

Do Outpatient Services Impact Length of Stay? A Preliminary Analysis through Deep Learning Classification

Giorgio Leonardi^{1,2,†}, Stefania Montani^{1,2,†} and Manuel Striani^{1,2,*,†}

¹DISIT, Computer Science Institute, University of Piemonte Orientale, 15121 Alessandria, Italy

²Laboratorio Integrato di Intelligenza Artificiale e Informatica in Medicina DAIRI; Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo, 15121 Alessandria, Italy

Abstract

Timely discharge of hospitalized patients can prevent complications and reduce costs. In this paper, we have investigated whether outpatient services, i.e., diagnostic exams or specialist consultations provided by external wards, have a clear impact on Length of Stay (LOS). In particular, we have worked on an event log of more than 7000 real patient traces, logging the sequence of outpatient services provided during hospitalization, and we have classified the traces into long (≥ 20 days) versus short (< 20 days) LOS, resorting to a deep learning approach that adopts a Long Short-Term Memory (LSTM) network. The very high quality of the classification results suggests that outpatient services play a significant role in determining LOS, and hospitals should work on optimizing their organization.

Keywords

Process trace classification, LSTM, Length of Stay,

1. Introduction

Length of Stay (LOS) is a clinical metric that measures the time elapsed between the hospital admission of a patient and her/his discharge. Unnecessary days in the hospital can lead to increased hospital-acquired patient complications (e.g., infections) and increased costs for patients and healthcare systems. Delays in hospital discharge may be related to unnecessary waiting, poor organization of care, delays in decision-making, or difficulties related to discharge planning [1].

In Italy, a patient is typically admitted to the ward where s/he will undergo her/his main procedure or intervention; however, during the stay, s/he very often needs a set of additional procedures, carried out by different wards, such as diagnostic exams or specialistic consultations, which are referred to as "outpatient services provided during hospitalization" (OSs henceforth). The admission ward has to request an OS, and wait until the external resources are available for its completion, possibly experiencing a delay.

In this paper, we aimed at verifying whether OSs have a major impact on LOS. To this end, we collected the data of more than 7000 patients admitted to Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo in Alessandria, Italy, over the years 2022-2023. By means of a pre-processing step, we transformed such data into an event log [2], where every patient trace is the sequence of the OSs s/he underwent during her/his hospital stay. We then classified the available traces as $LOS \geq 20$ days versus $LOS < 20$ days. The 20-day cutoff for defining "short" versus "long", was not arbitrary but was set in consultation with domain experts: they determined that, based on historical patterns and clinical relevance. This expert-driven threshold ensures the model's outputs align with real-world expectations and decision-making.

Classification was carried out resorting to a Long Short-Term Memory (LSTM) architecture.

4th International Workshop on Process Management in the AI era - ECAI 2025 - Bologna - October 25, 2025

*Corresponding author.

†Authors are listed in alphabetical order and contributed equally to this work.

✉ giorgio.leonardi@uniupo.it (G. Leonardi); stefania.montani@uniupo.it (S. Montani); manuel.striani@uniupo.it (M. Striani)

🌐 <https://upobook.uniupo.it/giorgio.leonardi> (G. Leonardi); <https://upobook.uniupo.it/stefania.montani> (S. Montani);

<https://upobook.uniupo.it/manuel.striani> (M. Striani)

🆔 0000-0002-9533-9722 (G. Leonardi); 0000-0002-5992-6735 (S. Montani); 0000-0002-7600-576X (M. Striani)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The very high quality of our classification results suggests that OSs play a significant role in determining *LOS*, and that a better organization in the request and provision of such external services may have a positive effect on patient management and healthcare costs.

2. Related work

Medical process traces constitute a very important source of information, that has been adopted to support different tasks within the field of process mining [2] specially in the healthcare domain [3, 4]. Trace classification is a classical task, whose state-of-the-art implementations rely on deep learning techniques. Different deep learning architectures have been proposed to this end. Due to the sequential nature of traces, however, Recurrent Neural Networks (RNNs) [5] represent a quite natural approach. RNNs indeed are able to capture both short and long term dependencies between the activities of a trace. Within RNNs, Long-Short Term Memory (LSTM) networks [6] constitute a particularly performing approach, due to their capability of learning the complex dynamics within the temporal ordering of input sequences, as they implement a long-term memory where the information flows from cell to cell with minimal variations, keeping certain aspects constant during the processing of all inputs. The works in [7, 8] are examples of applications of LSTMs to process traces. Another variant of RNNs is presented in [9], where a Memory Augmented Neural Network (MANN), able to learn even longer dependencies, is proposed; training is however more expensive.

