Aalborg Universitet

Patient Event Sequences for Predicting Hospitalization Length of Stay

Hansen, Emil Riis; Nielsen, Thomas Dyhre; Larsen, Thomas Mulvad; Strausholm, Mads Nibe; Sagi, Tomer; Hose, Katja

Published in:
Artificial Intelligence in Medicine

DOI (link to publication from Publisher):
10.1007/978-3-031-34344-5_7

Publication date:
2023

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Hansen, E. R., Nielsen, T. D., Larsen, T. M., Strausholm, M. N., Sagi, T., & Hose, K. (2023). Patient Event Sequences for Predicting Hospitalization Length of Stay. In J. M. Juarez, M. Marcos, G. Stiglic, & A. Tucker (Eds.), Artificial Intelligence in Medicine: 21st International Conference on Artificial Intelligence in Medicine, AIME 2023, Portorož, Slovenia, June 12–15, 2023, Proceedings (pp. 51-56). Springer. https://doi.org/10.1007/978-3-031-34344-5_7

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: March 22, 2026

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

Hospitalization Length of Stay Prediction using Patient Event Sequences

Emil Riis Hansen[1] (✉)[0000−0003−4103−1244], Thomas Dyhre Nielsen[1][0000−0002−4823−6341], Thomas Mulvad[2][0000−0002−4457−3965], Mads Nibe Strausholm[2][0000−0002−4099−3209], Tomer Sagi[1][0000−0002−8916−0128], and Katja Hose[1,3][0000−0001−7025−8099]

[1] Department of Computer Science, Aalborg University, Aalborg, Denmark
{emilrh,tdn,tsagi,khose}@cs.aau.dk
[2] Unit of Business Intelligence, North Denmark Region, Aalborg, Denmark
{tml,mns}@rn.dk
[3] TU Wien, Vienna, Austria
katja.hose@tuwien.ac.at

Abstract. Predicting patients' hospital length of stay (LOS) is essential for improving resource allocation and supporting decision-making in healthcare organizations. This paper proposes a novel approach for predicting LOS by modeling patient information as sequences of events. Specifically, we present a transformer-based model, termed Medic-BERT (M-BERT), for LOS prediction using the unique features describing patients' medical event sequences. We performed empirical experiments on a cohort of more than $45k$ emergency care patients from a large Danish hospital. Experimental results show that M-BERT can achieve high accuracy on a variety of LOS problems and outperforms traditional non-sequence-based machine learning approaches.

Keywords: length of stay prediction · transformers · sequence models

1 Introduction

Increasingly scarce hospital resources challenge (often oversaturated) hospital wards, with a negative impact on the quality of health care at the hospitals [1] Models for predicting the remaining time of patient admissions, i.e., patient length of stay (LOS), could be invaluable for healthcare facilities to plan the availability of beds, staff, and other essential resources. For instance, automatic prediction of discharge time could be used in administrative planning systems for preemptively freeing in-hospital resources to alleviate hospital ward oversaturation [20]. However, LOS prediction is a challenging problem, requiring methods for handling missing data [18] and integration of temporal event dependencies.

Previous work on LOS prediction models patient hospitalizations using tabular data with imputation techniques for replacing missing values [2]. While tabular data is the most common data representation in Machine Learning (ML) models, it has several drawbacks. Among others, it does not provide immediate

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

support for integrating the temporal dependencies between observations, such as the order of treatments, or the time of conducted procedures. Moreover, standard ML techniques for tabular data, such as artificial neural networks (ANN), gradient boosting (GB), and support vector machines (SVMs), require complete data, hence often relying on imputation techniques when data is incomplete. However, missing data observations in healthcare data are often not missing at random (NMAR), meaning that the mere fact that an observation is missing is in itself important information [11].

To alleviate the problem of temporal dependencies and missing data, attention models, also known as transformers, have recently been investigated for Electronic Health Record (EHR) data formatted as sequences of medical events [10]. However, embedding-based transformer approaches have, to the best of our knowledge, not previously been applied to LOS prediction. This work examines how attention models, can be utilized for this task.

