Kadri, Chaabane, Tahon, & Sun, 2015; Kadri, Chaabane et al., 2014; Kadri, Harrou et al., 2014). Thus, a precise forecast of patient demands in ED is necessary to mitigate overcrowding, decrease stress, ensure quality, and efficiently allocate resources.

There are various proposed methods in the literature to enhance the forecasting accuracy of ED patient flow. Times series methods are commonly employed to forecast ED visits (Afifal et al., 2016; Bergs, Heerinckx, & Verelst, 2014; Jones et al., 2009; Kadri, Harrou et al., 2014; Lin, 1989). For example, in Carvalho-Silva, Monteiro, de Sá-Soares, and Dória-Nóbrega (2018), a seasonal autoregressive moving

* Corresponding author.

E-mail addresses: fouzi.harrou@kaust.edu.sa (F. Harrou), abdelkader.dairi@univ-usto.dz (A. Dairi), farid.kadri@soprasteria.com (F. Kadri), ying.sun@kaust.edu.sa (Y. Sun).

average (SARMA) model has been implemented for ED visits forecasting at the Braga Hospital in Portugal. [Martín and Cáceres \(2005\)](#) introduced an approach using time-series models for predicting hourly patient arrival to Tenerife ED, Spain, for the period from 1997 to 2002. Authors in [Kadri, Harrou, and Sun \(2017\)](#) applied a multivariate ARIMA approach to model patient flow in Lille hospital, France. The study conducted in [Champion et al. \(2007\)](#) presented exponential smoothing and ARIMA to predict monthly ED demands in a hospital in Victoria, Australia, from 2000 to 2005. [Barişçi \(2008\)](#) analyzed the electromyographic (EMG) signals by using an adaptive autoregressive moving average model (ARMA). In [Mohamed and Mohamad \(2020\)](#), a forecasting method based on ARIMA is considered to forecast patient visits in an orthopedic clinic in Kuantan using historical data collected from January 2013 until June 2018. In [Zhang et al. \(2020\)](#), Zhang et al. investigated and compared three linear models (i.e., ARIMA, least absolute shrinkage, and Lasso) and three nonlinear models (i.e., linear-and-radial SVR, random forests, and adaptive boosting) to forecast patient visits in the radiology department of West China Hospital in Chengdu, China. They showed that Lasso outperforms the other models in terms of mean absolute percentage error. [Sun, Heng, Seow, and Seow \(2009\)](#) predicted the ED visits by employing ARMA techniques. The authors highlighted that the time-series models represent a helpful tool for patient flow prediction at ED and the workload of medical staff. [Aboagye-Sarfo et al. \(2015\)](#) considered vector-ARMA models for ED demand forecasting in Western Australia. The authors gathered data on monthly visiting patients from 2006 to 2012 from public hospital ED in Western Australia.

However, it is worth noticing that traditional time-series methods could lead to a satisfying performance when applied to data with regular variations. Still, the forecast accuracy is degraded if time series exhibit irregular fluctuations. To alleviate this deficiency, non-parametric methods and machine learning algorithms are employed in developing patients flow forecasting ([Benbelkacem, Kadri, Atmani, & Chaabane, 2019](#); [Handly, Thompson, Li, Chuirazzi, & Venkat, 2015](#); [Harrou, Sun, Hering, Madakyaru et al., 2020](#); [Xu, Wong, & Chin, 2013](#)). For instance, the technique in [Khatri and Tamil \(2017\)](#) used an artificial neural network for predicting the days with a peak demands of patients having respiratory diseases attending EDs in Dallas County of Texas. A neural network approach is combined with a regression model in [Gul and Guneri \(2016\)](#) to improve ED patient visits forecasting. Similarly, in [Jiang et al. \(2019\)](#), a deep learning model has been combined with regression models to improve the prediction of patient flow in Hong Kong. Yu et al. introduced an integrated method combining the advantages of wavelet decomposition (WD) and ANN to improve forecasting of ED visits ([Yu, Hang, Tang, Zhao, & Lai, 2017](#)). Specifically, the WD is applied to decompose monthly ED visits data into multiple components and residuals. Then, ANN is applied to every decomposed component for forecasting, and then the final forecast is obtained by a simple joining procedure. Besides the advantages of machine learning techniques, they are well employed in diverse healthcare sectors, including healthcare coverage, hospital length of stay, care demands, and COVID-19 prediction ([Benbelkacem et al., 2019](#); [Daghistani et al., 2019](#); [Ichikawa, Saito, Ujita, & Oyama, 2016](#); [Ma, Yu, Ye, Yao, & Zhuang, 2020](#); [rekha Hanumanthu, 2020](#); [Xie et al., 2016](#)).

Recently, with the advancement of deep learning technologies, several techniques have been designed to enhance the forecasting of time-series data in different fields of applications ([Dairi, Cheng, Harrou, Sun, & Leiknes, 2019](#); [Dairi, Harrou, Khadraoui, & Sun, 2021](#); [Guo, Zhou, Zhang, & Yang, 2018](#); [Mai, Tian, Lee, & Ma, 2019](#); [Maldonado & Harabagiu, 2019](#); [Pham, Tran, Phung, & Venkatesh, 2017](#); [Saba et al., 2019](#)). The core insight behind deep learning models is concatenating several layers into the neural network structures to better learn complex information from data ([Harrou, Sun et al., 2020](#); [Zeroual, Harrou, Dairi, & Sun, 2020](#)). The principal singularity of deep learning methodologies is their ability to extract pertinent information from large data in an

automatic way without any features engineering ([Zeroual et al., 2020](#)). In addition, they demonstrated to be effective in handling distinct kinds of datasets, such as time-series, videos, images, and audio ([Dairi, Harrou, Sun, & Senouci, 2018](#); [Harrou, Hittawe, Sun and Beya, 2020](#); [Wang, Lee, Harrou, & Sun, 2020](#)). Karsant et al. considered a Long Short Term Memory (LSTM) model for forecasting monthly visits in a hospital in South Tangerang City ([Karsanti, Ardiyanto, & Nugroho, 2019](#)). Results demonstrated that the LSTM approach excels ARIMA, linear regression, simple exponential smoothing, and traditional ANN. Similarly, Kadri et al. used an LSTM-based deep learning model to predict hospital ED visits ([Kadri, Baraoui, & Nouaouri, 2019](#)).

Overcrowding in EDs remains a primary concern for hospital administration. Managing patient flow has become a critical concern in most hospital systems around the world. Accurately forecasting patients visiting different departments at the hospital, such as radiology, biology, scanner, and echography, is of great importance for improved resource allocation, optimized appointment scheduling decisions, and more strategic decision making are in high demand ([Ganguly & Nandi, 2016](#); [Harrou et al., 2015](#); [Kadri, Harrou et al., 2014](#); [Kadri et al., 2013](#); [Kellermann, 2006](#); [Kolker, 2008](#); [Zeroual et al., 2020](#)). This study investigates and compares seven deep learning models named Deep Belief networks (DBN), Restricted Boltzmann machines (RBM), Long short-term memory (LSTM), and gated recurrent unit (GRU) the combined GRU and convolutional neural networks (CNN-GRU), CNN-LSTM, and the Generative Adversarial Network based on Recurrent Neural Networks (GAN-RNN), as well as two baseline models (i.e., Support Vector Regression (SVR) and ridge regression (RR)) to forecast patient flow in a hospital emergency department. The DBN, RBM, LSTM, and GRU were selected from non-hybrid models, whereas CNN-GRU, CNN-LSTM, and GAN-RNN were chosen from the class of hybrid deep learning models. As far as we know, this is the first time these models have been introduced to enhance patient flow forecasting. It is worth highlighting that we introduce a forecaster layer as output for DBN and RBM models to enable the forecasting of ED visits time-series data. Moreover, to our best knowledge, GRU-CNN deep learning-driven forecasting has not been investigated in the current literature. Patient flow data from different pediatric ED services (i.e., radiology, biology, scanner, and echography) in Lille regional hospital in France, is used as a case study in assessing the considered forecasting models. The data is smoothed via exponential smoothing to reduce the outlier effect and to improve forecasting accuracy. Four metrics of effectiveness are adopted for evaluating and comparing the forecasting methods. Forecasting results revealed the promising performance of deep learning models. The rest of the article is organized subsequently. Section 1 briefly describes the basic features of studied models. In Section 2, the adopted forecasting framework is described. Section 3 presents the used patient flow datasets and analyses the forecasting results obtained from the considered models. The conclusion and future lines are drawn in Section 4.

2. Methodology

This section introduces firstly an overview of the four deep learning models, namely DBN, RBM, CNN-GRU, and GAN-RNN. Then presents the deep learning-driven patient flow forecasting procedure.

2.1. Deep belief networks

DBNs are probabilistic generative models formed of many hidden layers. DBNs consist of stacking Restricted Boltzmann machines (RBM) plus directed Sigmoid Belief Networks (see Fig. 1). The main advantage of the DBN model is its capability of learning complex and hierarchical features of a high order via layer-by-layer learning ([Erfani, Rajasegaran, Karunasekera, & Leckie, 2016](#)). This approach was introduced by [Hinton, Osindero, and Teh \(2006\)](#), which makes each layer able to learn a higher-level representation of the layer below ([Hinton et al., 2006](#)).