The approaches in [10, 11], instead, rely on a Transformer, an architecture that substitutes the recurrence by the attention mechanism [12]. We plan to test transformers for trace classification in our future work.

3. Deep learning trace classification to assess LOS

The deep learning architecture we have adopted exploits a Long Short-Term Memory (LSTM) network, as shown in Figure 1. The input consists of a trace, i.e., a sequence of activities (namely, a sequence of OSs). Each OS is preliminarily converted into a numerical format using one-hot encoding.

The architecture features two LSTM layers that capture temporal dependencies within the trace. The LSTM's hyperparameters were chosen experimentally via small-scale tuning rather than a formal search. In particular, first LSTM layer has 256 units with a *tanh* activation function. It is followed by the second LSTM layer that consists of 128 units with *tanh* activation, that is then passed through a dropout layer with a 0.5 rate, which helps prevent overfitting by randomly deactivating half of the neurons during training. After the LSTM layers, a dense layer with 16 units and a *relu* activation function is applied to learn complex relationships and reduce dimensionality. Finally, a classification layer with a *sigmoid* activation function outputs a probability score for binary classification.

4. Experimental results

Our real world dataset encompasses a total of 7393 patients, representing a varied population with broad clinical profiles. The mean *LOS* is 11.10 days, with a standard deviation of 9.01 days, and 12 OS activities on average (ranging between 2 and 236). The dataset is imbalanced, with 89.06% of the traces categorized as class 0 ($LOS < 20$ days) and 10.94% classified as class 1 ($LOS \geq 20$ days). We then implemented an oversampling technique relying on SMOTE [13], augmenting the underrepresented class up to 2427 items.

The results are shown in Table 1. The average plots of loss and accuracy for each epoch are shown in Figure 2. The experimental outcomes describe a robust performance of the LSTM-based architecture. The precision of 0.91 and the recall of 0.90 in the short *LOS* group (< 20 days – class 0) show the network's ability to capture temporal dependencies relevant to shorter hospital stays. A precision of 0.75 and a recall of 0.78 for the long *LOS* group (≥ 20 days – class 1) underscore sound performance too, although there remains room for improvement.

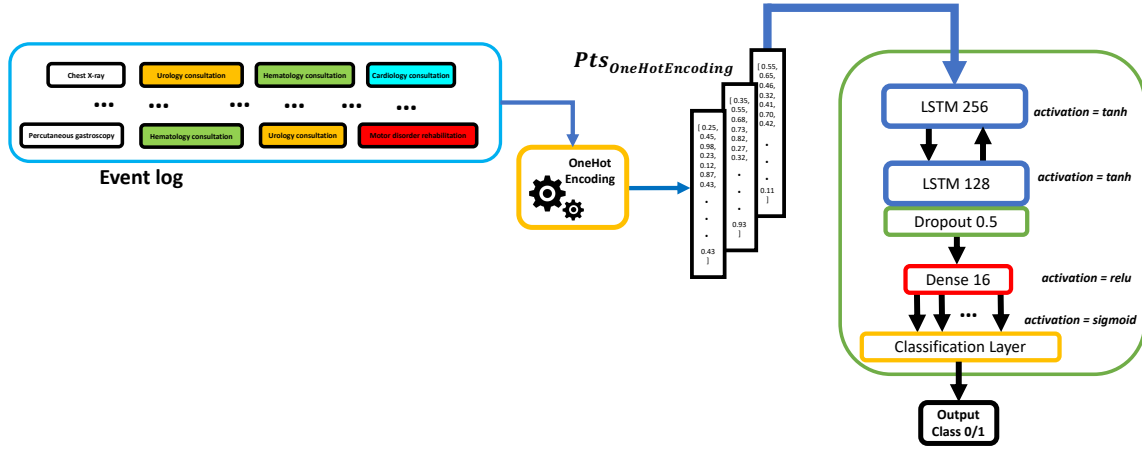


Figure 1: The proposed LSTM-based architecture

Overall, classification results are good, achieving a notable *accuracy* of 0.87 across 30 epochs with a *loss* value of 0.33. Such positive outcomes testify that OSs do play a major role in determining *LOS*, as OS traces classify very well according to the *LOS* dimension.

