Leveraging Temporal Learning with Dynamic Range (TLDR) for Enhanced Prediction of Outcomes in Recurrent Exposure and Treatment Settings in Electronic Health Records

Jingya Cheng, M.B.[1*]; Jonas Hügel, PhD.,[1,2*]; Jiazi Tian, M.Sc.[1] ; Alaleh Azhir, MD, MSc[1,3]; Shawn N. Murphy, MD, PhD[4] ; Jeffrey G. Klann, MEng, PhD[1#] ; Hossein Estiri, PhD[1†#]

[1] Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[2] University Medical Center Göttingen, Department of Medical Informatics, Göttingen, Germany and University of Göttingen, Campus Institute Data Science, Section of Medical Data Science, Göttingen, Germany

[3] Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

[4] Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

* These authors contributed equally as first authors

[#] These authors contributed equally as senior authors

[†] Corresponding author: Hossein Estiri

Email: hestiri@mgh.harvard.edu

399 Revolution Drive, Suite 790, Somerville, MA, 02145, USA

Abstract

Background: The temporal sequence of clinical events is crucial in outcomes research, yet standard machine learning (ML) approaches often overlook this aspect in electronic health records (EHRs), limiting predictive accuracy.

Methods: We introduce Temporal Learning with Dynamic Range (TLDR), a time-sensitive ML framework, to identify risk factors for post-acute sequelae of SARS-CoV-2 infection (PASC). Using longitudinal EHR data from over 85,000 patients in the Precision PASC Research Cohort (P2RC) from a large integrated academic medical center, we compare TLDR against a conventional atemporal ML model.

Results: TLDR demonstrated superior predictive performance, achieving a mean AUROC of 0.791 compared to 0.668 for the benchmark, marking an 18.4% improvement. Additionally, TLDR's mean PRAUC of 0.590 significantly outperformed the benchmark’s 0.421, a 40.14% increase. The framework exhibited improved generalizability with a lower mean overfitting index (-0.028), highlighting its robustness. Beyond predictive gains, TLDR's use of time-stamped features enhanced interpretability, offering a more precise characterization of individual patient records.

Discussion: TLDR effectively captures exposure-outcome associations and offers flexibility in time-stamping strategies to suit diverse clinical research needs.

Conclusion: TLDR provides a simple yet effective approach for integrating dynamic temporal windows into predictive modeling. It is available within the MLHO R package to support further exploration of recurrent treatment and exposure patterns in various clinical settings.

Introduction

In temporal outcomes research, particularly in settings involving repeated exposures or a combination of exposures and recurrent treatments, the sequential ordering of events associated with each exposure or intervention may hold significant prognostic value. The relative importance of these temporal patterns may vary depending on the duration and context of the observation period.

For instance, consider a patient with diabetes (exposure) receiving repeated insulin adjustments (treatments) with an outcome of interest, such as hypoglycemic episodes requiring medical intervention. In this scenario, temporal information—such as the occurrence of specific symptoms (e.g., dizziness or confusion) during the most recent insulin dose adjustment or the history of hypoglycemic events following prior adjustments—may offer greater predictive utility compared to static, atemporal data.

In epidemiology and biostatistics, repeated-measures designs are typically employed to model data collected from the same individuals over time, allowing for the assessment and longitudinal tracking of effects.[1–6] These studies utilize structured longitudinal panels with meticulously designed protocols for data collection, treatment administration, and outcome measurement at predefined intervals. This methodological framework is particularly effective for analyzing

structured data, facilitating a deeper understanding of the impact of various factors on clinical outcomes.

Electronic health records (EHRs) serve as a rich source of observational data encompassing patients' health status, treatments, diagnoses, and other clinical parameters, inherently structured within a temporal framework. However, leveraging this temporal information for analytical purposes is challenging due to two fundamental limitations. First, event timestamps correspond to the time of documentation rather than the actual occurrence of events, leading to potential temporal misalignment. Second, the recording of clinical events in EHRs is inherently asynchronous, occurring at irregular intervals rather than in a temporally uniform manner, posing methodological and computational challenges for conventional analytical approaches.[7,8] Patient heterogeneity and missing data further complicate the temporal usage of EHR data, as does the complexity of temporal problem formation.[9,10] Consequently, most standard machine learning (ML) approaches applied to EHR data often fail to account for its informative temporal structure, potentially limiting their ability to capture the underlying temporal dynamics of clinical events.[11]

Nevertheless, the temporal structure of EHRs can be effectively leveraged in downstream ML applications.[10] In particular, within contexts involving repeated exposures or a combination of exposures and recurrent treatments, the sequential ordering of events associated with each exposure or intervention may hold significant predictive value.[10] The relative importance of these temporal patterns varies depending on the observation period, and incorporating this information can enhance the performance of machine learning models applied to EHR data, improving their predictive accuracy and clinical utility.[12,13] This temporal knowledge enables the

identification of patterns and dependencies that are critical for forecasting future outcomes, such as cancer,[14,15] Alzheimer's Disease,[8] and sepsis.[16,17]

The dynamic integration of temporal data into feature engineering is crucial in these repeated-exposure/treatment settings. By tailoring temporal features to the evolving context of an exposure or intervention, researchers can enhance the accuracy and robustness of predictive models, thereby improving their clinical applicability.[18–20] While some studies have sought to refine methodologies for analyzing temporally stamped records to predict risk factors, the performance of machine learning models for recurrent outcomes in EHRs that incorporate temporal characteristics often remains modest (e.g., the Area Under the Receiver Operating Characteristic curve [AUROC] frequently hovers around 0.75).[17,21–26] This limitation may stem from the inherently non-stationary nature of temporal timestamps in EHRs, wherein their relevance and influence in modeling health outcomes dynamically shift in relation to the exposure or treatment of interest.[27]