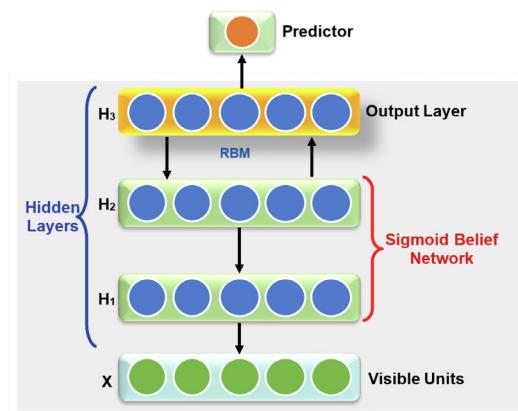


Fig. 1. DBN structure.

RBMs are trained separately through unsupervised learning using a greedy layer-wise procedure where the training starts from the lower level, namely the input layer X (observed data).

For a DBN composed of two hidden layers $\mathbf{h}_{(1)}, \mathbf{h}_{(2)}$, we define $P(\mathbf{x}, \mathbf{h}_{(1)}, \mathbf{h}_{(2)}; \Theta)$ as the joint distribution over \mathbf{x} , $\mathbf{h}_{(1)}$, and $\mathbf{h}_{(2)}$ expressed by DBN under the following form:

$$P(\mathbf{x}, \mathbf{h}_{(1)}, \mathbf{h}_{(2)}; \Theta) = P(\mathbf{x}|\mathbf{h}_{(1)}; \mathbf{W}_{(1)}) P(\mathbf{h}_{(1)}, \mathbf{h}_{(2)}; \mathbf{W}_{(2)}) \quad (1)$$

with \mathbf{W} weight matrix, and Θ the model parameter defined as $\Theta = \{\mathbf{W}_i\}^K$, K is the number of hidden layers in the DBN. The term $P(\mathbf{x}|\mathbf{h}_{(1)}; \mathbf{W}_{(1)})$ represents the directed sigmoid.

$$P(\mathbf{x}|\mathbf{h}; \mathbf{W}) = \text{sigmoid}(\sum_j \mathbf{W}_{ij} \mathbf{h}_j), \quad (2)$$

while the second term $P(\mathbf{h}_{(1)}, \mathbf{h}_{(2)}; \mathbf{W}_{(2)})$ express the joint distribution of the second-layer RBM.

$$P(\mathbf{h}_{(1)}, \mathbf{h}_{(2)}; \mathbf{W}_{(2)}) = \frac{1}{Z} e^{(\mathbf{h}_{(1)}^T \mathbf{W}_{(2)} \mathbf{h}_{(2)})} \quad (3)$$

Z represents the partition function (normalizing constant), and *sigmoid* is the logistic function. The model parameters θ are computed during the unsupervised learning phase, which is followed by a crucial step, namely the fine-tuning. The fine-tuning phase is essential, especially for optimizing the DBN parameters furthermore to improve the model performance. This phase uses the back-propagation (BP) algorithm, which is achieved via supervised learning, aiming to reach the global optimum utilizing labeled data for training the DBN.

DBN, as a deep learning model, has proved its ability to extract high-level abstractions through unsupervised learning, which makes it powerful data distribution approximators based on Gibbs sampling (Efani et al., 2016). DBNs have been applied to various research fields, mainly computer vision applications, as feature extractors. One significant advantage of DBN is its ability to learn an intricate hidden pattern and reproduce a similar sample as it is a generative model. Also, it can map a sequence to a value or sequence values, which is identical to the problem of univariate and multivariate forecasting even with multi-step forecasting problems.

In this paper, the DBN is employed PED visits time-series forecasting by introducing a forecaster layer at the output layer of the DBN (Fig. 1). Mainly, we used the forecaster layer for mapping the DBN output into a scalar data point, which makes it comparable to the original time-series data.

2.2. Restricted Boltzmann machine

RBM belong to famous Boltzmann machines (RBMs) energy-based, stochastic and generative models represented by an undirected fully connected graph. RBMs are shallow neural networks composed of visible

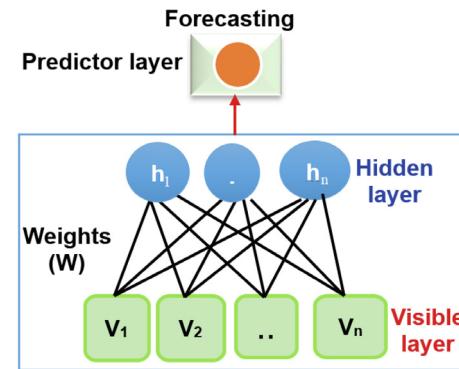


Fig. 2. RBM forecasting structure.

unit v and hidden units h , v is the input, and h represents a latent space representing extracted features from the input. RBMs aim to find model configurations that reduce the energy of the underlying system. RBMs are suitable for approximating unknown probability distribution. However, they present a high computation cost. A restricted version was proposed by Smolensky (1986) Restricted Boltzmann machines (RBMs) with no connections of the form: visible-to-visible and hidden-to-hidden. RBMs are trained using the Markov chain Monte Carlo (MCMC) method based on Gibbs sampling to obtain an estimator of the log-likelihood gradient (Bengio, 2009). Deep models like Deep Belief Networks (DBN) (Hinton et al., 2006) and the hierarchical probabilistic model deep Boltzmann machine (DBM) (Salakhutdinov & Hinton, 2009) can be built by stacking RBMs. The energy function representing the existing energy between visible v and hidden h units can be expressed (ignoring biases):

$$\text{Energy}(v, h) = - \sum_{i=1}^m \sum_{j=1}^n \mathbf{W}_{ij} v_i h_j, \quad (4)$$

with \mathbf{W} a weight matrix, $P(v)$ is the probability assigned by the model for a given visible unit v :

$$P(v) = \frac{\sum_h e^{-\text{Energy}(v,h)}}{\sum_{v,h} e^{-\text{Energy}(v,h)}}, \quad (5)$$

Thus mapping a visible vector v in the space of the hidden unit h is expressed:

$$P(h|v) = \prod_j p(h_j|v). \quad (6)$$

Then, sampling a new visible vector v from the space of the hidden unit h is expressed:

$$P(v|h) = \prod_i p(v_i|h), \quad (7)$$

RBM can be used as a probabilistic autoencoder, where the encoding process is achieved through inference from $P(h|v)$ and the decoding process in the other way via $P(v|h)$. Moreover, RBMs also can be stacked to build a stacked autoencoder.

Similarly to the DBN model, a forecaster layer is added to RBM as the output layer (Fig. 2); this enables using the RBM model to forecast daily visits at PED.

2.3. The hybrid GRU-CNN method

The gated recurrent unit (GRU) model is a deep learning model introduced to overcome the vanishing and explosion of gradients problem encountered with traditional recurrent neural networks during the training (Cho et al., 2014). Indeed, GRU is derived from the well-known Long Short Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997); both are based on gating mechanisms that regulate and store the information from the past that will impact and influence the future output.

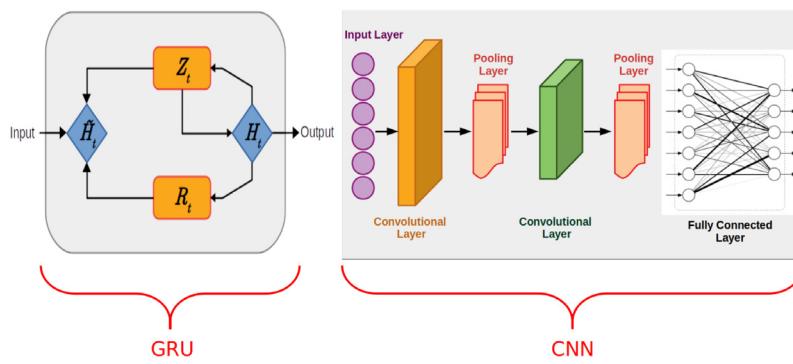


Fig. 3. Hybrid GRU-CNN structure.

The LSTM model was proposed with three gates (input, output, and forget); however, the GRU model comes with only two gates: reset gate and update gate, resulting in fewer parameters and good ability for time dependencies modeling. On the other hand, CNN, as a powerful tool for handling high-dimensional data, was widely applied to computer vision problems due to its high ability for features extraction and classification tasks from 2-dimensional (2D) input signals (LeCun, Bengio, & Hinton, 2015; Liu et al., 2019). CNN was adapted to deal with numerical 1D inputs, such as sensory data and time-series (LeCun et al., 2015).

The GRU is combined with CNN in a hybrid deep learning architecture called the GRU-CNN model (Fig. 3). In this hybrid model, GRU and CNN are stacked and trained together. The GRU output will feed the CNN. The forecasting task will be achieved by the CNN, which compacts the GRU temporal features space by extracting features and learn the mapping with the next value in the data sequence that was the input of GRU. Of course, the GRU-CNN is a hybrid architecture where we merge the desirable features of the GRU and the CNN models. This concatenation is dedicated to integrating the abilities of the GRU approach in modeling time series data and the capability of the CNN to extract important features from complex data.

2.4. The hybrid GAN-RNN method

Generative adversarial network (GAN) was introduced by Goodfellow et al. (2014), as a generative model able to learn data distribution via probability distribution approximation of a given input, and generate new data points, such as the generation of new images. The data generation process is performed through data sampling, which is useful for data augmentation of training datasets. The GAN structure is composed of two neural network models, which makes it different from traditional models based on only one model. This structure makes GANs very flexible. In contrast to the other generative models like RBM where the model play the roles of both generator (sampling) and discriminator (model and recognize the actual data), GAN delegate these two roles to a separate model called: generative and discriminative arranged in an adversarial way (Fig. 4). The discriminative model, D , learning is based on two sources of data: the training dataset (true data) and noisy data (simulated) produced by the generative model G . After the training process, the discriminator will be able to distinguish between true and simulated data.

