An LSTM-based Deep Learning Approach with Application to Predicting Hospital Emergency Department Admissions

Farid Kadri
Analytics and Big Data dept.
Aeroline Agence PCS, Sopra Steria
Colomiers, France
farid.kadri@soprasteria.com

Merouane Baraoui
Analytics and Big Data dept.
Aeroline Agence PCS, Sopra Steria
Colomiers, France
merouane.baraoui@soprasteria.com

Issam Nouaouri
LGI2A Laboratory
Univ. Artois, EA 3926 Arras,
F-62400, France
issam.nouaouri@univ-artois.fr

***Abstract*—Since the need for medical cares has significantly increased all over the last years, the efficient management of patient flow becomes a core element for hospitals and particularly in emergency departments (EDs). With the high demand for ED services, overcrowding can be generated and thus the quality of medical services could be degraded. In this regards, forecasting daily attendances at the ED is vital to mitigate the overcrowding problems. Specifically, ED demands forecasting provides relevant information for ED's managers to appropriately use the available resources. This paper presents a Long Short-Term Memory (LSTM)-based deep learning approach for forecasting daily admissions at an ED. Experimental data from the pediatric emergency department at Lille regional hospital center, France, are used to test the efficiency of the proposed approach. Results show the good potential of the LSTM-based deep learning approach in forecasting ED admissions.**

***Keywords*— Emergency departments, Patient flows, ED demands, Prediction, Deep learning, LSTM.**

I. Introduction

Since the need for medical care's has significantly increased all over the last years, it makes patient flows and medical staff management more and more important. The control of ED demands is a challenge faced by many hospital systems. The EDs represent the main mission of the healthcare establishments. Also, EDs considered as the main gateway to the hospital because they constitute an almost mandatory passage for all patients before their admission to most hospital services [1]–[3]. Hence, EDs must be able to receive patient flows sometimes very important for medical and surgical treatments. To meet their missions; EDs must incorporate in their operating mode the ability to anticipate and to forecast the occurrence of these perturbations.

Forecasting the patient arrivals at an ED may provide useful information's for ED managers. This information's are important for best allocating human and material resources, and then reducing the patient waiting time and length of stay.

The ability to forecast ED attendances is more important because it allows ED managers to design a best strategies aimed at avoiding overcrowding in their establishments [4], [5]. The classical ED demands forecasting approaches include regression models and times series models (Exponential smoothing, ARMA and its variants), are limited in term of capturing more random variations, nonlinear characteristics, accuracy and prediction horizon (long term forecasting) [6]–[10]. Recurrent Neural Networks (RNNs) have been successfully used in machine learning and deep learning problems. These models have been proposed to address time-dependent learning problems. Long Short Term Memory (LSTM) is one of the most used RNN models for time series data predictions, which is perfectly suited to the ED demands forecasting problems [11]. Several case studies show that LSTM network presented better performance than classical time series models and traditional networks [12], [13]. In this paper, we applied the Long Short term memory (LSTM) recurrent neural network (RNN) model to predict daily admissions at the pediatric emergency department (PED) in Lille regional hospital center, France. The remainder of this paper is organized as follows. Section II introduces the basics of deep learning, its applications in the case of ED demands, and presents an LSTM-based approach to forecast the patient flows at EDs. Section III presents and discusses the results, and section IV concludes this study.

II. Deep learning and predicting ED admissions

Emergency departments (EDs) suffer from several constraints (structural, human, material, financial and organizational) [14], [15]. Due to these constraints, high complexity of such organizations and the increasing demands for cares, the importance of forecasting of ED admissions is absolutely critical for resource (human and material) planning in the hospital [16]–[19].

Over the last decades, many studies have been dedicated to the forecasting problems in several application domains. In the case of hospital systems, many forecast models have been proposed in order to study and develop appropriate predictions for ED demands. According to the literature, most of the developed techniques/models can be categorized into two categories: i) statistical and time series models such as regression and time series models: ARIMA model Holt–Winters exponential smoothing, and ii) artificial intelligence techniques such as neural networks, thanks to the development of computer science, Deep Learning approaches have become one of the most popular research in academics, industries, and services sectors.