Attention models using self-attention alleviate the inefficiency of recurrence networks for long sequences [19]. However, they still capture significant sequential information by learning from the order of tokens in the sequence. In medical data, multiple observations may be given the same timestamp, with no meaning assigned to their individual order within the corresponding event. For example, a blood panel drawn from a patient contains several individual measurements whose internal order is insignificant. Based on layers of transformer encoders, we propose a revised attention model, henceforth called Medic-BERT (M-BERT), based on the Natural Language Processing (NLP) model BERT [6] and its revised semi-supervised training method. We employ the model for LOS prediction based on sequences of patient specific medical events happening during hospitalization and which exhibit the event concurrences common in patient data. We evaluate our method on a cohort of more than $45k$ patient admissions from a large Danish hospital with diverse medical events, such as vital measurements, medication administration, laboratory tests, and conducted procedures.

The rest of this paper is structured as follows. Section 2 reviews related work on LOS prediction. Section 3 presents our model for representing hospitalizations as event sequences and our proposed M-BERT model for this unique data. Section 4 describes the large dataset used to evaluate this work and Section 5 the evaluation and its results. We conclude with Section 6. This work is an extended version of Hansen et al. [7] published at AIME 2023.

2 Related Work

The LOS prediction problem can be stated in different ways varying in the required resolution of the prediction, applied methods, and structure of patient data. Iwase et al. [9] investigate the use of Machine Learning (ML) methods for LOS prediction of intensive care unit patients (ICU). Using Random Forest (RF), Gradient Boosting (GB), and Artificial Neural Networks (ANN) technologies, on binary stratification of patient admissions into long (more than one week) and short (less than one week) stays. RF showed high predictive performance. In contrast to the tabular structure of their data, we instead attempt to utilize

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

the temporal dependencies within the data by structuring the patient data as medical event sequences. In Batista et al. [4], patients are stratified into three categories (LOS < 3, $3 <$ LOS < 10, LOS $>= 10$) using RF and Support Vector Machines (SVM). The highest performance was achieved for RF models. While the nature of their ML models requires manual feature selection and missing data imputation, our method does not require any feature selection for in-hospital measurements, nor does it rely on data imputation.

The most challenging LOS prediction task is regression-based LOS, where the precise hospitalization LOS as a real value is to be predicted. Barsasella et al. [3] investigated a range of classical AI models, such as Decision Trees (DTs), RFs, Logistic Regression (LR), and SVMs for real-valued LOS prediction. While standard ML models work for tabular structured patient data, they neglect the important temporal dependencies between in-hospital patient events. While much work has been done in LOS prediction and regression using standard ML methods, only a few works attempt the usage of sequence models.

Seminal work in the area of transformer models for patient sequence data includes that of Li et al. [12] and Rasmy et al. [16], where sequences of diagnosis codes for consecutive hospitalizations are used as input to a BERT like transformer model for diagnosis code prediction tasks. While Rasmy et al. [16] investigate prolonged LOS > 7 days event prediction as a model pre-training task, they do not predict future LOS for new patients. Furthermore, while their approaches are only investigated for sequences consisting of diagnosis codes, we evaluate sequences consisting of multiple patient event types.

In Meng el al. [13], a transformer model is used to predict the onset of depression. They integrate various medical types for patient event sequences, including diagnosis and medication events. While their sequences include code-based medical events, we extend with medical measurement events, such as laboratory tests and vital measures while encoding the event value as part of the event token.

The work closest to ours is that of Song et al. [19], where an attention model is used to predict categorical LOS based on patient sequences. Whereas their approach targets dense time-series data, where the same measurements are present at every timestep, instead, we investigate a learned embedding approach where the input sequences consist of codes with different types of medical events.

3 Transformer Models for EHR

In this section, we describe how EHR data can be modeled as medical event sequences and our usage of transformer models for learning from these sequences.