Post-acute sequelae of SARS-CoV-2 (PASC) represent a health outcome that can be examined within the framework of recurring exposures and outcomes. In this context, individuals may experience multiple COVID-19 infections, with PASC potentially associated with a specific infection episode. This allows for the study of temporal patterns and the cumulative impact of recurrent exposures on the development and progression of PASC over time.[27–29]

Here, we introduce a time-sensitive machine learning approach, Temporal Learning with Dynamic Range (TLDR), to identify risk factors for PASC. Our approach, while straightforward, integrates dynamic temporal windows of patient data into the predictive framework. Each window delineates distinct chronological phases: the first encapsulates the patient’s medical history, the second encompasses events occurring between the historical period and

exposure/treatment, and the third captures the interval between exposure/treatment—specifically, COVID-19 infection in this study—and the subsequent development of post-acute sequelae of SARS-CoV-2 (PASC). By creating patient-relative dynamic time windows, we significantly enhance model performance and uncover deeper insights into PASC risk factors. We present the TLDR method and compare its efficacy to traditional approaches that do not integrate temporal data.

Method

The validated Precision PASC Research Cohort (P2RC)[29] from Mass General Brigham (MGB) comprises integrated longitudinal clinical data from over 85,000 patients within 12 months of COVID-19 infection. This data was collected from eight hospitals and community health centers across Massachusetts. The inclusion criteria restricted the dataset to COVID-19 infections that occurred between March 6, 2020, and June 7, 2022, with clinical follow-up data extending up to June 7, 2023.

All patients who tested positive for SARS-CoV-2 infection were included in this study. We conducted a retrospective cohort study utilizing electronic health records (EHR) to track the incidence of SARS-CoV-2 infection from March 2020 to June 2022. Patients who developed Post-Acute Sequelae of SARS-CoV-2 (PASC) within 12 months following infection were classified as PASC group, whereas those who did not develop PASC were designated as Non-PASC group.

The utilization of data for this study was authorized under Mass General Brigham (MGB) institutional review board protocol #2020P001063, which included a specified waiver of consent

for a data-only retrospective study. Mass General Brigham is an integrated healthcare delivery system that serves approximately 1.5 million patients annually.

TLDR: With Outcome date

We introduce the Temporal Learning with Dynamic Range (TLDR) approach, which categorizes clinical data across patient records into distinct categories. This method accommodates multiple exposures or treatments and offers the flexibility to handle recurrent outcomes by treating each outcome individually.

- Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of all clinical events for an individual where each x_k is associated with a date $d(x_k)$

- $T = \{t_1, t_2, \ldots, t_m, t_{outcome}\}$ be a sequence of time points corresponding to patient exposures or treatments, where $t_1$is the time points of first exposure or treatment. Note that t_m is the last exposure/treatment before the outcome, and $t_{outcome}$ is the observed outcome.

- $W = \{W_h, W_P, W_l, W_{outcome}\}$ be the time buffers assigned to the categories *history*, *past*, *last*, and *outcome* respectively. A time buffer is a time interval that is excluded from the label assignment. For example, W_h is the time until the beginning of the observation period.

We categorize X into ($X_{history}$, X_{past}, X_{last}) according to these three criteria:

1. **History**: All clinical events before the first exposure or treatment after the *history* buffer:

$$X_{history} = \{x_k \in X: d(x_k) < t_1 + W_h\}$$

2. **Past**: Clinical events are labeled as *past* if their times of occurrence are from the first exposure or treatment and up to the final exposure or treatment before the observed outcome. This adjusted by the *past* buffer, meaning that there must be an amount of time between the final exposure or treatment and the outcome.

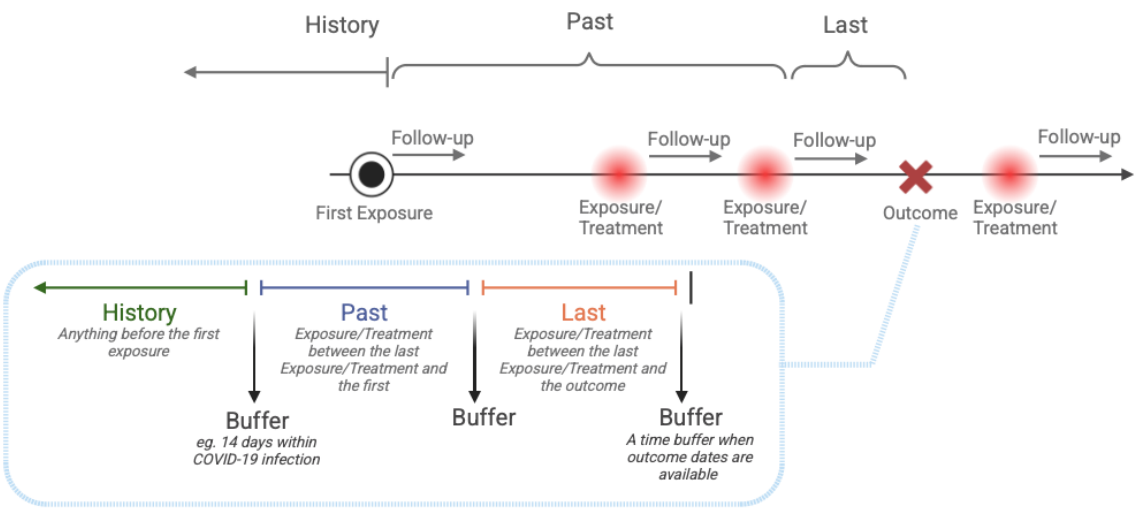
$$X_{past} = \{x_k \in X: t_1 + W_h < d(x_k) \ \leq t_{m-1} + \ W_p\}$$