The Deep Learning is the result of addition of more layers into the neural network mechanisms. The most used deep learning models include Boltzmann machines, Deep Belief Networks (DBN) and Recurrent Neural Networks (RNNs). RNN is a type of neural networks that exploit the sequential nature of input data. RNNs are used to model time-dependent

78-1-7281-1566-5/19/$31.00 ©2019 European Union

data, and they give good results in the time series data, which have proven successful in several applications domains [12], [13]. Long Short-Term Memory Networks (LSTM) is a type of RNNs that is able to deal with remembering information's for much longer periods of time [11]. It is also considered as one of the most used RNN models for time series data predictions, which is perfectly suited to the ED demands forecasting problems.

The two next sub-sections introduce the basic model of RNN-LSTM and how it can be designed and implemented, then present an approach for modeling and predicting the daily patient arrivals at an ED.

A. Long Short-Term Memory models

LSTM models were initially proposed by Hochreiter and Schmidhuber [11] and were improved and popularized by many other researchers [14]–[17]. LSTM models have an excellent ability to memorize long-term dependencies [10], they are developed to improve the problem of vanishing gradient when training traditional RNNs. A common RNN-LSTM model is composed of a cell blocks in place of a standard neural network layers as shown in the figure 1.

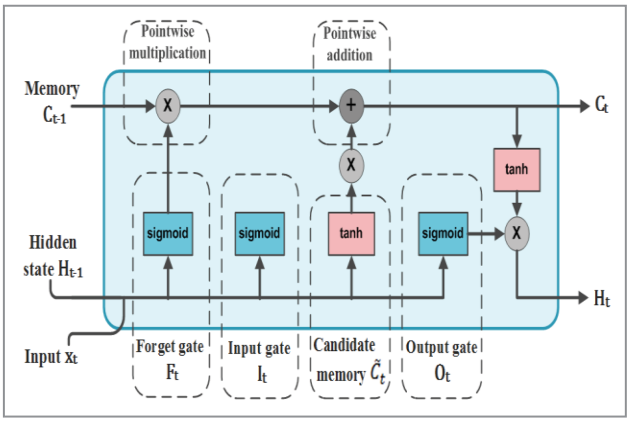

Fig.1. Basic structure of the Long Short-Term Memory (LSTM)

According the figure 1, the RNN-LSTM has two input features at each time, which include the current time step input X_t (input vector) and the hidden state of the previous time step H_{t-1} (previous input vector). The output is computed by the fully connected layer with its activation function (tanh, sigmoid, Softmax, Adam…).

Let us denote input time series as X_t, the number of hidden units as h, the hidden state of the last time step as H_{t-1}, and the output time series as H_t. The mathematical relationship between inputs and outputs of the RNN-LSTM can be described as follows:

$$\begin{aligned} I_t &= \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i) \\ F_t &= \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f) \\ O_t &= \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o) \end{aligned} \tag{1}$$

$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c) \tag{2}$$

$$C_t = F_t \; o \; C_{t-1} \; o \; \tilde{C}_t \tag{3}$$

$$H_t = O_t \; o \tanh(C_t) \tag{4}$$

where:

- I_t, F_t, O_t are respectively the input gate, forget gate, and output gate, W_{xi}, W_{xf}, W_{xo} and W_{hi}, W_{hf}, W_{ho} are weight parameters and b_i, b_f, b_o are bias parameters. All these gates have the same dimensions and equations with different parameter. They are called gates because the activation function transforms the element values between ranges ([0, 1], [-1, 1]).

- $\tilde{C}_t$ is the candidate memory cells, W_{xc}, W_{hc} are weight parameters and b_c is a bias parameter. LSTM model needs to compute the $\tilde{C}_t$, its computation is similar to the input, forget and output gates, but using a *tanh* function as an activation function with a value range between [-1, 1].

- C_t is the memory cells, o is an operator which expresses element-wise multiplication. The computation of the current C_t combines the information of the previous C_{t-1} and the current candidate $\tilde{C}_t$.

- H_t is the hidden states, the information flow of the hidden state H_t is controlled through the output gate. The *tanh* function ensures that the hidden state element value is between [-1, 1].

B. Proposed approach

The proposed approach in this paper aims to predict the daily patient arrivals at an emergency department. This methodology is based on LSTM deep-learning model. The figure 2 summarized the main steps of the proposed methodology. According to the figure 2, the proposed approach includes four key steps: i) data collecting from an ED database, ii) data pre-processing and statistical analysis, iii) data standardization, used to normalize the original data, iv) train, validate and test the LSTM model. Finally the validated LSTM model will used for the prediction.