3.1 Hospitalizations as Event Sequences

Patient hospitalizations can be naturally modeled as sequences of medical procedures for determining, measuring, or diagnosing the patient's condition. Other medical procedures are therapeutic and intend to treat or cure the patient. To

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

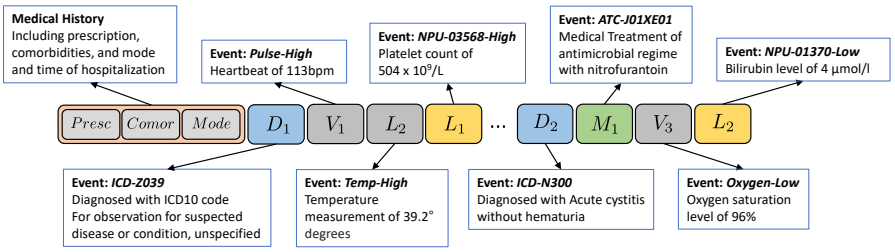


Fig. 1. Illustration of a patient event sequence. Starting with the patient's medical history as summarized in Table 1, the patient is initially diagnosed with the ICD-10 code *Z039*. Subsequent vital measurements and laboratory tests are performed to monitor the patient's state and determine the patient's underlying condition. Consequently, the patient is diagnosed with acute cystitis without hematuria (ICD-10 code *N300*), and antibiotic treatment is initiated with nitrofurantoin (ATC code *J01XE01*). After additional procedure and therapeutic medical events, the patient is released from the hospital.

standardize how medical procedures are described, medical facilities code procedure concepts using accepted medical taxonomies often used in ML applications [8], such as the Anatomical Therapeutic Classification (ATC) [17] for medication administration and the International Classification of Diseases and Related Health Problems (ICD) [5] for condition diagnosis. Hence, a patient hospitalization can be described as the sequence of concept tokens detailing medical procedures pertaining to a patient coded using medical concepts from accepted medical taxonomies. An example patient hospitalization sequence is illustrated in Figure 1.

Furthermore, as a patient's medical history is crucial for correct management and treatment, we pre-pend the patient's medical history to the hospitalization sequence as a tokenized vector as illustrated in Figure 1. The vector consists of 38 tokens describing the patient's medical history, including comorbidities from the Charlson Index [21], five years of medications prescription history grouped by the first level of the ATC hierarchy [17], and the mode, time, and initial triage category [23] of hospitalization. The historical information included in the vector is summarized in Table 1.

3.2 Hospital Measurement Events

For some medical procedures, such as vital measurements and laboratory tests, a numerical measurement value accompanies the procedure. While other works in transformer models for EHR data disregard the numerical values of measurements [13,16,12], we instead integrate this

Table 1. Patient medical history.

Data	#Tokens
Comorbidities	18
Prescription history	14
Mode, time, & triage	6

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

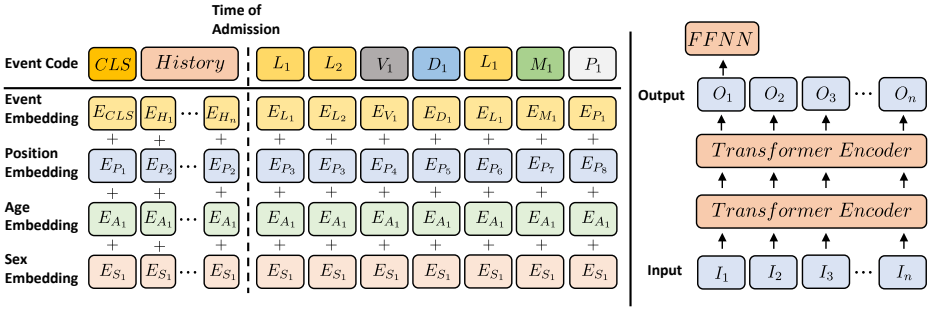


Fig. 2. Medical event sequence pre-pended with the patient's medical history.

information as part of the patient input sequences because numerical measurement values add important information regarding the state of a patient. For example, the knowledge that a temperature measurement was performed is naturally important information. Still, from the measurement value of $40.1°C$, we learn that the patient has a fever. Using patient-specific threshold values for measurements based on age, gender, and pregnancy status, we map measurement values into tokens representing either normal, abnormal-low, or abnormal-high findings. For example, given that we measure an albumin level of 56 g/L for a 31-year-old male patient, we would create the token *albumin-high* to reflect that the value of the measurement was above what is expected (36-48 g/L) for a patient with the given demography.