GAN training optimize a cost function $\mathcal{V}(G, D)$ as:

$$\min_G \max_D \mathcal{V}(G, D) = \mathbb{E}_{p_{data}(\mathbf{x})} \log D(\mathbf{x}) + \mathbb{E}_{p_g(\mathbf{z})} \log (1 - D(G(\mathbf{z}))) \quad (8)$$

where $p_z(z)$ describe a prior on input noise variables and $p_{data}(x)$ represents the probability distribution of the true data. The distribution of samples is denoted by $p_g(x)$ produced by the generator. During the GAN training, the generator is encouraged to generate a distribution $p_g(x)$ similar to $p_{data}(x)$ of the real data. Indeed the generator helps the discriminator to classify new data points to (true, generated). After completing the training, the $p_g(x)$ becomes similar to $p_{data}(x)$ of the real

data, which is the distribution of the historical data learned during the training by generator and discriminator.

GANs are hybrid models consisting of a generative and discriminative model; those models can be any kind of neural network, such as recurrent networks (e.g., RNN, LSTM, or GRU), which makes GAN very flexible. For instance, in the GAN-RNN, the generative and discriminative models are recurrent neural networks, hence GAN-RNN. Indeed, the GANs are not designed for time-series modeling; however, their ability for the data distribution approximation makes them able to predict the next values of a given data sequence and thus can be investigated for improving the accuracy of forecasting. In this study, we have two GRUs arranged in an adversarial way in one architecture denoted by Generative is G-GRU, and the discriminative is D-GRU. The training procedure remains the same G-GRU feed D-GRU with noisy data combined with the training data to distinguish between real and generated data.

2.5. The proposed deep learning forecasting strategy

The PED visits time-series datasets are first preprocessed and then used to train deep learning models. This study adopted the exponential smoothing algorithm to smooth the PED visits time-series data, discard outliers, and improve data quality. The PED datasets are smoothed as follows:

$$\mathbf{e}_t = \alpha \mathbf{x}_t + (1 - \alpha) \mathbf{e}_{t-1}, \quad (9)$$

where \mathbf{x}_t denotes the number of PED demands, \mathbf{e}_t refers to the filter output, \mathbf{e}_0 is selected to be the starting point \mathbf{x}_0 and $\alpha \in [0, 1]$ denotes the smoothing parameter that defines the memory depth of the exponential smoothing scheme. Here, we use a large value of α (i.e., $\alpha = 0.7$) to slightly smooth the data and keep the most variance in the original data. Next, we normalize the smoothed data, \mathbf{e} , by min–max normalization within the interval $[0, 1]$.

$$\tilde{\mathbf{e}} = \frac{(\mathbf{e} - e_{min})}{(e_{max} - e_{min})} \quad (10)$$

where e_{min} and e_{max} refer respectively to the minimum and maximum of the smoothed measurements; a reverse procedure is allied after forecasting.

Here, we adopted three hybrid models (i.e., GRU-CNN, LSTM-CNN, and GAN-RNN), two non-hybrid models (i.e., DBN, RBM, LSTM, and GRU), and two base models (SVR and RR) to forecast daily patient flow recorded in different departments, such as radiology, biology, and echography. To this end, we first split the smoothed data into training and test sets. Fig. 5 shows the flowchart of the whole framework of the deep learning-based daily PED visits forecasting procedure. Firstly, the forecast models are established using a training set, and the parameters of each model are computed. For DBN and RBM, a new layer having only single output is incorporated for forecasting purposes. The whole structure of each model is trained for optimizing its parameters to get suitable performance. For instance, the training in the coupled DBN and

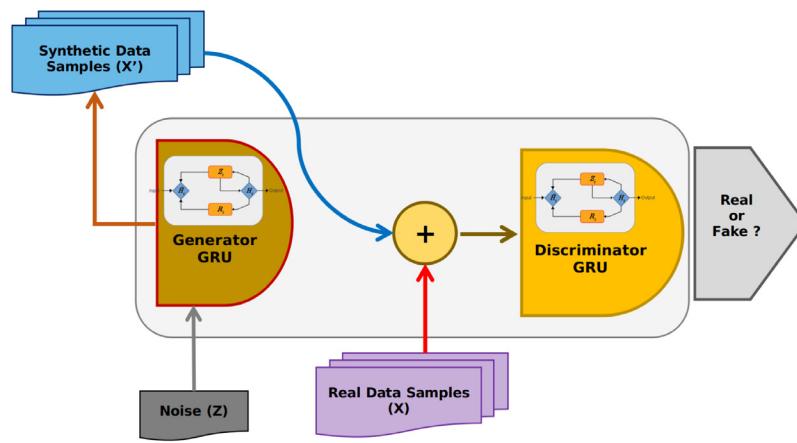


Fig. 4. The GAN-RNN structure.

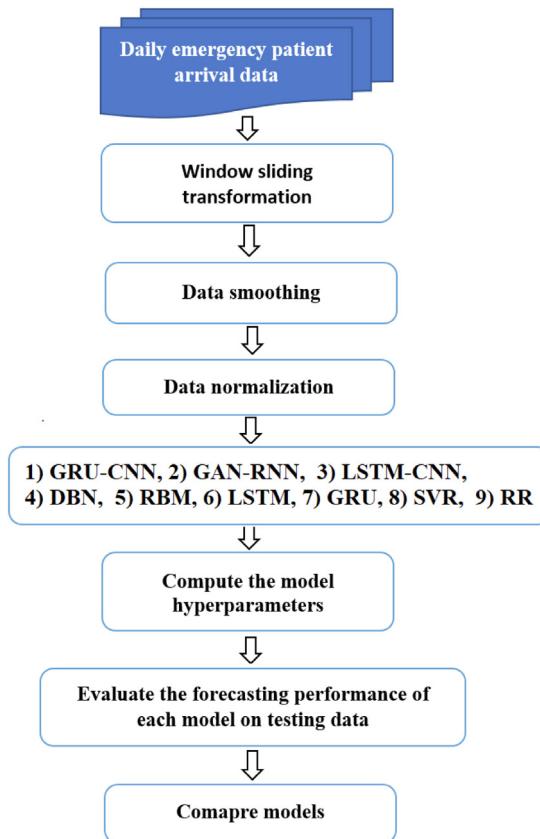


Fig. 5. Deep learning-driven daily PED visits forecasting procedure.

forecaster layer is performed in an unsupervised way. At first, the first layer is trained, and the latent variables at every layer are employed as the input for training the following layer. After this greedy layer-wise training, supervised fine-tuning is employed to tune weight matrices and bias of the network to minimize the loss function. Essentially, the crucial role of fine-tuning is adjusting the network parameters to minimize information losses.

2.6. Measurements of effectiveness

In this study, we assess the accuracy of the forecasting models using six commonly used statistical scores: Coefficient of determination (R^2), Root Mean Square Error (RMSE), mean absolute error (MAE),

Table 1
Definition of measurements of effectiveness.

Metric	Definition
R^2	$\frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})^2}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}$
RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
MAE	$\frac{\sum_{i=1}^n y_i - \hat{y}_i }{n}$
MAPE	$\frac{100}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \%$
EV	$1 - \frac{\text{Var}(\hat{y} - y)}{\text{Var}(y)}$
NRMSE	$\left(1 - \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \right) \cdot 100\%$

explained variance (EV), and the normalized RMSE (NRMSE) are used (Table 1). In Table 1, y_t refers to the number of PED demands, \hat{y}_t is its corresponding forecasted values. Of course, a more accurate model is characterized by higher R^2 , EV, MAPE, and NRMSE values and lower RMSE and MAE values.

3. Results and discussion

A daily patient visit dataset recorded from the PED in CHRU-Lille is used to assess the considered methods. The Lille hospital serves about four million residents in Nord-Pas-de-Calais in France, and its PED admits, on average, 23,900 demands yearly. The PED process care at the CHRU-Lille includes the following main stages (Fig. 6): (1) administrative registration, (2) hostess management, (3) nurse and medical consultations, (4) admission to the vital emergency room, (5) admission to the short-term hospitalization unit, (6) and additional examinations.

The patient care process within PED begins when the patient arrives and ends when the patient exits the PED (Fig. 6). When the patient does not present a vital emergency, he first goes through the administrative registration. Then he will be supported by the hostess before starting the nurse and medical consultations. If the patient presents a vital emergency, he will be directly admitted to a vital emergency room without administrative registration (in general, the administrative registration is after the treatment).