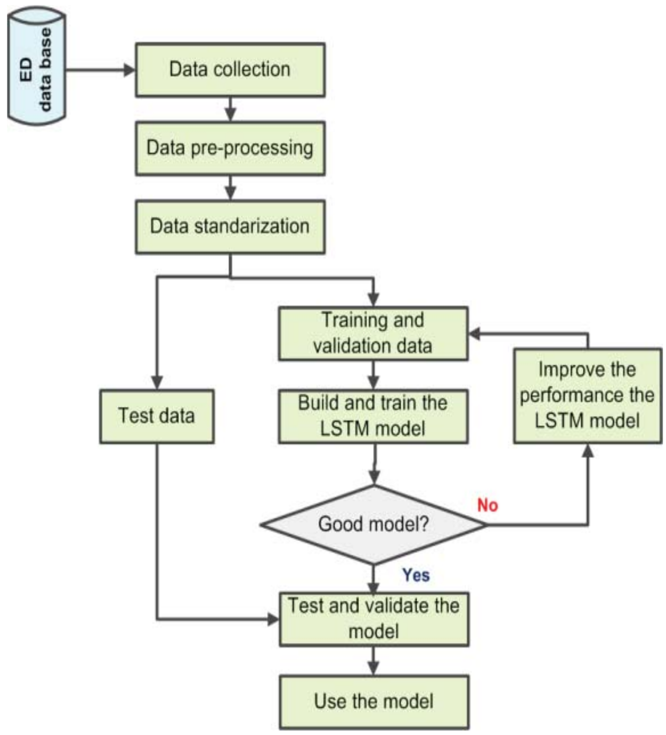

Fig.2. Flow chart of the proposed approach.

1) Metrics for evaluating the forecasting models

Several metrics are proposed in the literature in order to evaluate the predictive abilities of the forecasting models. The popular metrics include; mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), Coefficient of determination (R^2), and mean absolute percentage error (MAPE). In this paper, MAE, RMSE, and R^2 are used to evaluate the accuracy of ED admissions forecasting, and the MSE is used as a loss function of LSTM model. The MAE, RMSE, and R^2 are formulated as follows:

$$MAE = \frac{1}{n}\sum|\hat{x} - x| \quad (5)$$

$$RMSE = \sqrt{\frac{\sum(\hat{x}-x)^2}{n}} \quad (6)$$

where x are the measured values, $\hat{x}$ are the corresponding predicted values and n is the number of samples. MAE and RMSE are two of the most common metrics used to measure accuracy for continuous variables. These statistics are easier to interpret.

The coefficient of determination (R^2) corresponds to the percentage of variability explained by the model. R^2 is used to evaluate the goodness of fit (real and predicted values) and defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i-\hat{x}_i)^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \quad (6)$$

where, x_i are the real values, $\hat{x}_i$ are the predicted values obtained from the selected model, and $\bar{x}$ is the mean of the observed data.

2) Improve the performance of LSTM model

The accuracy of deep learning models depends not only on the quality and the quantity of training datasets but also on their architectures (configuration of the network), their hyper-parameters, and the used optimizers. To improve the RNN models performance, we can act on:

- *Activation Functions:* activation functions are used to determine the output of neural network. It can be divided into two types: 1) linear activation functions: the output of these functions is linear and it's not confined between any range, and ii) non-linear activation functions: are the most used activation functions, the output is non-linear and they confined between range. For example: rescale data to values between [0,1] for sigmoid, Softmax, and Rectified Linear Unit (ReLU) activation functions, and between [-1,1] for Hyperbolic Tangent (tanh) activation functions.

- *Optimization and Loss:* optimization algorithms are used in order to minimize the objective function of the neural network. A common used algorithms are: stochastic gradient descent (SGD), RMSProp, ADAM, and NADAM [24], [25].

- *Dropout:* is a popular stochastic regularization technique for reducing over-fitting and improving the RNN models performance [26], [27].

- *Epochs and Batches:* In case of deep learning implementations we need to act on the number of epochs and batch. The literature shows that the use of large epochs and small batch sizes give good results.

- *Weight regularization:* is a technique for imposing constraints on the weights within RNN nodes. It reduces over-fitting and improves the model performance. There are three different regularization techniques: 1) L_1: calculated as the sum of absolute values, 2) L_2 calculated as the sum of the squared values, and 3) both L_1L_2 calculated as the sum of absolute and sum of the squared values.