3.3 Transformer Models for EHR Data

To investigate the challenge of LOS, we examine a modified version, henceforth termed Medic-BERT (M-BERT), of the Bidirectional Encoder Representations from Transformers (BERT) [6] model for EHR data. BERT is an NLP model based on a stack of encoder layers from the transformer architecture introduced by Vaswani et al. [22]. We argue that sequence models, such as BERT exhibit properties beneficial for solving medical tasks based on EHR data. The transformer encoder naturally handles the complex long-term dependencies that occur between medical concepts through its utilization of multi-head self-attention. BERT can naturally integrate disparate modalities, such as diagnostic end therapeutic events, as each event is encoded as an n-dimensional vector token. Furthermore, BERT naturally operates in domains with irregular intervals between events, as is the case with EHR data. We, therefore, investigate our modified version of BERT for LOS prediction for patient event sequences. The M-BERT architecture is illustrated in Figure 2.

M-BERT learns an embedding for each medical event token while trained toward LOS prediction. The position embedding enables the model to learn from the temporal dependencies within a sequence. We use a static position embedding as described in Waswani et al. [22]; however, modified for the usage on medical event sequences as described in Section 3.4. As patient demographics are a vital part of any medical prediction model, we pay special attention to

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

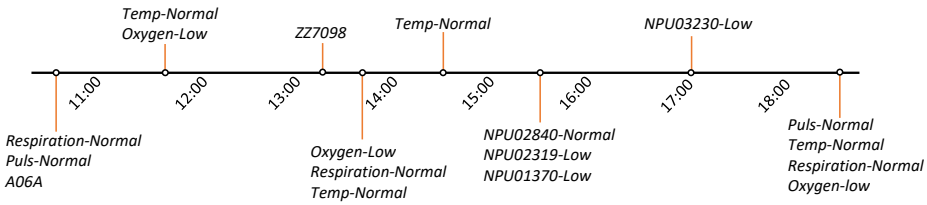


Fig. 3. Patient event sequence illustrating events grouping together.

this information by adding a trainable age and sex embedding at each medical event [12]. Furthermore, as in the original BERT model, we use the special classification (CLS) token as a final aggregate representation of the sequence. Hence, as illustrated in Figure 2, the CLS representation is fed to a linear output layer for LOS classification and regression tasks.

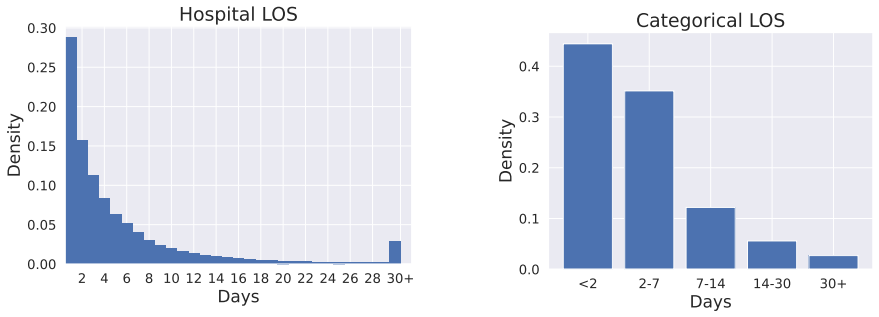
3.4 Position Embedding

Due to the nature of patient care and hospital administration, some measurement events tend to chunk together. Clinicians will often conduct various measurements, such as blood pressure, temperature, and heart rate, over a short time and later persist the information into the patient's EHR. Hence, we are sometimes prevented from knowing the specific times and order of medical events. This effect is most frequent for vital measurements and laboratory tests. Multiple laboratory tests are often conducted on the same patient sample, such as a single blood sample, making it impossible to know the chronological order for such measurement events. An example sequence of patient events chunking together is illustrated in Figure 3. To enable the model to understand that these events have no fixed ordering, we assign the same position embedding to events co-located in time as illustrated in the position embedding of Figure 2. For example, the L_1 and L_2 events are both mapped to the E_{P_3} position embedding.