3. **Last**: Clinical events are labeled as *last* if their timestamps fall within the interval after the last exposure or treatment (adjusted by the *last* buffer) up to the observed outcome (adjusted by the outcome buffer):

$$X_{last} = \{x_k \in X: t_{m-1} + W_P \leq d(x_k) \leq t_{outcome} + W_l + W_{outcome}\}$$

Temporal Learning with Dynamic Range (TLDR)

When outcome exists:



A detailed figure explaining how the exposures/treatments are labeled for each outcome.

When outcome does not exist:

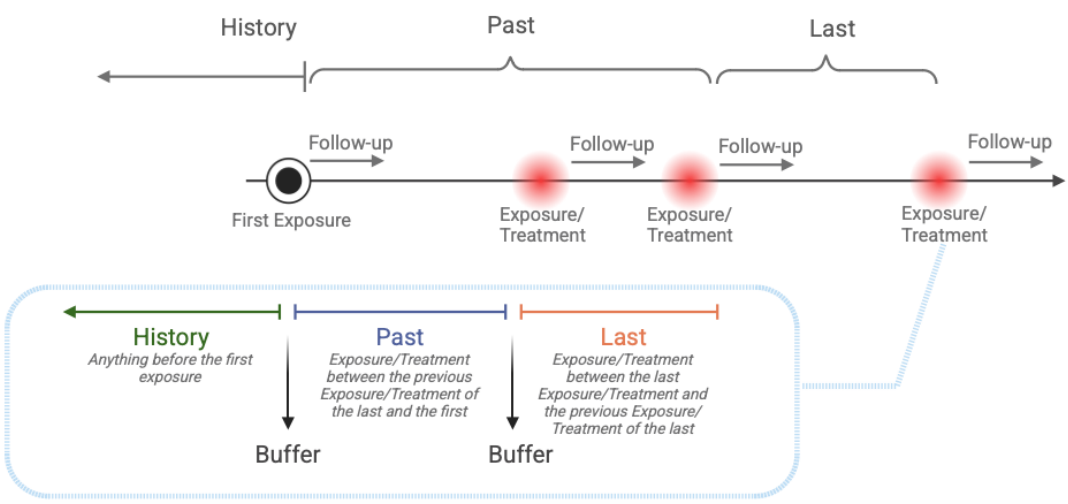


Figure 1. Detailed explanation of the TLDR methodology. All clinical events occurring prior to the initial exposure or treatment are classified as history events. Events that occurred between the first exposure and the time of the last exposure or treatment are labeled as past events. Events happening from the last exposure or treatment until the observed outcome are referred to as the last events. A time buffer is incorporated between each label assignment interval to accommodate a window period for exposure or treatment. TLDR accommodates recurrent outcomes by treating each outcome individually. When no outcome is observed, the only difference is that events occurring from the previous to the last exposure or treatment are labeled as the last events.

This algorithm provides a more precise way to categorize and aggregate features by each outcome date adjusted for a window time (time buffer) for a specific exposure or treatment. With the dynamic window, we are able to capture the temporal features more precisely. Figure 1 shows how the algorithm works to label each exposure or treatment in a dynamic approach. All events that happened before the first exposure or treatment are labeled as *history* events. The events that happened from the first exposure to the time of the last exposure or treatment are labeled as *past* events. The events that happened from the last exposure or treatment to the outcome date are labeled as the *last* events. We allow for flexibility to add a time interval as a parameter of time buffer on each exposure or treatment before or after the *history*, *past*, and *last* labels, and the outcome dates. For example, a buffer can be the 14 days of COVID-19 infection by each time label. For any exposure or treatment where no outcome is observed, the only modification is that events occurring from the previous exposure or treatment to the last exposure or treatment are classified as last events. All other conditions remain unchanged.

Once the data are labeled, all labels (*history, past, last*) are merged and reformatted into a wide format, where the occurrence of the labeled events is recorded in the incidence level (See appendix 2). We use the `MSMR.lite` function from the MLHO R package[30] to perform a series of transformations and feature selections by labeling each feature by categorizing them to "history", "past", and "last". The `buffer` parameter gives flexibility to add a time interval within the infection period. After this step, the transformed data can be fed into a variety of advanced algorithms, such as XGBoost.

Benchmark: Without Label assignment

Our TLDR algorithm provides flexibility to account for temporal relationships between each record by assigning them with *history*, *past*, and *last* labels. To compare our TLDR algorithm with the atemporal method, we also included the benchmark approach, which does not consider the effect of temporal label assignments. The key difference between this benchmarking approach and the TLDR approach is that this process ignores temporal relationships and simply counts all occurrences of each feature for each patient across their entire events. For example, for each patient i, and each clinical record x, we calculate the occurrences of x across all time points of the patient's exposures or treatments, regardless of the label assignments. This method assigns each feature to a distinct column in the output data structure. The value in each of these columns corresponds to the total count of occurrences of that specific feature across all time points of exposures or treatments for each patient. Therefore, instead of obtaining segmented counts like *history*, *past*, and *last* for each encounter, the output simply records the feature occurrence without timestamps for each patient.