3) Implementation

The general program of RNN LSTM model may be designed with four steps: define the LSTM model, compile the LSTM network, fit the LSTM model, and use the validated LSTM model to forecast. The partial codes used for building the RNN-LSTM model are show in the table I.

TABLE I. PARTIAL CODES USED FOR BUILDING THE LSTM NETWORK

Steps	Action
Step 1: Define LSTM network	*from keras.layers.recurrent import LSTM from keras.models import Sequential from keras.layers.core import Activation, Dense, Dropout model = Sequential() model.add(LSTM(units=nb_neural, return_sequences=True, input_shape=(Xtrain.shape[1], 1)))*
Step 2: Compile the LSTM network	*model.compile(loss="mse", optimizer="adam", metrics=[rmse, 'mae', Rsquare])*
Step 3: Fit the LSTM network	*history = model.fit(Xtrain, ytrain, batch_size=batch_size, epochs=num_epochs, validation_data=(Xval, yval), verbose=2)*
Step 4: Forecasting	*ypred = model.predict(Xtest)*

III. CASE STUDY

Lille Regional Hospital Centre (CHRU-Lille) serves four million inhabitants in Nord-Pas-de-Calais, France. The pediatric emergency department (PED) in CHRU receives around 24 000 patients by year. Besides its internal capacity, the PED shares many resources with adult emergency department.

The data used in this paper was extracted from the database of the PED of Lille. It represents two years (2011 and 2012). Figure 3 presents the monthly arrivals at the PED. According to this figure, the patient flow varies between winter/epidemic (November to March) period and normal period (April to October). Figure 4 presents the patient daily arrivals at the PED. Figure 5 presents the patient arrivals by day of week. According to this figure, the number of patients varies considerably according to the day of the week. We also observed that Sunday, Monday and Saturday are the most overloaded.

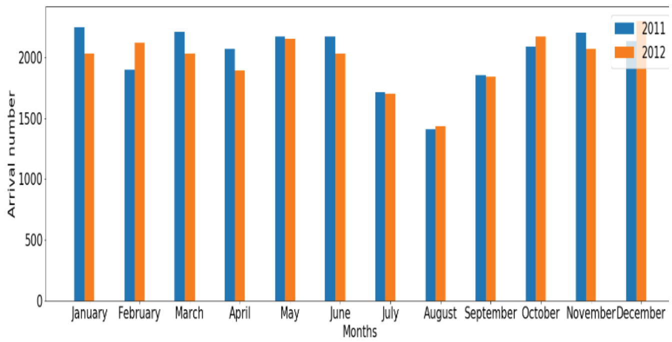

Fig.3. Patient by month (2011 to 2012).

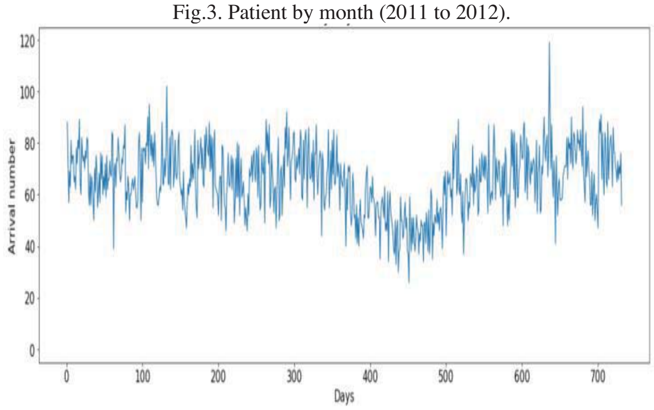

Fig.4. Daily PED arrivals (2011 to 2012).

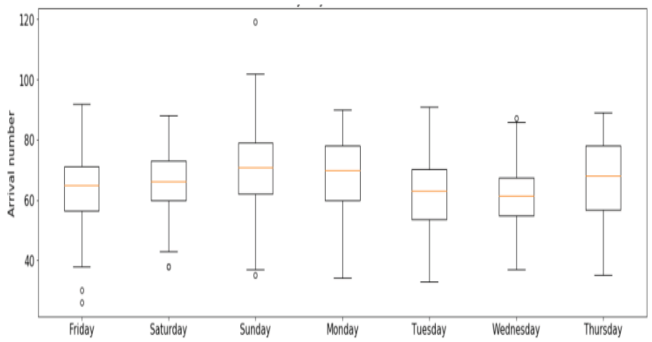

Fig.5. Box plots of the patient arrivals by day of the week for the entire studied period (January 2011 to December 2012).