3.1. Data analysis

This study was carried out using datasets extracted from the pediatric emergency department (PED) database in Lille hospital center, France. The data used in this study are the time series of daily patient attendances at the PED, from January 2011 to December 2012. It concerns the seven following time series variables CMU1, CCMU2,

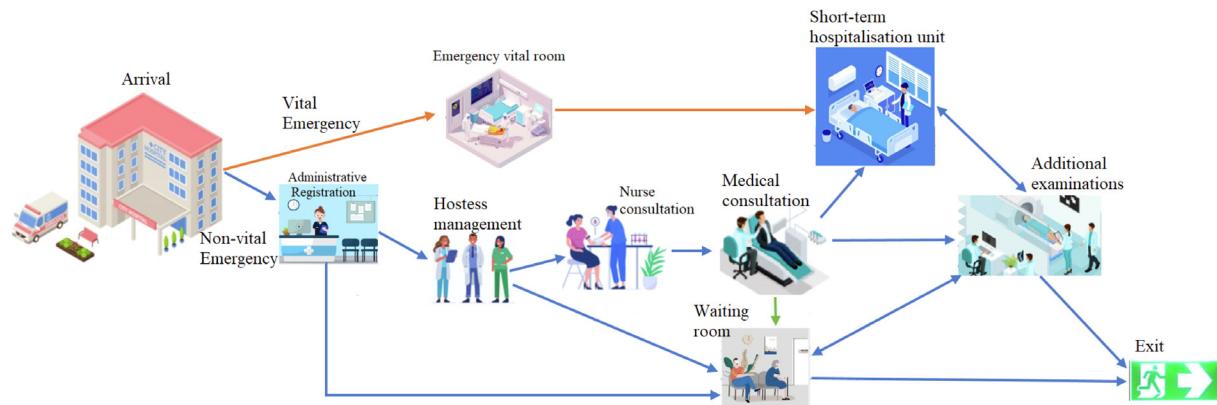


Fig. 6. Main stages of the PED process care at CHRU-Lille.

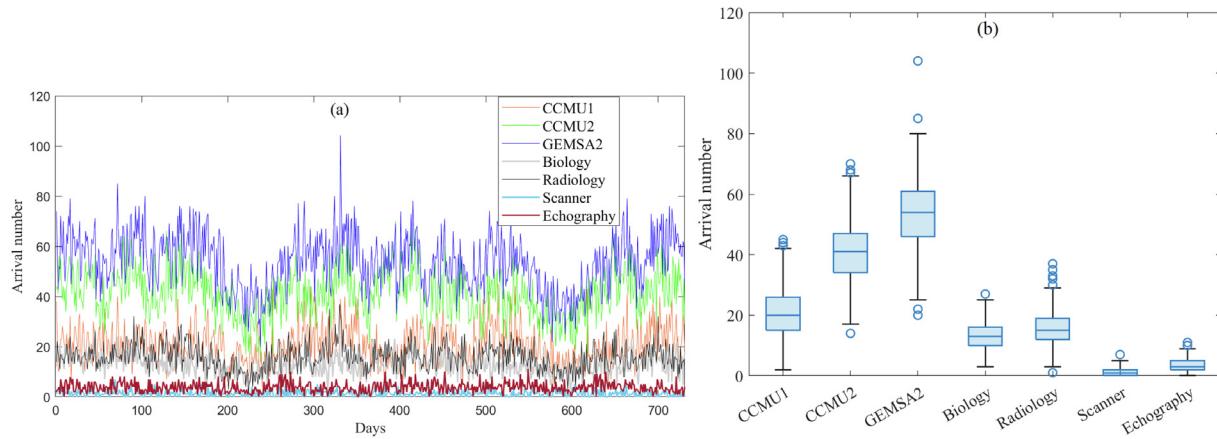


Fig. 7. (a) Daily PED arrivals from January 2011 to December 2012. (b) Box plots of PED visits time-series.

Table 2
Time series variables collected from the PED database at Lille hospital centre (France).

Variable	Definition
CCMU1	Daily number of non-urgent arrivals to the PED
CCMU2	Daily number of patient arrivals to the PED with a lesion state or stable functional prognosis; possibility of complimentary examination and therapeutic treatment (plaster, suture) by staff from the PED
GEMSA2	Daily number of unexpected patient arrivals, like patient not convened, and unexpected arrival to the PED
Biology	Daily number of patient arrivals for biology
Radiology	Daily number of patient arrivals for radiology
Scanner	Daily number of patient arrivals for Scanner
Echography	Daily number of patient arrivals for Echography

GEMSA2, Biology, Radiology, Scanner, and Echography, as shown in Table 2.

Fig. 7(a) depicts the daily visits at the PED of the time-series variables collected from January 2011 to December 2012. The box plots of the data in Fig. 7(b) are displayed in Fig. 7(b). We observe that the GEMSA2 time series, which is the daily number of unexpected visits to the PED, has large variability compared to the other datasets; CCMU2 and CCMU2, respectively, follow it. The other time series have relatively small variability (i.e., the boxplot is compact).

We compute the heatmap of the Pearson correlation to check the correlation among the six time-series (Fig. 8). We can see that there is a relatively weak correlation between GEMSA2, CCMU1, and CCMU2. For instance, some of the non-urgent visits to the PED (i.e., CCMU1) can arrive with no convocation (i.e., they can be categorized as GEMSA2). Also, we can observe a very weak correlation between the other variables.

The sample autocorrelation function (ACF) of the daily patient flow is depicted in Fig. 9. Fig. 9 reveals the presence of short-term autocorrelation in ED visits data and the absence of a periodic cycle.

3.2. Forecasting results

Modeling and predicting patient visits at the PED are relatively challenging because they depend on multiple factors, such as meteorological variations and epidemic events. Here, we explore the capacity of hybrid and non-hybrid deep learning models to forecast the number of daily visits to the emergency department at the PED of Lille hospital. Forecasting daily visits to different departments, including radiology, biology, and echography, are considered for optimizing the planning of nursing rosters and the management of available staff within the emergency department. Mainly, the forecasting performance of seven models, namely DBN, RBM, CNN-GRU, GAN-RNN, LSTM-CNN, LSTM, GRU, and two base-line methods, namely SVR and RR, are examined in forecasting patient visits from different departments at the Lille PED. To this end, we first train the nine models using training datasets collected from January 1, 2011, to May 7, 2012. The values of the tuned parameters of the trained models are listed in Table 3. These parameters are computed by the minimization of the cross-entropy of the reconstructed error during the training phase.

We also assessed the impact of the length of the input sequence on the forecasting output. To this end, we forecasted patient flow based on data of the previous 3 days, 6 days, 9 days, 12 days, 15 days, 18 days, and 21 days. Fig. 10 shows the barplot of the R^2 values obtained by the DBN model when applied to training data of CCMU1, CCMU2, and GEMSA2 series. The results from the other series are omitted because they all produce roughly similar outcomes. Results revealed that high

Table 3
Tuned parameters in the considered methods.

Model	Parameters
RR	alpha=1.0, solver='SVD'
SVR	kernel=rbf, nu=0.01,gamma=0.2
GRU	GRU(32, activation='relu') Dense(1, activation='sigmoid'), Epochs=1000, Loss=Cross Entropy, Optimizer=Rmsprop
LSTM	LSTM(32, activation='relu') Dense(1, activation='sigmoid'), Epochs=1000, Loss=Cross Entropy, Optimizer=Rmsprop
DBN	Layers: (visible layer: 7, hidden layers(32, 32)), Dense(1, activation='sigmoid'), Epochs=500, Loss=Cross Entropy, Optimizer=Rmsprop
RBM	Layers: (visible layer 7, hidden layer 32), Dense(1, activation='sigmoid'), Epochs=1000, Loss=Cross Entropy, Optimizer=Rmsprop
GRU-CNN	Layers: GRU(hidden units 32, activation='tanh'), CNN (Conv1D (32, activation='Relu'), Conv1D (32), MaxPooling1D(pool length=2), Dense(1, activation='sigmoid')), Epochs=1000, Loss=Cross Entropy, Optimizer=Rmsprop
LSTM-CNN	Layers: LSTM(hidden units 32, activation='tanh'), CNN (Conv1D (32, activation='Relu'), Conv1D (32), MaxPooling1D(pool length=2), Dense(1, activation='sigmoid')), Epochs=1000, Loss=Cross Entropy, Optimizer=Rmsprop
GAN-RNN	Generator: GRU(32, activation='relu'), BiGRU(32, activation='relu'), Dense (latentDim=16, activation='sigmoid'), Discriminator: GRU(64, activation='relu'), BiGRU(64, activation='relu'), Dense(latentDim=16, activation='sigmoid'), Dense(1, activation='sigmoid'), Epochs=500, Loss=Cross Entropy, Optimizer=Rmsprop

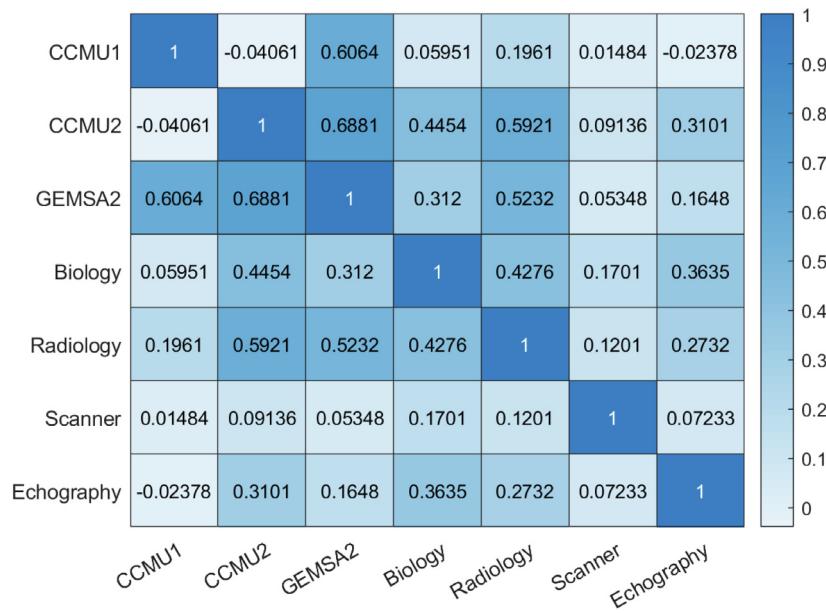


Fig. 8. The correlation matrix of the PED visits time-series data.