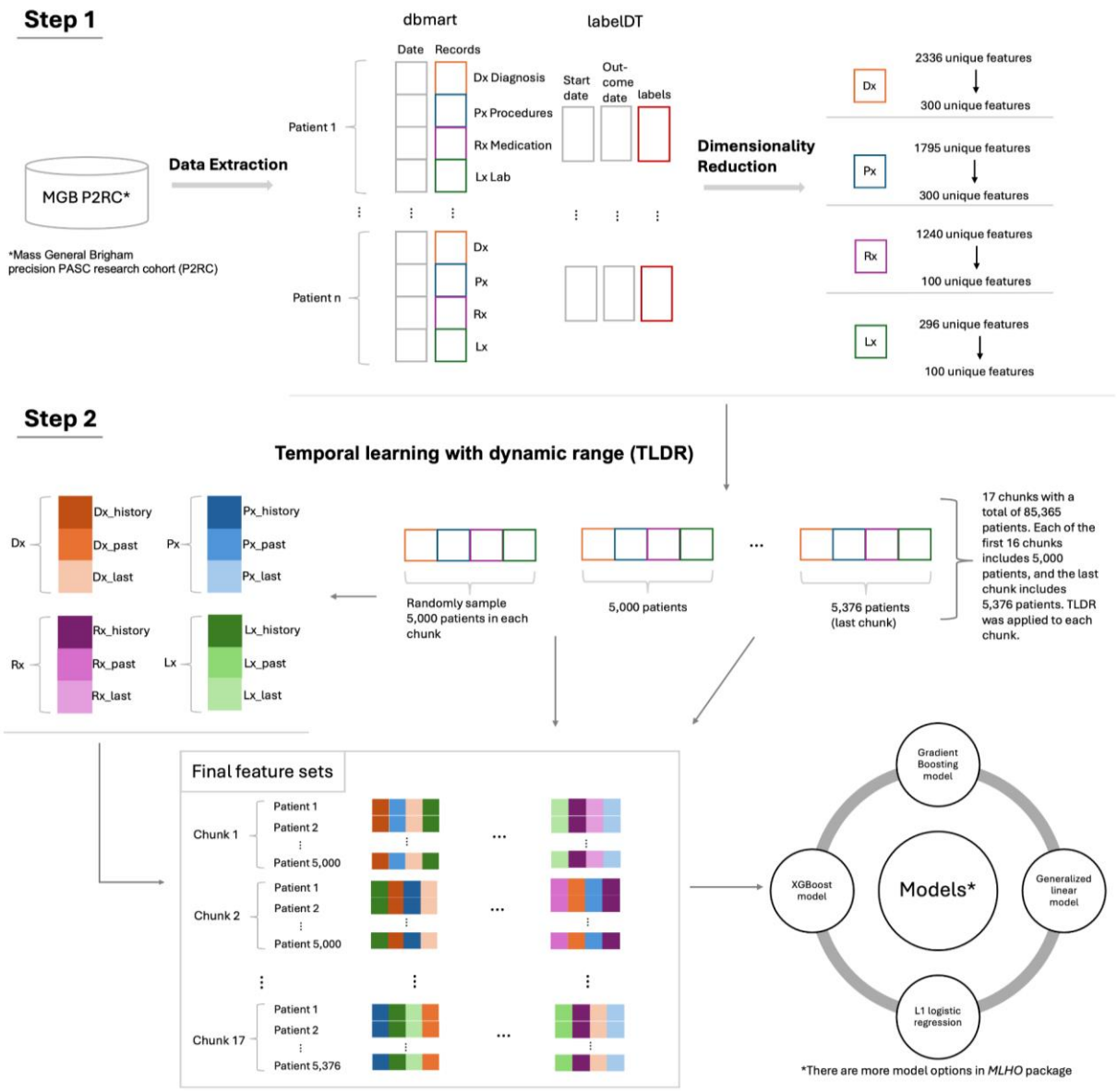


Figure 2. A figure that shows the process of TLDR implementation. The feature dataset encompasses diagnostic, procedural, laboratory, and medication data, while the label dataset includes the dates of exposure or treatment alongside the observed outcomes. The P2RC

cohort was divided into 17 subsets, with each subset subjected to both the TLDR and benchmark approaches using four distinct learning models.

This study incorporated diagnostic, procedural, laboratory, and medication data (Figure 2). The labeled dataset extracted both the date of COVID-19 infection and the date of long COVID development from the MGB P2RC cohort.[29]

Dimensionality reduction

To identify top features for model training, we adopted the Minimize Sparsity, Maximize Relevance (MSMR)[7,12,13] dimensionality reduction technique provided within the MLHO R Package.[30] The MSMR algorithm employs high-performance dimensionality reduction to refine EHR representations. Beginning with an initial set of *N* temporal or atemporal features curated by the TLDR algorithm, it systematically reduces this set to fewer than 200 (a user-defined number) features through a three-step process. This process incorporates frequency-based filtering and information-based selection, including joint mutual information[31–33] analysis, to retain the most predictive and informative features.

Experimental design

The study data were methodically partitioned into 17 subsets, with the initial 16 subsets containing approximately 5,000 patients each. We applied the TLDR technique on all 17 subsets. Subsequent analyses utilized advanced learning models, including the gradient boosting model, generalized linear model, L1 logistic regression, and the XGBoost model, applied individually to each subset to ensure a rigorous evaluation of the data. As a result, we developed and tested 68 (17×4) predictive models.