A. LSTM Model selection

The LSTM model was building based on the parameters/metrics presented above (see section 2.A and 2.B). Data were split into training, validation and test datasets (80% , 10%, and 10% respectively). Table II summarizes the algorithms and parameters used in the modeling and fitting of LSTM model. In this study the LSTM model selected were evaluated by using MAE, RMSE and R^2 (see table III).

TABLE II. HYPERPARAMETERS AND LEARNING ALGORITHMS

Hyperparameters and Learning algorithms	Value/name
Droupout	0.2
Batch size	32
Neural number	50
Optimisation method	ADAM
Loss function	MSE
Output activation	ReLu
Metrics	MAE, RMSE, R^2
Epochs	50
Sequence step	30
Number of LSTM layers	1
Regularization, L1 and L2	1.10^{-2}
Learning rate	2.10^{-4}

TABLE III. VALIDATION MEASURES APPLIED TO STUDIED DATA

Parameters	Value
R^2	0.972
$RMSE$	0.089
MAE	0.068
Loss Function	0.016

Figure 6 presents the evolution of the loss function, RMSE, MAE and R2 in function of the number of epochs. As shown in Table 3, the high R^2 (0.982), the low RMSE (0.089) and the low MAE (0.068) of the best-fit LSTM model show that the selected LSTM model closely represented the observed time series.

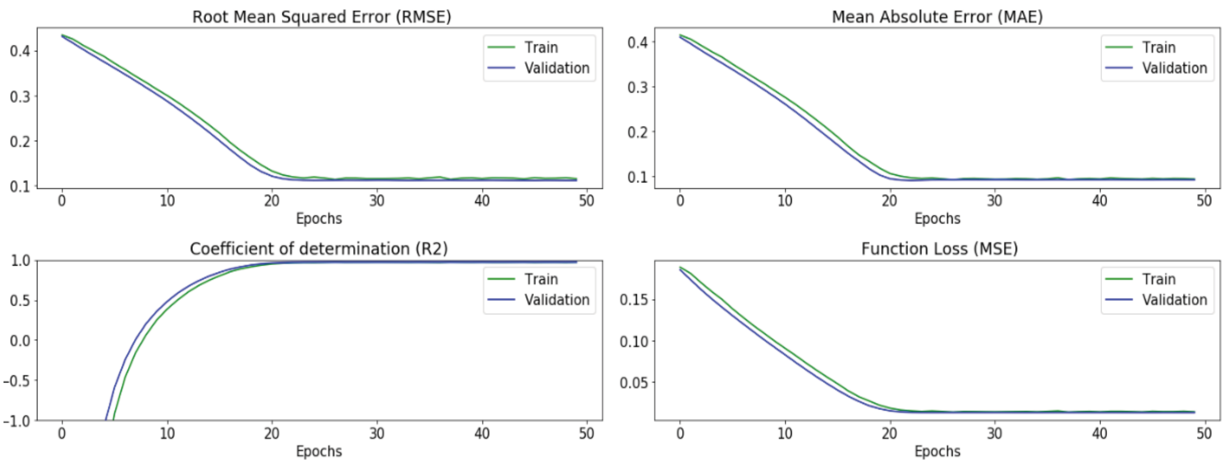

Fig.6. Evolution of the loss function, RMSE, MAE and R^2 in function of the number of Epochs

B. *Forecasting*

Once the RNN-LSTM model has been fitted to the real data, future values of the patient arrival at the PED can forecast. The fitted LSTM model from the above sub-section was examined for its predictive capability. The output of the model was compared with the actual data values. The results of the predicted and actual values (the predicted and actual number of patient arrival at the PED) are shown in Figure 7, recorded for horizon H = 78 days (more than two months).

Table IV presents the statistical descriptive of the forecasting errors (difference between the real and predict data). We remarked that the mean forecast error for daily patient arrivals is 5.133 (5 patients) with a standard deviation of the forecast errors of 4.08. According to this study, the fitted LSTM model has a good ability of prediction and can be used to forecast the patient arrival numbers at the PED.