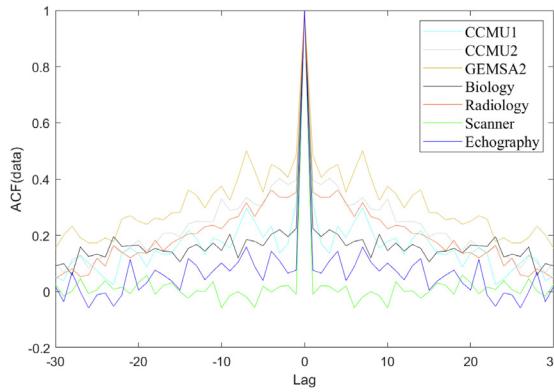


Fig. 9. Box plots of the seven time-series collected from the PED database.

forecasting accuracy in terms R^2 is obtained by using the previous three days to forecast the next day.

Reliable daily forecasting of ED demands offers helpful information to support medium-term planning (e.g., the designation of rotas).

Furthermore, it permits the ED managers to improve staff management settlements and obtain an earlier understanding of awareness situations. Essentially, precise forecasting is needed to give a guide for decisions. All models are built using the training set and then assessed using the testing set. In this work, we investigate the performance of deep learning-based methods on datasets of limited size. Fig. 11 displays the forecasting results from the trained models based on test data. Fig. 11 visually indicates that the investigated methods can follow the future trend of the patient visits data.

Now, we quantitatively examine the forecasting accuracy by computing five commonly used statistical scores based on the test data. The heat map of the statistical indicators for each time series is shown in Fig. 12. Firstly, we observe from Fig. 12 that the seven deep learning models are providing satisfying forecasting results. This means that the forecasted PED visits closely follow the recorded patients' visit trends. Secondly, we notice that the hybrid models GRU-CNN and GAN-RNN, LSTM, GRU, SVR, and RR provide unsuited results with R^2 of 0.734, 0.723, 0.56, 0.91, 0.754, and 0.755, respectively, for the scanner's daily PED visits. At the same time, the hybrid LSTM-CNN and non-hybrid, DBN, and RBM can achieve much better outcomes for all time series. Lastly, the DBN excels all the considered models in terms of the computed statistical scores (R^2 , EV, RMSE, MAE, and MAPE). This could be attributed to the greedy layer-wise learning procedure used to

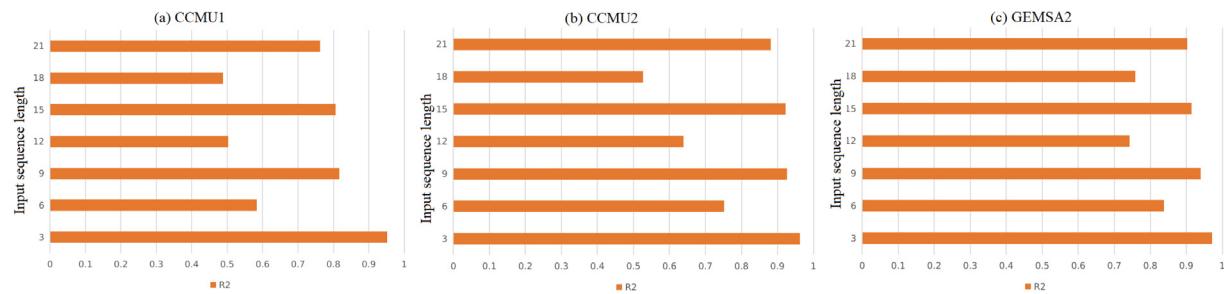


Fig. 10. DBN forecasting accuracy in function of the length of input sequence based on training set (a) CCMU1, (b) CCMU2, and (c) GEMSA2.

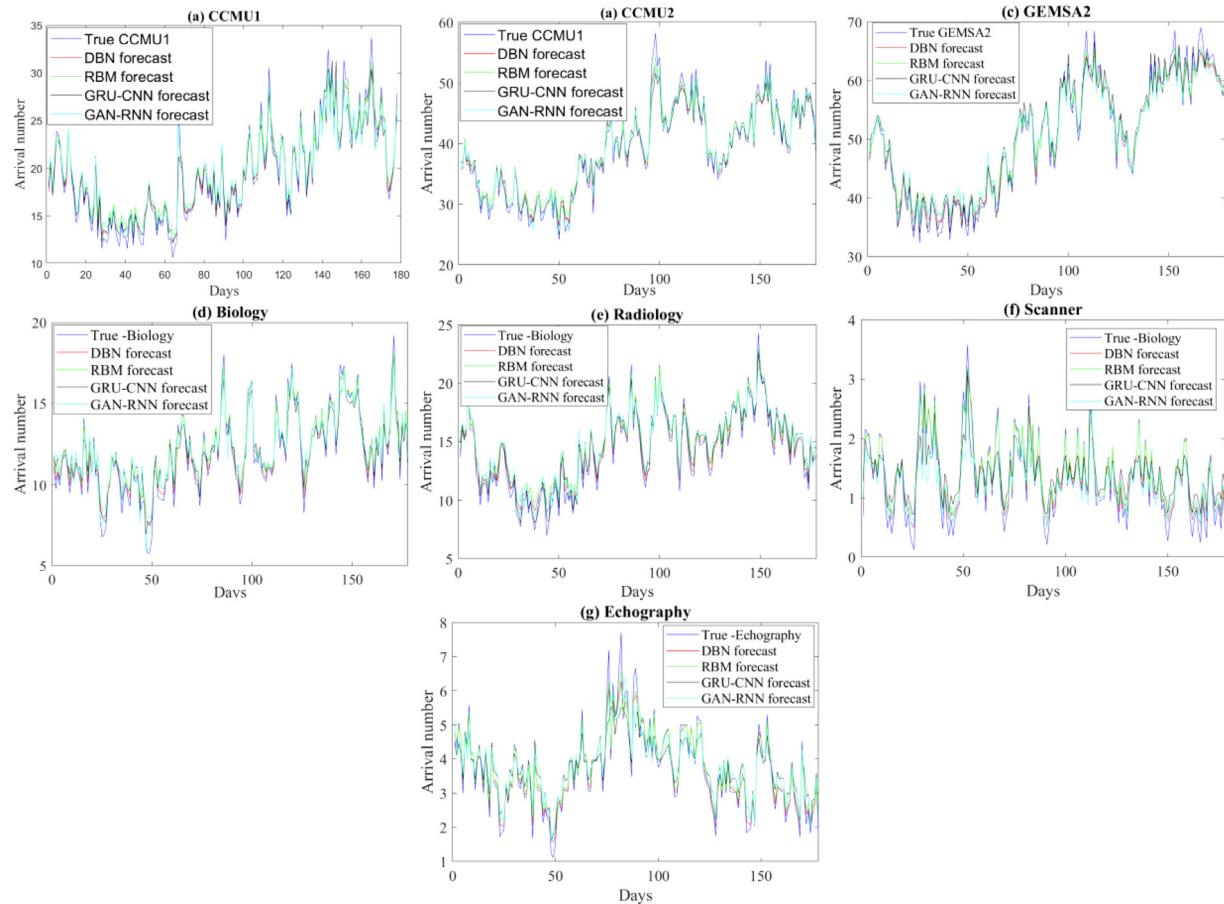


Fig. 11. Forecasting results of ED visits by the four models for each time series: (a) CCMU1, (b) CCMU2, (c) GEMSA2, (d) Biology, (e) Radiology, (f) Scanner, and (g) Echography.

train DBN, which allows deeply learning pertinent information in PED time-series data.

A boxplot of forecast errors is employed to visually verify the quality of the forecasting results of the investigated models. Fig. 13 illustrates the boxplots of the forecasting errors of each deep learning model based on testing daily PED data. The forecasting errors describe the difference separating the recorded and forecasted PED visits from the trained models. Crucially, in comparing boxplots, wider distribution is characterized by a large box with large ranges. In contrast, small and compact boxes with a central line (median) near zero reveal small forecasting errors. Visually, from Fig. 13 we can see also that DBN, RBM, GRU-CNN, and GAN-RNN with narrower boxes and whiskers reach superior performance compared to the other models (LSTM, GRU, LSTM-CNN, SVR, and RR). We also noticed that the DBN reaches accurate PED visits forecasting (i.e., having the shortest boxes with a median around zero) compared to all the other models.

Fig. 14 displays the NRMSE (%) derived with the above-considered models for test data. The results confirm that the DBN model efficiently

captures the linear and nonlinear features in patient flow data and appropriately forecasting patient flow time-series datasets.

Table 4 lists the averaged statistical scores of every model. The DBN achieved the best accuracy in terms of R^2 , EV, RMSE, MAE, and MAPE, in comparison to the other methods. It reached an R^2 of 0.973 and MAPE of 4.097%, which implies that DBN fitted well the PED visits time series data. It is followed by RBM, GRU-CNN, and GAN-RNN with R^2 of 0.954, 0.90, and 0.889. Table 5 tabulates the averaged evaluation scores of each series. R^2 and MAPE indicate that all series are well forecasted except the daily visits for the scanner that is predicted with an R^2 of 0.77. The daily visits for the scanner time series data are likely very dynamic, and therefore a large dataset is needed in training to capture the most variability in the data.