4 Data

We compiled a Danish dataset of patients admitted to a large hospital in northern Jutland from the period 2018-2021 to investigate transformer models for LOS prediction. The dataset consists of $48,177$ emergency care patients with admissions longer than one day. Figure 4 illustrates the distribution of the remaining length of stay for the patient cohort. In a clinical setting of resource allocation for emergency care patients, we are mostly interested in the planning of patient care for a couple of weeks in advance. Hence, we clip long patient stays to 30 days of admission to better fit and optimize for the clinical setting, as we are not interested in precisely predicting the long-tail distribution of the data.

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7



(a) Histogram of length of stay.

(b) Categorical length of stay.

Fig. 4. Illustrations for length of stay distribution over patient hospitalizations.

In Denmark, each person can be uniquely identified by an identification number from the central person register (CPR) henceforth referred to as a CPR-number. As medical events are each associated with a unique CPR-number, it is possible to connect the medical data pertaining to a patient from disparate databases. As mentioned in Section 3.1, we divide patient medical data into historic information and admission-specific information. Historic information includes prescriptions, comorbidities, and mode and time of hospitalization, whereas admission information includes laboratory tests, vital measurements, hospital-administered medicine, and procedure codes.

Danish laboratory tests are coded using the Nomenclature for Properties and Units (NPU) terminology [15]. NPU ensures that laboratory tests are standardized and patient examinations can be used and understood by all clinicians. As detailed in Section 3.2, we integrate the semantics of laboratory test results as part of the event token by appending either high, low, or normal with respect to the patient demographics to the laboratory event token, thus encoding the meaning of the result. Vital measurements are extracted from a system called Clinical Suite and consist of the 7 most common vital observations, including temperature, oxygen saturation, BMI, pulse, respiration rate, blood pressure (systolic and diastolic), and oxygen supplement. As with laboratory tests, patient-specific thresholds are used to encode the meaning of the result into event tokens. In-hospital administered medication is coded using the ATC taxonomy [17] and consists of more than 5k chemical substances. Procedure codes specify in-hospital procedures performed

Table 2. Concept types with their occurrences in the dataset.

Event Type	Tokens	Occurrences
Lab Tests	748	2,774,790
Vital Measures	22	837,931
Medication	1,441	376,591
Procedures	2049	247,924
History	81	1,880,580

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

on patients. While diagnosis events are an important medical modality, the dataset does not contain the time of such events. Hence, we omit diagnosis events from patient sequences. The event types with distinct tokens and total occurrence in the data are summarized in Table 2.

5 Empirical Evaluation and Results

In this section, we explain the experimental settings and results of the empirical evaluation.

5.1 Experimental Setting

Transformer models are trained, as a rule, using unsupervised pre-training for learning general token embeddings, followed by supervised fine-tuning targeting a specific downstream task. As we are only focusing on a single prediction task, we directly train model parameters and token embeddings toward the downstream task of LOS prediction without pre-training. The experimental code is available online[4].

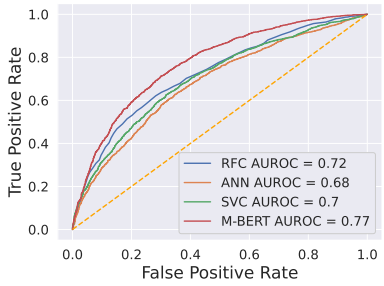
We use the admission data gathered within the first 24 hours of admissions for LOS predictions. Hence, we remove sequence events happening after 24 hours of admission. We evaluate our approach on three LOS experiments of increasing complexity. The first two experiments are a **Binary** classification of LOS > 2 days, and a three-class **Category** task of LOS > 2, $2 \leq$ LOS ≤ 7, and LOS > 7 days with class balances as illustrated in Figure 4b. The last experiment, termed **Real**, is a regression task with the objective of predicting the LOS as the real number, with a histogram of admission times as illustrated in Figure 4a.

We evaluate our approach against three ML models: RF, ANN, and SVM. We use implementations of the models from the Sklearn library [14] with standard model hyperparameters. In preparing samples for these models, we use the latest measured value for each event type (within 24 hours of admission) as input features [9]. Subsequently, the mean of variables is used for imputation of missing data values, and variables are scaled to values between 0 and 1. Lastly, a chi^2 test is used in feature selection for the selection of the 50 most relevant features.