Model evaluation metrics

We evaluated the discrimination performance by comparing the mean Area Under the Receiver Operating Characteristic (AUROC) and the Precision-Recall Area Under the Curve (PRAUC) between the benchmark approach and the TLDR approach across all 17 data subsets and for each learning model employed. We used a 70-30 train-test split to compute out of sample AUROC and PRAUC. Additionally, to assess the potential for overfitting, we calculated the difference in AUROC performance between the validation and test sets. This analysis allowed us to compare the mean AUROC, PRAUC, and the overfitting metric across the various models applied to the 17 chunks, providing a comprehensive overview of model robustness and predictive accuracy.

Results

In the MGB Precision PASC Research Cohort (P2RC), which encompasses 85,376 patients diagnosed with 98,069 episodes/incidents of SARS-CoV-2 infection, 24,360 individuals developed post-acute sequelae of SARS-CoV-2 (PASC) within 12 months following their infection, while 61,016 did not exhibit PASC (Table 1). For additional details regarding cohort development and validation, refer to Azhir et al (2024).

The cohort's mean age was 53.6 years. Approximately 62.5 percent of the participants were female, 71.4 percent were White, 10 percent were Black or African American, and 7.2 percent were Hispanic. Among those classified as long-haulers, the mean age was 57.1 years. Approximately 64.5 percent of this subgroup were female, 71.4 percent were White, 10.4 percent were Black or African American, and 6.6 percent were Hispanic. The mean Charlson Comorbidity Index was 2.18 for the overall cohort and 3.06 among the long-haulers.

Table 1. Summary statistics of the study population

	Long-haulers (N=24,360)	Overall (N=85,376)
Age		
Mean (SD)	57.1 (17.4)	53.6 (17.2)
Median [Min, Max]	58.0 [19.0, 106]	54.0 [19.0, 106]
Sex		
F	15,707 (64.5%)	53,398 (62.5%)
M	8,653 (35.5%)	31,978 (37.5%)
Race/Ethnicity		
White	17,385(71.4%)	60,936 (71.4%)
Black	2,538 (10.4%)	8,409 (9.8%)
Asian	700 (2.9%)	2,936 (3.4%)
Other	3,059 (12.6%)	10,250(12.0%)
Unknown	678 (2.8%)	2,845 (3.3%)
Hispanic	1,612 (6.6%)	6,110 (7.2%)
Charlson Index		
Mean (SD)	3.06 (2.94)	2.18 (2.48)
Median [Min, Max]	2.00 [0, 19.0]	1.00 [0, 19.0]