In this study, the DBN model demonstrated the best performance in forecasting PED visits time-series data. Those predicted PED visits time-series are illustrated in Fig. 15. Figs. 15 and 16 indicate that the forecasting result from the DBN is very satisfying.

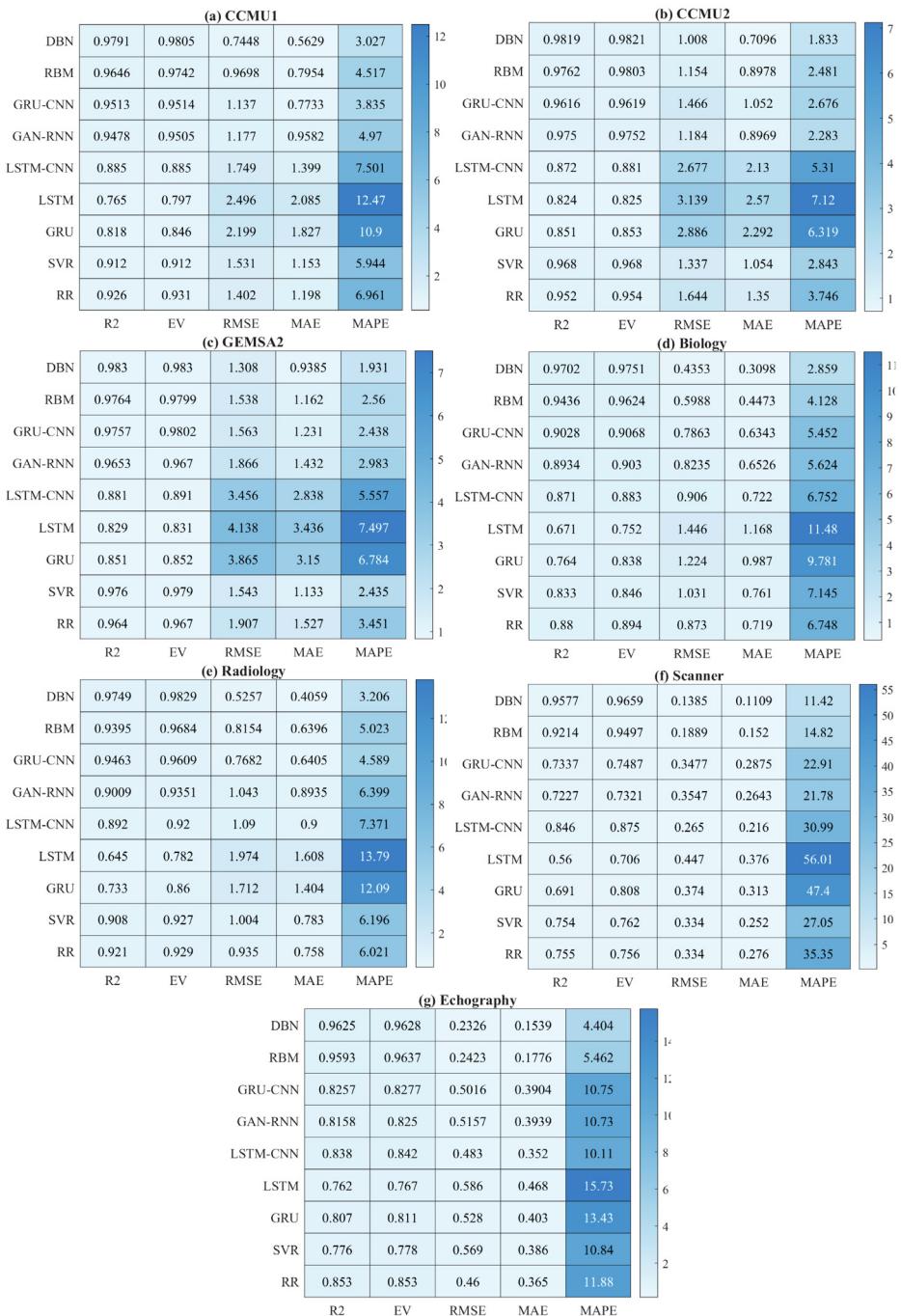


Fig. 12. Heatmap of the statistical indicators for each time series: (a) CCMU1, (b) CCMU2, (c) GEMSA2, (d) Biology, (e) Radiology, (f) Scanner, and (g) Echography.

Table 4
Averaged evaluation metrics per method.

Model	R2	EV	RMSE	MAE	MAPE
DBN	0.973	0.976	0.628	0.456	4.097
RBM	0.954	0.968	0.787	0.610	5.570
GRU-CNN	0.900	0.905	0.939	0.716	7.522
GAN-RNN	0.889	0.898	0.995	0.784	7.824
LSTM-CNN	0.869	0.882	1.518	1.222	10.512
LSTM	0.722	0.780	2.032	1.673	17.730
GRU	0.788	0.838	1.827	1.482	15.243
SVR	0.875	0.882	1.050	0.789	8.921
RR	0.893	0.898	1.079	0.885	10.594

Table 5
Averaged evaluation metrics per dataset.

Dataset	R2	EV	RMSE	MAE	MAPE
CCMU1	0.91	0.91	1.49	1.19	6.68
CCMU2	0.93	0.93	1.83	1.44	3.85
GEMSA2	0.93	0.94	2.35	1.87	3.96
Biology	0.86	0.88	0.90	0.71	6.66
Radiology	0.87	0.92	1.1	0.89	7.19
Scanner	0.77	0.81	0.31	0.25	29.75
Echography	0.84	0.85	0.46	0.34	10.37

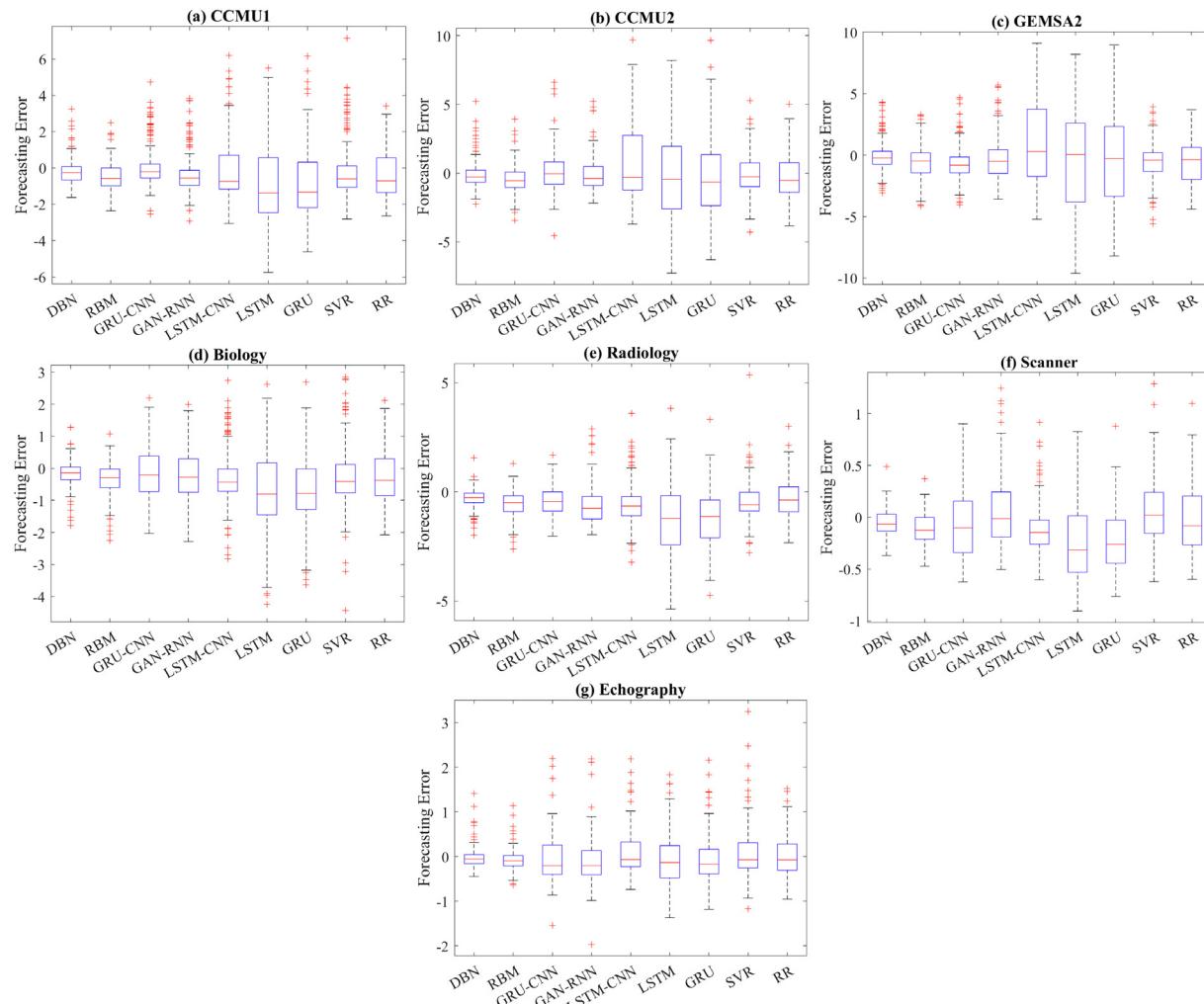


Fig. 13. Forecasting error for each time-series: (a) CCMU1, (b) CCMU2, (c) GEMSA2, (d) Biology, (e) Radiology, (f) Scanner, and (g) Echography.