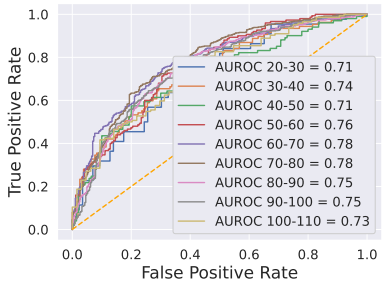
Our model is trained on a random split distributed as 80/10/10 of all patient samples for training, validation, and testing. We use the loss of the evaluation data for early stopping training if the loss does not go down within ten training epochs. The model architecture has six hidden layers with an intermediate layer size of 288, eight attention heads, and input token embeddings have a size of 288. We truncate sequences to 256 tokens, as most sequences adhere to this limit. To counter overfitting, we add a dropout layer with a probability of 10% after the output of the final encoder layer, attention dropout at every layer with a probability of 10%, and weight decay of $0,003$. Furthermore, experiments were performed with a learning rate of 1e-5.

[4] `https://github.com/dkw-aau/medic_transformer`

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7



(a) AUROC curves for the **Binary** LOS experiment over all methods.



(b) **Binary** experiment AUROC curves for different age groups over M-BERT.

Fig. 5. AUROC plots for experimental results on the binary prediction task.

5.2 Results

Table 3 presents the Area Under the Receiver Operating Characteristics (AUROC) and harmonic mean of precision and recall (F1) values for the **Binary** and **Category** experimental settings and Mean Absolute Error (MAE) and Mean Squared Error (MSE) for the LOS regression task. Furthermore, Figure 5a presents the AUROC curves for the **Binary** experimental setting. M-BERT outperforms the traditional ML techniques in all experimental settings. The results indicate that transformer models may be able to leverage the temporal dependencies inherent to patient EHR data for increased predictive accuracy.

Table 3. Experimental results.

	Binary		**Category**		**Real**	
	AUROC	**F1**	**AUROC**	**F1**	**MAE**	**MSE**
RF	0.72	0.70	0.66	0.45	4.18	39.08
ANN	0.67	0.68	0.63	0.43	4.09	38.10
SVM	0.70	0.70	0.65	0.38	3.56	43.36
M-BERT	**0.78**	**0.77**	**0.74**	**0.54**	**3.42**	**37.48**

Furthermore, being a transformer-based model, M-BERT overcomes the challenge of missing data and imputation, as patient sequences are not required to contain the same medical events, nor the same sequence length. Figure 5b presents the AUROC for binary LOS prediction stratified by age groups for those with more than 100 patient samples. Interestingly, the model performance is stable for different age groups, with an AUROC of 0.78 for the age groups 60-70 and 70-80 years. Furthermore, stratification based on the sex of patients yielded similar results, pointing to the robustness of M-BERT for LOS classification.

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

6 Conclusion

We have presented a novel approach for predicting LOS by modeling patient information as event sequences. Our approach adapts the transformer machine-learning approach for sequence prediction, which is able to handle the unique features of medical event sequences, namely grouped events and a variety of data types. Our empirical evaluation on a large cohort of emergency care patients from a Danish hospital demonstrates that our model can achieve high accuracy on various LOS problems, while outperforming traditional non-sequence machine learning approaches. Future work could include pre-training of the transformer-based model on a medical task to further improve its performance. Overall, the proposed approach has the potential to improve resource allocation and support decision making in healthcare organizations by providing accurate predictions of LOS. All experimental code is available online[4].

Acknowledgments

This work is partially supported by the Poul Due Jensen Foundation and the Region North Denmark Health Innovation Foundation.