TABLE IV. STATISTICAL DESCRIPTIVE OF THE FORECASTING ERRORS

Metrics	Value
Minimum	0.021
Mean	5.133
Maximum	28.79
Standard deviation	4.081

IV. CONCLUSION

An LSTM-based deep-learning approach has been proposed for reliable forecasting of daily patient arrivals at the pediatric emergency department (PED). This approach employs the desirable properties of LSTM, which is a powerful tool for modeling dependency in data. The forecasting quality of this approach has been verified using data from January 2011 to December 2012 collected from PED in Lille regional hospital center, France. Promising results have been achieved by the proposed LSTM-based approach to forecasting daily demands at the PED.

As future work, to further enhance the forecasting quality we plan implement and test the performance of other RNN models like Gated recurrent unit (GRU) model, and to incorporate other information such as meteorological data, pollution peaks, and epidemic events.

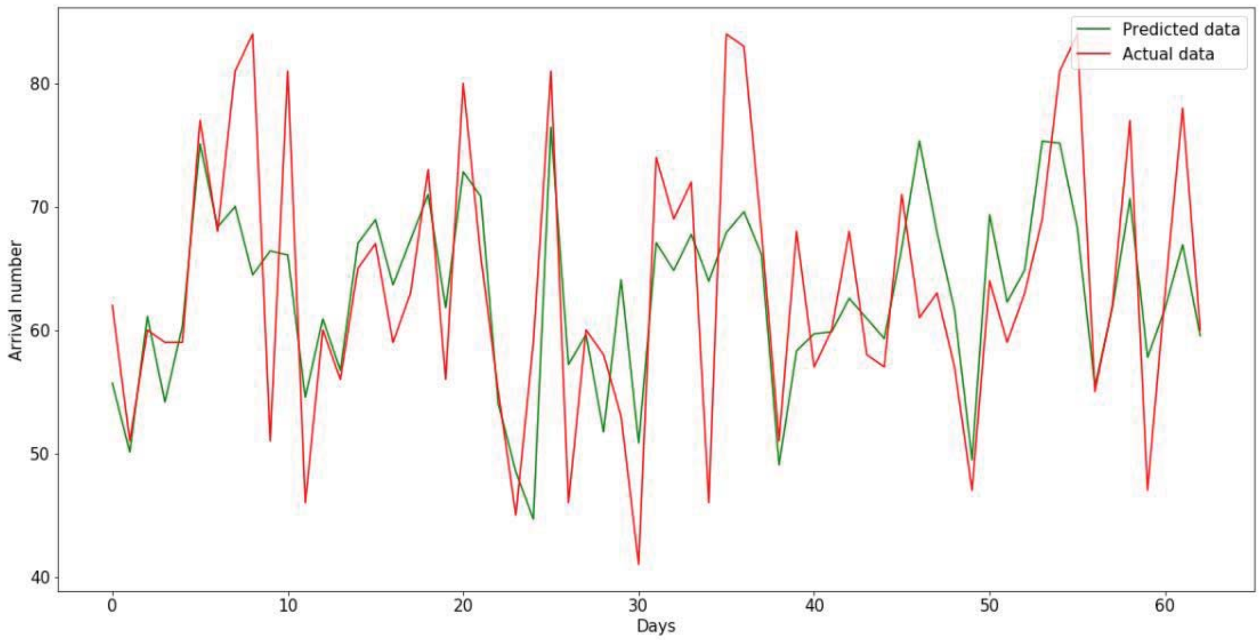

Fig.7. Actual daily patient arrivals and its forecast with the fitted LSTM Model

REFERENCES

[1] J. A. Gordon, J. Billings, B. R. Asplin, and K. V. Rhodes, “Safety net research in emergency medicine: proceedings of the Academic Emergency Medicine Consensus Conference on ‘The Unraveling Safety Net," Acad Emerg Med, vol. 8, no. 11, pp. 1024–1029, 2001.

[2] A. Boyle, K. Beniuk, I. Higginson, and P. Atkinson, “Emergency Department Crowding: Time for Interventions and Policy Evaluations,” Emergency Medicine International, vol. 2012, pp. 1–8, 2012.

[3] F. Kadri, F. Harrou, S. Chaabane, and C. Tahon, “Time series modelling and forecasting of emergency department overcrowding,” Journal of medical systems, vol. 38, no. 107, pp. 1–20, 2014.