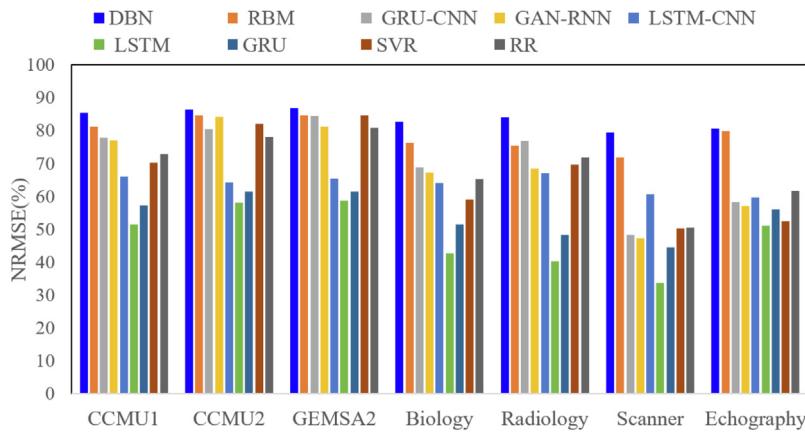


Fig. 14. NRMSE values from the examined models using test data.

In summary, we explored the feasibility of deep learning to forecast patient flows from different departments in PED, both with hybrid and non-hybrid deep methods. We construct three-hybrid models called the GRU-CNN, LSTM-CNN and the GAN-RNN, four non-hybrid models (DBN, RBM, LSTM, and GRU), and two shallow models (SVR and RR). Despite the sophistication of hybrid deep learning models, this study showed that the non-hybrid models (DBN and RBM) exhibit the highest accuracy forecasting of daily PED visits. This result is maybe because

the hybrid model requires large training data compared to the non-hybrid model to capture the most variability in data. The actual daily PED visits data is relatively small, which makes the non-hybrid models more suited. As expected, the DBN-based forecast method provides improved forecasting compared to the RBM-based forecast method. It is to be noted that the DBN model is built by stacking several RBMs, making it more efficient than the standalone RBM. Furthermore, the greedy layer-wise training procedure performed to build the DBN

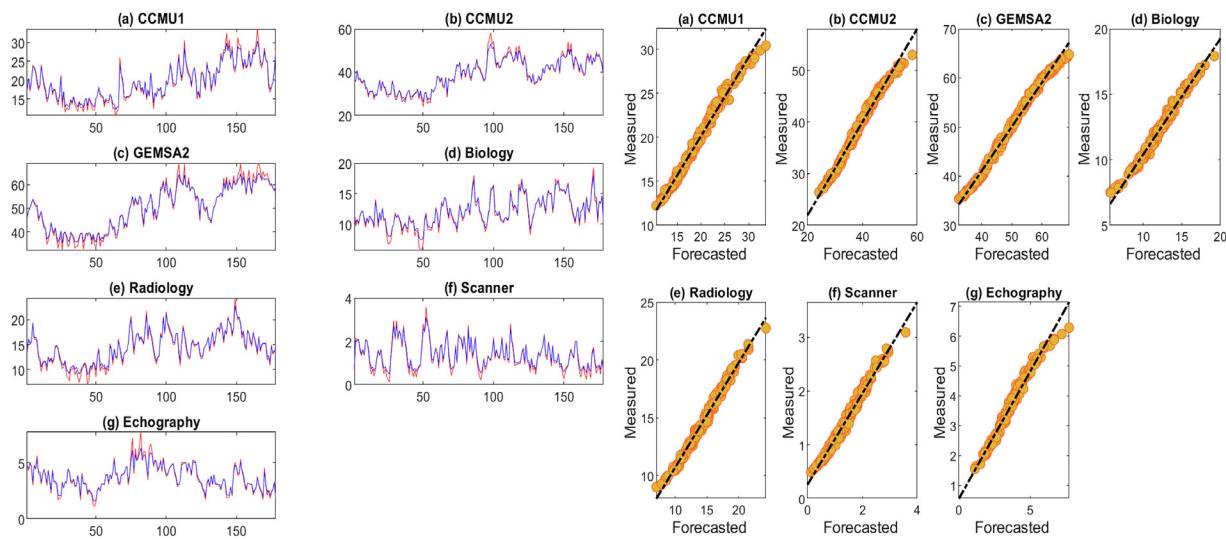


Fig. 15. Forecasted PED visits from the DBN model for each time-series:(a) CCMU1, (b) CCMU2, (c) GEMSA2, (d) Biology, (e) Radiology, (f) Scanner, and (g) Echography.

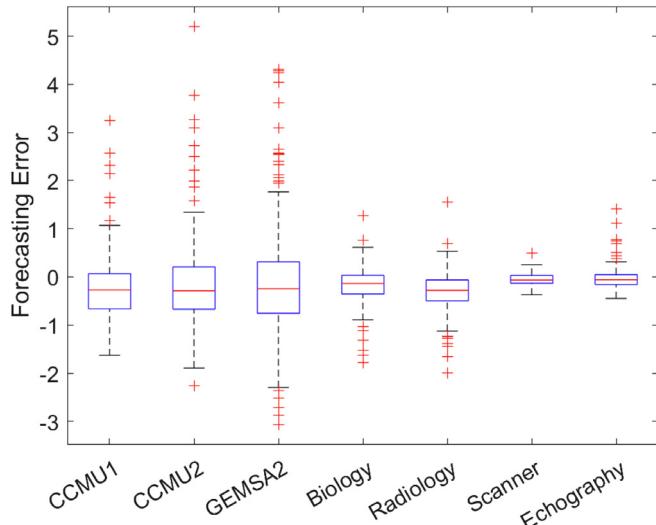


Fig. 16. Distribution of forecasting error of the investigated methods based on testing data.

model permits us to get an efficient model able to learn nonlinear variations and mimicking the patient flow dynamics. The DBN learning is also called greedy-layer wise training hierarchical learning, where each layer (represented by RBM) is trained separately. Also, the training results in DBN are based on a simple feed-forward neural network (no gates or memory cells used by RNNs).

4. Conclusion

The accurate forecast of daily visits in different departments at PED is crucial for optimizing appointment scheduling and facilitate managing staff deployment and allocation of resources. The rapid advancement in deep learning technology has generated high opportunities in data-based applications, especially healthcare. This paper investigates the feasibility of deep learning-driven methods for daily PED visits forecasting. Seven hybrid and non-hybrid deep learning models (DBN, RBM, GRU-CNN, LSTM-CNN, and GAN-RNN, LSTM, and GRU.), as well as two baseline models (SVR and RR), have been constructed and examined for patient flow forecasting. At first, the multivariate PED visits data from different departments (e.g., biology,

radiology, echography, and scanner) has been smoothed using the exponential smoothing approach to eliminate outliers and then normalized to train the four deep learning models. Then, the built methods are adopted for PED visits forecasting. This study is conducted using daily visits at the PED of Lille Hospital, France. Results show that the DBN model has a much better performance than the other models. This result is mainly due to the extended capacity of the DBN model for learning higher-level features via the greedy-layer wise training hierarchical learning procedure, which permits good forecasting accuracy.

Despite the satisfactory patient flow forecasting results using the deep learning models, there is still plenty of room for improvement. One direction of improvement is to incorporate external factors that affect the patient flow evolution, such as meteorological data and epidemic events, in building deep learning-based models. Another direction is to design an early detection system of overcrowding in EDs to foster reactive control, avoiding strain situations, and reduced consequences.

CRediT authorship contribution statement

Fouzi Harrou: Methodology, Conceptualization, Formal analysis, Data Curation, Writing, Validation, Reviewing. **Abdelkader Dairi:** Methodology, Formal analysis, Conceptualization, Data Curation, Writing, Validation, Reviewing. **Farid Kadri:** Data Curation, Formal analysis, Writing, Validation, Methodology, Conceptualization, Reviewing. **Ying Sun:** Supervision, Methodology, Conceptualization, Funding acquisition, Reviewing.

Declaration of competing interest

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

Acknowledgment

This publication is based upon work supported by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2019-CRG7-3800.