Model outputs

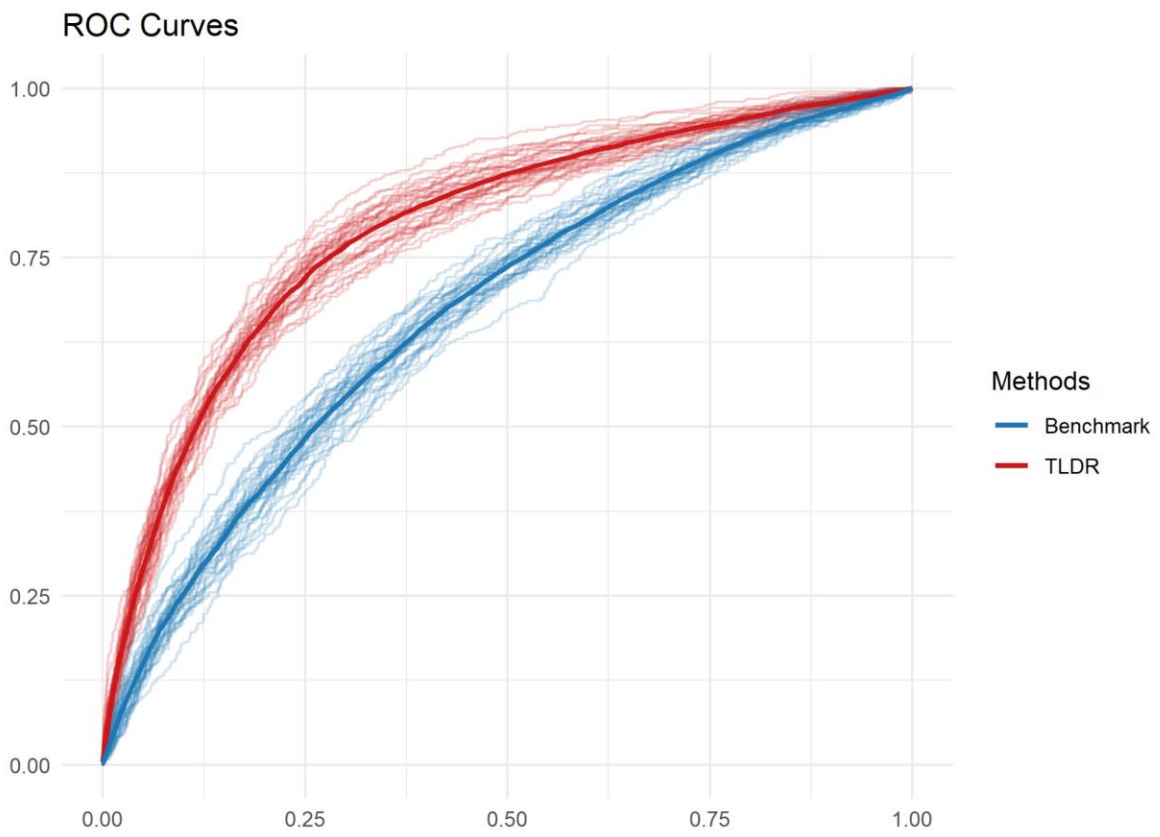


Figure 3. ROC curves comparing the benchmark approach and the TLDR approach

Our analysis utilized 2,336 unique diagnostic features, 1,795 procedural features, 1,240 medication-related features, and 296 laboratory features before dimensionality reduction. We compared the mean AUROC, mean PRAUC, and the mean overfitting measure between the benchmark approach and the TLDR approach (Table 2). The mean AUROC for the TLDR approach was recorded at 0.791, compared to 0.668 for the benchmark (Figure 3), representing an 18.4% improvement. Similarly, the mean PRAUC for the TLDR approach was 0.590, significantly higher than the benchmark's 0.421, yielding a 40.14% increase. Both the benchmark and TLDR approaches demonstrated minimal overfitting, as indicated by negative values in the overfitting indices. Although the benchmark approach displayed an average

overfitting index near zero, suggesting negligible overfitting, the TLDR approach exhibited superior generalization on the test set with a mean overfitting index of -0.028.

Table 2. Model performance metrics for the benchmark approach and the TLDR approach

	Mean AUC (Mean 95% CI)	Delta Mean AUC (P)	Mean PRAUC	Delta Mean PRAUC (P)	Mean Overfitting*	Delta Mean Overfitting (P)
Benchmark	0.668 (0.64, 0.668)	0.123 (<0.001)	0.421	0.169 (<0.001)	-0.001	-0.028 (<0.001)
TLDR	0.791 (0.767, 0.791)		0.590		-0.029	

To evaluate the statistical significance of these differences, we employed the Wilcoxon test with corrections for multiple comparisons across the three metrics. The results confirmed that the improvements observed with the TLDR approach were statistically significant, underscoring its enhanced performance in both discrimination and generalization compared to the benchmark methodology.

Table 3. Model performance metrics by models for the benchmark approach and the TLDR approach

Model	Approach	Mean		
		AUC	PRAUC	Overfitting
GBM*	Benchmark	0.681	0.434	0.003

	TLDR	0.810	0.611	-0.018
GLM**	Benchmark	0.652	0.404	0.006
	TLDR	0.769	0.565	-0.029
L1 logistic regression	Benchmark	0.662	0.416	0.00
	TLDR	0.784	0.578	-0.040
XGBoost	Benchmark	0.678	0.431	-0.012
	TLDR	0.803	0.605	-0.029

* Gradient boosting machine, ** generalized linear regression

We compared the mean AUC, PRAUC, and the overfitting measures for both TLDR and benchmark approach among the gradient boosting model, generalized linear model, the L1 logistic regression model, and the XGBoost model (Table 3). For the gradient boosting model, the TLDR approach yielded a mean AUC of 0.810, compared to 0.681 for the benchmark approach. Similarly, the PRAUC for the TLDR approach registered at 0.611, surpassing the benchmark's 0.434. The overfitting measure indicated a mean of -0.018 for the TLDR approach and 0.003 for the benchmark.

In the generalized linear model, the TLDR approach demonstrated a mean AUC of 0.769, exceeding the benchmark's 0.652. The mean PRAUC was 0.565 for the TLDR approach, compared to 0.404 for the benchmark. The overfitting measure for the TLDR approach was -0.029, while it was 0.006 for the benchmark.

For the L1 logistic regression model, the TLDR approach achieved a mean AUC of 0.784, outperforming the mean AUC of 0.662 of the benchmark. The mean PRAUC for the TLDR approach was 0.578, versus 0.416 for the benchmark. The overfitting measure was notably lower at -0.040 for the TLDR approach, compared to 0.00 for the benchmark.

Regarding the XGBoost model, the TLDR approach reported a mean AUC of 0.803, higher than the benchmark's 0.678. The mean PRAUC was 0.605 for the TLDR approach and 0.431 for the benchmark. The overfitting measure was -0.029 for the TLDR approach and -0.012 for the benchmark.

Feature Examples

The temporal features create an overall stronger predictive model by considering the feature only when it is important. Top features appear in the model in the temporal window that matters. It also can sometimes mean that temporal features have a higher relative model ranking than their equivalent non-temporal benchmark feature (See Appendix 1). Z79 Long Term (Current) Drug Therapy [last] is one such example. Similarly, surgical procedures on arteries and veins [past and last] appear near the top of the list, while [history] appears near the bottom and the feature is not even part of the benchmark model.

Discussion

Despite the inherent temporal variability and noise present in electronic health record (EHR) data, advanced methodological approaches have demonstrated the capacity to extract informative features that enhance predictive modeling. In this study, we addressed a persistent challenge in modeling recurrent exposure and treatment effects using real-world EHR data and standard machine learning (ML) techniques. Our findings indicate that the proposed Temporal Learning with Dynamic Ranges (TLDR) framework outperformed a conventional atemporal modeling approach by 18.4% in predicting multi-episode exposure/outcome associations, specifically in the context of

post-acute sequelae of SARS-CoV-2 infection. Furthermore, the TLDR models exhibited improving generalizability and robustness in real-world applications.