[4] F. Kadri, S. Chaabane, and C. Tahon, “A simulation-based decision support system to prevent and predict strain situations in emergency department systems,” Simulation Modelling Practice and Theory, vol. 42, pp. 32–52, Mar. 2014.

[5] F. Kadri et al., “Modelling and management of strain situations in hospital systems using an ORCA approach,” in Proceedings of 2013 International Conference on Industrial Engineering and Systems Management (IESM), 2013, pp. 1–9.

[6] F. Harrou, F. Kadri, S. Chaabane, C. Tahon, and Y. Sun, “Improved Principal Component Analysis for Anomaly Detection: Application to an Emergency Department,” Computers & Industrial Engineering, vol. 88, pp. 63–77, Jun. 2015.

[7] S. S. Jones, A. Thomas, R. S. Evans, S. J. Welch, P. J. Haug, and G. L. Snow, “Forecasting daily patient volumes in the emergency department,” Acad Emerg Med, vol. 15, no. 2, pp. 159–170, 2008.

[8] A. Earnest, M. I. Chen, D. Ng, and L. Y. Sin, “Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore,” BMC Health Services Research, vol. 5, no. 1, p. 36, May 2005.

[9] R. Champion et al., “Forecasting emergency department presentations,” Aust Health Rev, vol. 31, no. 1, pp. 83–90, 2007.

[10] T. Kim, K. Hong, S. Shin, G. Park, S. KIM, and N. Hong, “Forecasting respiratory infectious outbreaks using ED-based syndromic surveillance for febrile ED visits in a Metropolitan City,” Am J Emerg Med, 2018.

[11] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, no. 9, pp. 1735–1780, 1997.

[12] U. Ugurlu, I. Oksuz, and O. Tas, “Electricity Price Forecasting Using Recurrent Neural Networks,” Energies, Apr. 2018.

[13] R. Fu, Z. Zhang, and L. Li, “Using LSTM and GRU neural network methods for traffic flow prediction,” in 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), 2016, pp. 324–328.

[14] F. Kadri, S. Chaabane, A. Bekrar, and C. Tahon, “Resilience-Based Performance Assessment of Strain Situations in Emergency Departments,” presented at the Industrial Engineering and Systems Management (IESM), 2015 International Conference on, Seville, 2015, pp. 609–618.

[15] F. Kadri, S. Chaabane, and C. Tahon, “Reactive Control System to Manage Strain Situations in Emergency Departments,” presented at the 13th International Conference on Informatics in Control, Automation and Robotics, 2016.

[16] S. Benbelkacem, F. Kadri, S. Chaabane, and B. Atmani, “A data mining-based approach to predict strain situations in hospital emergency department systems,” in 10the International Conference of Modeling and Simulation - MOSIM’14, Nancy - France, 2014, pp. 1–8.

[17] F. Harrou, Y. Sun, F. Kadri, S. Chaabane, and C. Tahon, “Early detection of abnormal patient arrivals at hospital emergency department,” in 2015 International Conference on Industrial Engineering and Systems Management (IESM), 2015, pp. 221–227.

[18] F. Harrou, Y. Sun, and F. Kadri, “Enhanced monitoring of abnormal emergency department demands,” in 15th International Conference on Intelligent Systems Design and Applications (ISDA), 2015, pp. 692–696.

[19] F. Kadri, F. Harrou, and Y. Sun, “A multivariate time series approach to forecasting daily attendances at hospital emergency department,” in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1–6.

[20] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, “LSTM: A Search Space Odyssey,” Mar. 2015.

[21] K. Cho et al., “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” arXiv:1406.1078 [cs, stat], Jun. 2014.

[22] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton, *Speech Recognition with Deep Recurrent Neural Networks*. 2013.

[23] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, “A Novel Connectionist System for Unconstrained Handwriting Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855–868, May 2009.

[24] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Dec. 2014.

[25] M. C. Mukkamala and M. Hein, “Variants of RMSProp and Adagrad with Logarithmic Regret Bounds,” *arXiv:1706.05507 [cs, stat]*, Jun. 2017.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[27] Y. Gal and Z. Ghahramani, “A Theoretically Grounded Application of Dropout in Recurrent Neural Networks,” *arXiv:1512.05287 [stat]*, Dec. 2015.