Table 1
Results obtained by the LSTM-based classifier

Class	Precision	Recall	F1-score	Specificity	MCC	K-Stat	Accuracy
< 20 Days (class 0)	0.91	0.90	0.91	0.78			
≥ 20 Days (class 1)	0.75	0.78	0.76	0.90			
Weighted / Overall	0.87	0.87	0.87	0.81	0.67	0.67	0.87

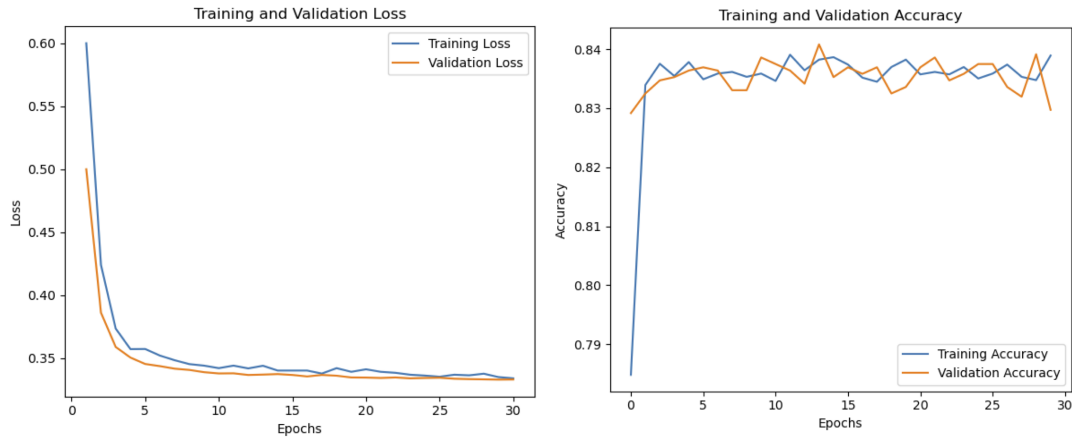


Figure 2: Loss / Accuracy per epoch of the LSTM-based classifier

5. Conclusions

In this paper, we have investigated whether OSs play a significant role in determining patient *LOS*. To this end, we have collected a large real-world event log, where each trace is the sequence of the OSs applied to the patient at hand during her/his hospital stay. The very good classification results of these traces along the dimension of *LOS* (≥ 20 days versus < 20 days) prove that the sequence of the OSs is strongly related to the stay duration, and should be the object of proper organizational optimization strategies. In the future, we will therefore further analyze the OS traces, e.g., by means of process

mining techniques [2], in order to learn intra-ward and/or intra-disease process models, in search of bottlenecks and other useful information for organizational improvements.

As regards the classification model, we have currently exploited a LSTM architecture; in the future, we would like to test a Transformer-based network, which is progressively becoming the state of the art also in process classification and prediction. Moreover, given the intrinsic "black-box" nature of deep learning approaches, we plan to adopt proper explainability techniques, possibly along the lines we investigated in [14].