Our findings further demonstrated that, beyond their enhanced predictive performance, the time-stamped features in the TLDR framework provide greater precision in characterizing individual electronic health records, thereby improving interpretability. Notably, a record of Long Term Drug Therapy documented after the last infection episode likely signifies the clinical relevance of ongoing pharmacologic treatment for chronic conditions, which may indicate underlying health vulnerabilities that compromise resilience to post-acute sequelae of SARS-CoV-2 infection. Likewise, Sleep Disorders and Abnormalities of Heart Rate emerged as particularly influential when recorded in historical data, highlighting their potential role as early indicators of long-COVID risk.

Interpreting differences in feature importance between models can be challenging, as time-stamped features in the TLDR framework highlight specific records that may carry varying levels of importance across TLDR and benchmark approaches. However, given that our results demonstrate the clear superiority of the TLDR approach, the necessity of interpreting these feature importance differences is mitigated. Instead, the focus should remain on the TLDR model's enhanced predictive performance and its ability to provide more precise and clinically meaningful insights.

The models predominantly prioritized diagnostic features, with laboratory results and medication data contributing less frequently. Notably, IV solutions demonstrated significantly greater predictive utility when historical data were excluded. This is likely attributable to its strong association with hospitalization, where a post-COVID-19

hospitalization serves as a more robust predictor of PASC than a history of prior hospitalizations alone.

A well-documented limitation of recurrent deep learning models has been their constrained capacity for long-term memory retention, which can hinder their efficacy in modeling complex temporal dependencies. The advent of transformer-based architectures has substantially mitigated these limitations by leveraging self-attention mechanisms that enable more effective contextualization of sequential information over extended time horizons. Notably, transformer models have demonstrated robust applicability beyond natural language processing, extending to structured clinical and biomedical data, including electronic health records, physiological time-series data, and genomic sequences.[34–36] Given these advancements, future research should rigorously evaluate the predictive performance and clinical utility of transformer-based models in structured healthcare datasets. Comparative analyses with traditional recurrent architectures will be essential to elucidate the extent to which transformers enhance predictive accuracy, interpretability, and generalizability in real-world clinical applications.

Implementing a fully dynamic approach for selecting time ranges presents inherent challenges, as researchers may require different time-stamping strategies depending on their specific use cases. To address this, we have incorporated flexibility into the TLDR framework by allowing for dynamic ranges through the use of optional buffer periods and benchmark dates. Future applications across diverse clinical scenarios can provide

valuable insights to further refine and enhance the algorithm, ensuring its adaptability to a broader range of predictive modeling tasks.

While the TLDR methodology introduced in this study is inherently simple, it offers a straightforward yet effective means of integrating dynamic temporal windows of patient data into predictive modeling, with each window delineating distinct clinical periods. Specifically, we sought to differentiate between a patient's medical history prior to their initial exposure (i.e., pre-COVID-19 infection), the interval between this history and their most recent exposure, and the period between the last exposure and the outcome, defined as the onset of PASC. To facilitate broader applicability, we have provided the complete function within the MLHO R package, allowing for modification to support diverse investigations of recurrent treatments or exposures and enabling rigorous evaluation against established temporal modeling approaches.

Data Sharing

The R package associated with this study can be accessed at the following GitHub repository: https://github.com/clai-group/MLHO

Declaration of Interests

The authors declare no competing interests.

Ethics approval

Use of patient data in this study was approved by the Mass General Brigham Institutional Review Board (protocol 2020P001063).

Contributions

H.E., J.C., J.G.K., and J.H. conceived, designed, and planned this study. H.E., J.C., and J.G.K., collected and acquired the data. H.E., J.C., and J.G.K. performed data preparation. J.C. and H.E. analyzed the data. H.E., J.C., A.A., S.N.M., J.H., and J.G.K., interpreted the data. H.E., J.C., J.G.K., and J.H. drafted the paper. All authors critically reviewed the paper and revised the final paper. H.E., J.G.K., and J.C. had access to all the data in the study. All authors approved the decision to submit for publication.

Acknowledgments

This study has been supported by grants from the National Institute of Allergy and Infectious Diseases (NIAID) R01AI165535.

J.Hügel's work was partially funded by a fellowship within the IFI programme of the German Academic Exchange Service (DAAD) and by the Federal Ministry of Education and Research (BMBF) as well by the German Research Foundation (426671079).

References

1. Schober, P. & Vetter, T. R. Repeated measures designs and analysis of longitudinal data: If at first you do not succeed-try, try again. *Anesth. Analg.* **127**, 569–575 (2018).
2. Albert, P. S. Longitudinal data analysis (repeated measures) in clinical trials. *Stat. Med.* **18**, 1707–1732 (1999).
3. Schober, P. & Vetter, T. R. Repeated measures designs and analysis of longitudinal data: If at first you do not succeed—try, try again. *Anesth. Analg.* **127**, 569–575 (2018).
4. Herbert, R. D., Kasza, J. & Bø, K. Analysis of randomised trials with long-term follow-up. *BMC Med. Res. Methodol.* **18**, (2018).
5. Link, M. W. Measuring compliance in mobile longitudinal repeated-measures design study. *Surv. Pract.* **6**, 1–8 (2013).
6. Butera, N. M., Zeng, D., Heiss, G. & Cai, J. Modeling longitudinal change in biomarkers using data from a complex survey sampling design: An application to the Hispanic Community Health Study/Study of Latinos. *Stat. Med.* **42**, 632–655 (2023).
7. Estiri, H., Strasser, Z. H. & Murphy, S. N. High-throughput phenotyping with temporal sequences. *J. Am. Med. Inform. Assoc.* **28**, 772–781 (2021).
8. Estiri, H. *et al.* Temporal characterization of Alzheimer's Disease with sequences of clinical records. *EBioMedicine* **92**, 104629 (2023).
9. Xie, F. *et al.* Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. *J. Biomed. Inform.* **126**, 103980 (2022).
10. Pungitore, S. & Subbian, V. Assessment of prediction tasks and time window selection in temporal modeling of electronic health record data: A systematic review. *J. Healthc. Inform. Res.* **7**, 313–331 (2023).