References

- Aboagye-Sarfo, P., Mai, Q., Sanfilippo, F. M., Preen, D. B., Stewart, L. M., & Fatovich, D. M. (2015). A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. *Journal of Biomedical Informatics*, 57, 62–73.
- Afifal, M., Yalaoui, F., Dugardin, F., Amodeo, L., Laplanche, D., & Blua, P. (2016). Forecasting the emergency department patients flow. *Journal of Medical Systems*, 40(7), 1–18.
- Barişçi, N. (2008). The adaptive ARMA analysis of EMG signals. *Journal of Medical Systems*, 32(1), 43–50.
- Benbelkacem, S., Kadri, F., Atmani, B., & Chaabane, S. (2019). Machine learning for emergency department management. *International Journal of Information Systems in the Service Sector (IJSSS)*, 11(3), 19–36.
- Bengio, Y. (2009). *Learning deep architectures for AI*. Now Publishers Inc.
- Bergs, J., Heerinckx, P., & Verelst, S. (2014). Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. *International Emergency Nursing*, 22(2), 112–115.
- Boujemaa, R., Jebali, A., Hammami, S., Ruiz, A., & Bouchriha, H. (2018). A stochastic approach for designing two-tiered emergency medical service systems. *Flexible Services and Manufacturing Journal*, 30(1–2), 123–152.
- Carvalho-Silva, M., Monteiro, M. T. T., de Sá-Soares, F., & Dória-Nóbrega, S. (2018). Assessment of forecasting models for patients arrival at emergency department. *Operations Research for Health Care*, 18, 112–118.
- Champion, R., Kinsman, L., Lee, G., Masman, K., May, E., Mills, T., et al. (2007). Forecasting emergency department presentations. *Australian Health Review*, 31(1), 83–90.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
- Daghistani, T. A., Elshawi, R., Sakr, S., Ahmed, A. M., Al-Thwayee, A., & Al-Mallah, M. H. (2019). Predictors of in-hospital length of stay among cardiac patients: A machine learning approach. *International Journal of Cardiology*, 288, 140–147.
- Dairi, A., Cheng, T., Harrou, F., Sun, Y., & Leiknes, T. (2019). Deep learning approach for sustainable WWTP operation: A case study on data-driven influent conditions monitoring. *Sustainable Cities and Society*, 50, Article 101670.
- Dairi, A., Harrou, F., Khadraoui, S., & Sun, Y. (2021). Integrated multiple directed attention-based deep learning for improved air pollution forecasting. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–15.
- Dairi, A., Harrou, F., Sun, Y., & Senouci, M. (2018). Obstacle detection for intelligent transportation systems using deep stacked autoencoder and k-nearest neighbor scheme. *IEEE Sensors Journal*, 18(12), 5122–5132.
- Erfani, S. M., Rajasegarar, S., Karunasekera, S., & Leckie, C. (2016). High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, 58, 121–134.
- Ganguly, A., & Nandi, S. (2016). Using statistical forecasting to optimize staff scheduling in healthcare organizations. *Journal of Health Management*, 18(1), 172–181.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).
- Gul, M., & Guneri, A. F. (2016). Planning the future of emergency departments: Forecasting ED patient arrivals by using regression and neural network models. *International Journal of Industrial Engineering*, 23(2).
- Guo, Z., Zhou, K., Zhang, X., & Yang, S. (2018). A deep learning model for short-term power load and probability density forecasting. *Energy*, 160, 1186–1200.
- Handly, N., Thompson, D. A., Li, J., Chuirazzi, D. M., & Venkat, A. (2015). Evaluation of a hospital admission prediction model adding coded chief complaint data using neural network methodology. *European Journal of Emergency Medicine*, 22(2), 87–91.
- Harrou, F., Dairi, A., Kadri, F., & Sun, Y. (2020). Forecasting emergency department overcrowding: A deep learning framework. *Chaos, Solitons & Fractals*.
- Harrou, F., Hittawe, M. M., Sun, Y., & Beya, O. (2020). Malicious attacks detection in crowded areas using deep learning-based approach. *IEEE Instrumentation & Measurement Magazine*, 23(5), 57–62.
- Harrou, F., Kadri, F., Chaabane, S., Tahon, C., & Sun, Y. (2015). Improved principal component analysis for anomaly detection: Application to an emergency department. *Computers & Industrial Engineering*, 88, 63–77.
- Harrou, F., Kadri, F., Sun, Y., & Khadraoui, S. (2021). Monitoring patient flow in a hospital emergency department: ARMA-based nonparametric GLRT scheme. *Health Informatics Journal*, 27(2), Article 14604582211021649.
- Harrou, F., Sun, Y., Hering, A. S., Madakyaru, M., et al. (2020). *Statistical process monitoring using advanced data-driven and deep learning approaches: Theory and practical applications*. Elsevier.
- Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527–1554.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Ichikawa, D., Saito, T., Ujita, W., & Oyama, H. (2016). How can machine-learning methods assist in virtual screening for hyperuricemia? A healthcare machine-learning approach. *Journal of Biomedical Informatics*, 64, 20–24.
- Jiang, S., Xiao, R., Wang, L., Luo, X., Huang, C., Wang, J.-H., et al. (2019). Combining deep neural networks and classical time series regression models for forecasting patient flows in Hong Kong. *IEEE Access*, 7, 118965–118974.
- Jones, S. S., Evans, R. S., Allen, T. L., Thomas, A., Haug, P. J., Welch, S. J., et al. (2009). A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of Biomedical Informatics*, 42(1), 123–139.
- Kadri, F., Baraoui, M., & Nouaouri, I. (2019). An LSTM-based deep learning approach with application to predicting hospital emergency department admissions. In *2019 international conference on industrial engineering and systems management (IESM)* (pp. 1–6). IEEE.
- Kadri, F., Chaabane, S., & Tahon, C. (2014). A simulation-based decision support system to prevent and predict strain situations in emergency department systems. *Simulation Modelling Practice and Theory*, 42, 32–52.
- Kadri, F., Harrou, F., Chaabane, S., & Tahon, C. (2014). Time series modelling and forecasting of emergency department overcrowding. *Journal of Medical Systems*, 38(9), 107.
- Kadri, F., Harrou, F., & Sun, Y. (2017). A multivariate time series approach to forecasting daily attendances at hospital emergency department. In *2017 IEEE symposium series on computational intelligence (SSCI)* (pp. 1–6). IEEE.
- Kadri, F., Pach, C., Chaabane, S., Berger, T., Trentesaux, D., Tahon, C., et al. (2013). Modelling and management of strain situations in hospital systems using an ORCA approach. In *Industrial engineering and systems management (IESM), Proceedings of 2013 international conference on* (pp. 1–9). IEEE.
- Karsanti, H. T., Ardianto, I., & Nugroho, L. E. (2019). Deep learning-based patient visits forecasting using long short term memory. In *2019 international conference of artificial intelligence and information technology (ICAIT)* (pp. 344–349). IEEE.
- Kellermann, A. L. (2006). Crisis in the emergency department. *New England Journal of Medicine*, 355(13), 1300–1303.
- Khatri, K. L., & Tamil, L. S. (2017). Early detection of peak demand days of chronic respiratory diseases emergency department visits using artificial neural networks. *IEEE Journal of Biomedical and Health Informatics*, 22(1), 285–290.
- Kolker, A. (2008). Process modeling of emergency department patient flow: Effect of patient length of stay on ED diversion. *Journal of Medical Systems*, 32(5), 389–401.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Lin, W. (1989). Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches. *International Journal of Forecasting*, 5(2), 195–208.
- Liu, W., Song, Y., Chen, D., He, S., Yu, Y., Yan, T., et al. (2019). Deformable object tracking with gated fusion. *IEEE Transactions on Image Processing*, 28(8), 3766–3777.
- Ma, F., Yu, L., Ye, L., Yao, D. D., & Zhuang, W. (2020). Length-of-stay prediction for pediatric patients with respiratory diseases using decision tree methods. *IEEE Journal of Biomedical and Health Informatics*.
- Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743–758.
- Maldonado, R., & Harabagiu, S. M. (2019). Active deep learning for the identification of concepts and relations in electroencephalography reports. *Journal of Biomedical Informatics*, 98, Article 103265.
- Martín, R., & Cáceres, H. (2005). A method for ascertaining the seasonal pattern of hospital emergency department visits. *Revista Española de Salud Pública*, 79(1), 5–15.
- Mohamed, B., & Mohamad, M. (2020). Forecasting patient admission in orthopedic clinic at a hospital in Kuantan using autoregressive integrated moving average (ARIMA) models. In *Journal of physics: Conference series*. IOP Publishing, Article 052090.
- Pham, T., Tran, T., Phung, D., & Venkatesh, S. (2017). Predicting healthcare trajectories from medical records: A deep learning approach. *Journal of Biomedical Informatics*, 69, 218–229.
- rekha Hanumanthu, S. (2020). Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. *Chaos, Solitons & Fractals*, Article 109947.
- Saba, L., Biswas, M., Kuppili, V., Godia, E. C., Suri, H. S., Edla, D. R., et al. (2019). The present and future of deep learning in radiology. *European Journal of Radiology*, 114, 14–24.
- Salakhutdinov, R., & Hinton, G. (2009). Deep boltzmann machines. In *Artificial intelligence and statistics* (pp. 448–455).
- Smolensky, P. (1986). *Information processing in dynamical systems: Foundations of harmony theory: CU-CS-321-86*.
- Sun, Y., Heng, B., Seow, Y., & Seow, E. (2009). Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emergency Medicine*, 9(1), 1.
- Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of parkinson's disease using deep learning and machine learning. *IEEE Access*, 8, 147635–147646.

- Warden, G., Griffin, R., Erickson, S., Mchugh, M., Wheatley, B., Dharshi, A., et al. (2006). *Hospital-based emergency care: at the breaking point*. Washington, DC: National Academy of Science.
- Xie, Y., Schreier, G., Hoy, M., Liu, Y., Neubauer, S., Chang, D. C., et al. (2016). Analyzing health insurance claims on different timescales to predict days in hospital. *Journal of Biomedical Informatics*, 60, 187–196.
- Xu, M., Wong, T.-C., & Chin, K.-S. (2013). Modeling daily patient arrivals at emergency department and quantifying the relative importance of contributing variables using artificial neural network. *Decision Support Systems*, 54(3), 1488–1498.
- Yu, L., Hang, G., Tang, L., Zhao, Y., & Lai, K. (2017). Forecasting patient visits to hospitals using a WD&ANN-based decomposition and ensemble model. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(12), 7615–7627.
- Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-series data: A comparative study. *Chaos, Solitons & Fractals*, Article 110121.
- Zhang, Y., Luo, L., Zhang, F., Kong, R., Yang, J., Feng, Y., et al. (2020). Emergency patient flow forecasting in the radiology department. *Health Informatics Journal*, Article 1460458220901889.