References

1. af Ugglas, B., Djärv, T., Ljungman, P.L., Holzmann, M.J.: Association between hospital bed occupancy and outcomes in emergency care: a cohort study in stockholm region, sweden, 2012 to 2016. Ann. Emerg. Med. **76**(2), 179–190 (2020)
2. Bacchi, S., Tan, Y., Oakden-Rayner, L., Jannes, J., Kleinig, T., Koblar, S.: Machine learning in the prediction of medical inpatient length of stay. Intern. Med. J. **52**(2), 176–185 (2022)
3. Barsasella, D., Bah, K., Mishra, P., Uddin, M., Dhar, E., Suryani, D.L., et al.: A machine learning model to predict length of stay and mortality among diabetes and hypertension inpatients. Medicina **58**(11), 1568 (2022)
4. Batista, R.W., Sanchez-Arias, R.: A methodology for estimating hospital intensive care unit length of stay using novel machine learning tools. In: ICMLA '20. pp. 827–832. IEEE (2020)
5. Cartwright, D.J.: Icd-9-cm to icd-10-cm codes: what? why? how? (2013)
6. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT '19. pp. 4171–4186 (2019). https://doi.org/10.18653/v1/n19-1423
7. Hansen, E.R., Nielsen, T.D., Mulvad, T., Strausholm, M.N., Sagi, T., Hose, K.: Patient event sequences for predicting hospitalization length of stay. In: Proceedings of the 21st International Conference on Artificial Intelligence in Medicine (2023)
8. Hansen, E.R., Sagi, T., Hose, K., Lip, G.Y.H., Larsen, T.B., Skjøth, F.: Assigning diagnosis codes using medication history. Artif. Intell. Medicine **128**, 102307 (2022)
9. Iwase, S., Nakada, T.a., Shimada, T., Oami, T., Shimazui, T., Takahashi, N., et al.: Prediction algorithm for icu mortality and length of stay using machine learning. Scientific reports **12**(1), 1–9 (2022)

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections.
The Version of Record is available online at: https://doi.org/10.1007/978-3-031-34344-5_7

10. Lequertier, V., Wang, T., Fondrevelle, J., Augusto, V., Duclos, A.: Hospital length of stay prediction methods: A systematic review. Med. Care **59**(10), 929–938 (2021)
11. Li, J., Yan, X.S., Chaudhary, D., Avula, V., Mudiganti, S., Husby, H., et al.: Imputation of missing values for electronic health record laboratory data. NPJ digital medicine **4**(1), 1–14 (2021)
12. Li, Y., Rao, S., Solares, J.R.A., Hassaine, A., Ramakrishnan, R., Canoy, D., Zhu, Y., Rahimi, K., Salimi-Khorshidi, G.: Behrt: transformer for electronic health records. Scientific reports **10**(1), 1–12 (2020)
13. Meng, Y., Speier, W., Ong, M.K., Arnold, C.W.: Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression. IEEE J. Biomed. Inform. **25**(8), 3121–3129 (2021)
14. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.e.a.: Scikit-learn: Machine learning in Python. JMLR **12**, 2825–2830 (2011)
15. Pontet, F., Petersen, U.M., Fuentes-Arderiu, X., Nordin, G., Bruunshuus, I., Ihalainen, J., et al.: Clinical laboratory sciences data transmission: the npu coding system. Stud. Health Technol. Inform. **150**, 265 (2009)
16. Rasmy, L., Xiang, Y., Xie, Z., Tao, C., Zhi, D.: Med-bert: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. NPJ digital medicine **4**(1), 1–13 (2021)
17. Ronning, M.: A historical overview of the atc/ddd methodology. WHO drug information **16**(3), 233 (2002)
18. Sha, Y., Wang, M.D.: Interpretable predictions of clinical outcomes with an attention-based recurrent neural network. In: ACM BCB '17. pp. 233–240 (2017)
19. Song, H., Rajan, D., Thiagarajan, J., Spanias, A.: Attend and diagnose: Clinical time series analysis using attention models. In: AAAI '18. vol. 32 (2018)
20. Stone, K., Zwiggelaar, R., Jones, P., Mac Parthaláin, N.: A systematic review of the prediction of hospital length of stay: Towards a unified framework. PLOS Digital Health **1**(4), e0000017 (2022)
21. Sundararajan, V., Henderson, T., Perry, C., Muggivan, A., Quan, H., Ghali, W.A.: New icd-10 version of the charlson comorbidity index predicted in-hospital mortality. Journal of clinical epidemiology **57**(12), 1288–1294 (2004)
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., el al.: Attention is all you need. Adv. Neural Inf. Process. Syst. **30** (2017)
23. Wireklint, S.C., Elmqvist, C., Göransson, K.E.: An updated national survey of triage and triage related work in sweden: a cross-sectional descriptive and comparative study. SJTREM '21 **29**(1), 1–8 (2021)