Acknowledgements This work was supported by the project *PNRR-NODES Bando per Imprese - 2023-2025 - TOT-AL: Ottimizzazione della Transizione Ospedale Territorio presso l'Ospedale di Alessandria*. We are grateful to R. Bellini and C. Zanelli for their work on data provision.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] A. Rojas-García, S. Turner, E. Pizzo, E. Hudson, J. Thomas, R. Raine, Impact and experiences of delayed discharge: A mixed-studies systematic review, *Health expectations : an international journal of public participation in health care and health policy* 21 (2018) 41–46. doi:10.1111/hex.12619, © 2017 The Authors Health Expectations Published by John Wiley & Sons Ltd.
- [2] W. M. P. van der Aalst, *Process Mining - Data Science in Action*, Second Edition, Springer, 2016. URL: <https://doi.org/10.1007/978-3-662-49851-4>. doi:10.1007/978-3-662-49851-4.
- [3] J. Munoz-Gama, N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana, D. Aloini, I. A. Amantea, R. Andrews, M. Arias, I. Beerepoot, E. Benevento, A. Burattin, D. Capurro, J. Carmona, M. Comuzzi, B. Dalmas, R. de la Fuente, C. Di Francescomarino, C. Di Ciccio, R. Gatta, C. Ghidini, F. Gonzalez-Lopez, G. Ibanez-Sanchez, H. B. Klasky, A. Prima Kurniati, X. Lu, F. Mannhardt, R. Mans, M. Marcos, R. Medeiros de Carvalho, M. Pegoraro, S. K. Poon, L. Pufahl, H. A. Reijers, S. Remy, S. Rinderle-Ma, L. Sacchi, F. Seoane, M. Song, A. Stefanini, E. Sulis, A. H. ter Hofstede, P. J. Toussaint, V. Traver, Z. Valero-Ramon, I. van de Weerd, W. M. van der Aalst, R. Vanwersch, M. Weske, M. T. Wynn, F. Zerbato, *Process mining for healthcare: Characteristics and challenges*, *Journal of Biomedical Informatics* 127 (2022) 103994. URL: <https://www.sciencedirect.com/science/article/pii/S1532046422000107>. doi:<https://doi.org/10.1016/j.jbi.2022.103994>.
- [4] L. Aversano, M. Iammarino, A. Madau, G. Pirlo, G. Semeraro, *Process mining applications in healthcare: a systematic literature review*, *PeerJ Comput. Sci.* 11 (2025) e2613. URL: <https://doi.org/10.7717/peerj-cs.2613>. doi:10.7717/PEERJ-CS.2613.
- [5] R. Pascanu, Ç. Gülçehre, K. Cho, Y. Bengio, *How to construct deep recurrent neural networks*, in: Y. Bengio, Y. LeCun (Eds.), *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014*. URL: <http://arxiv.org/abs/1312.6026>.
- [6] S. Hochreiter, J. Schmidhuber, *Long short-term memory*, *Neural Comput.* 9 (1997) 1735–1780. URL: <https://doi.org/10.1162/neco.1997.9.8.1735>. doi:10.1162/NECO.1997.9.8.1735.
- [7] J. Evermann, J. Rehse, P. Fettke, *Predicting process behaviour using deep learning*, *Decis. Support Syst.* 100 (2017) 129–140. URL: <https://doi.org/10.1016/j.dss.2017.04.003>. doi:10.1016/J.DSS.2017.04.003.
- [8] N. Tax, I. Verenich, M. La Rosa, M. Dumas, *Predictive Business Process Monitoring with LSTM*

- Neural Networks, Springer International Publishing, 2017, p. 477–492. URL: http://dx.doi.org/10.1007/978-3-319-59536-8_30. doi:10.1007/978-3-319-59536-8_30.
- [9] M. A. Khan, H. Le, K. Do, T. Tran, A. Ghose, K. H. Dam, R. Sindhgatta, Memory-augmented neural networks for predictive process analytics, CoRR abs/1802.00938 (2018). URL: <http://arxiv.org/abs/1802.00938>. arXiv:1802.00938.
 - [10] Z. A. Bukhsh, A. Saeed, R. M. Dijkman, Processtransformer: Predictive business process monitoring with transformer network, 2021. URL: <https://arxiv.org/abs/2104.00721>. arXiv:2104.00721.
 - [11] I. Donadello, J. Ko, F. M. Maggi, J. Mendling, F. Riva, M. Weidlich, Knowledge-driven modulation of neural networks with attention mechanism for next activity prediction, 2023. URL: <https://doi.org/10.48550/arXiv.2312.08847>. doi:10.48550/ARXIV.2312.08847. arXiv:2312.08847.
 - [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, volume 30, 2017, pp. 5998–6008. URL: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.
 - [13] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.
 - [14] G. Leonardi, S. Montani, M. Striani, Explainable process trace classification: An application to stroke, J. Biomed. Informatics 126 (2022) 103981. URL: <https://doi.org/10.1016/j.jbi.2021.103981>. doi:10.1016/J.JBI.2021.103981.