11. Yang, S., Varghese, P., Stephenson, E., Tu, K. & Gronsbell, J. Machine learning approaches for electronic health records phenotyping: a methodical review. *J. Am. Med. Inform. Assoc.* **30**, 367–381 (2023).
12. Estiri, H. *et al.* Transitive sequencing medical records for mining predictive and interpretable temporal representations. *Patterns (N. Y.)* **1**, 100051 (2020).
13. Estiri, H., Vasey, S. & Murphy, S. N. Transitive sequential pattern mining for discrete clinical data. in *Artificial Intelligence in Medicine* 414–424 (Springer International Publishing, Cham, 2020).
14. Baudrier, C. *et al.* Identifying homogeneous healthcare use profiles and treatment sequences by combining sequence pattern mining with care trajectory clustering in kidney cancer patients on oral anticancer drugs: A case study. *Health Informatics J.* **28**, 14604582221101526 (2022).
15. Hügel, J. *et al.* Temporal characterization and visualization of revolving therapy-events in lung cancer patients. *Stud. Health Technol. Inform.* **316**, 1642–1646 (2024).
16. Jazayeri, A., Yang, C. C. & Capan, M. Frequent temporal patterns of physiological and biological biomarkers and their evolution in sepsis. *Artif. Intell. Med.* **143**, 102576 (2023).
17. van den Berg, M. A. M. *et al.* Development and clinical impact assessment of a machine-learning model for early prediction of late-onset sepsis. *Comput. Biol. Med.* **163**, 107156 (2023).
18. Singh, V., Rana, R. K. & Singhal, R. Analysis of repeated measurement data in the clinical trials. *J. Ayurveda Integr. Med.* **4**, 77–81 (2013).
19. King, B. M. Analysis of Variance. in *International Encyclopedia of Education* 32–36 (Elsevier, 2010).
20. Muhammad, L. N. Guidelines for repeated measures statistical analysis approaches with basic science research considerations. *J. Clin. Invest.* **133**, (2023).
21. Nogues, I.-E. *et al.* Semi-supervised Double Deep Learning Temporal Risk Prediction

(SeDDLeR) with Electronic Health Records. *J. Biomed. Inform.* **157**, 104685 (2024).

22. Ahuja, Y. *et al.* Leveraging electronic health records data to predict multiple sclerosis disease activity. *Ann. Clin. Transl. Neurol.* **8**, 800–810 (2021).
23. Abraham, A. *et al.* Dense phenotyping from electronic health records enables machine learning-based prediction of preterm birth. *BMC Med.* **20**, 333 (2022).
24. Xu, J. *et al.* Data-driven discovery of probable Alzheimer's disease and related dementia subphenotypes using electronic health records. *Learn. Health Syst.* **4**, e10246 (2020).
25. Jorge, A. M. *et al.* Exploration of machine learning methods to predict systemic lupus erythematosus hospitalizations. *Lupus* **31**, 1296–1305 (2022).
26. Song, W. *et al.* Improved accuracy and efficiency of primary care fall risk screening of older adults using a machine learning approach. *J. Am. Geriatr. Soc.* **72**, 1145–1154 (2024).
27. Dagliati, A. *et al.* Characterization of long COVID temporal sub-phenotypes by distributed representation learning from electronic health record data: a cohort study. *eClinicalMedicine* **64**, (2023).
28. Estiri, H., Strasser, Z. H. & Murphy, S. N. Individualized prediction of COVID-19 adverse outcomes with MLHO. *Sci. Rep.* **11**, 5322 (2021).
29. Azhir, A. *et al.* Precision phenotyping for curating research cohorts of patients with unexplained post-acute sequelae of COVID-19. *Med (N. Y.)* 100532 (2024).
30. Jingya Cheng, Jiazi Tian, Jonas Hugel, Hossein Estiri. *MLHO*. (Github, 2021).
31. Bennasar, M., Hicks, Y. & Setchi, R. Feature selection using Joint Mutual Information Maximisation. *Expert Syst. Appl.* **42**, 8520–8532 (2015).
32. Salem, O. A. M., Liu, F., Chen, Y.-P. P. & Chen, X. Feature selection and threshold method based on fuzzy joint mutual information. *Int. J. Approx. Reason.* **132**, 107–126 (2021).
33. Yang, H. & Moody, J. Feature selection based on joint mutual information. in *Proceedings of international ICSC symposium on advances in intelligent data analysis* vol. 1999 22–25 (Citeseer, 1999).

34. Kraljevic, Z. *et al.* Foresight-a generative pretrained transformer for modelling of patient timelines using electronic health records: a retrospective modelling study. *Lancet Digit. Health* **6**, e281–e290 (2024).
35. Wang, L. *et al.* Transformer-based deep learning model for the diagnosis of suspected lung cancer in primary care based on electronic health record data. *EBioMedicine* **110**, 105442 (2024).
36. Antikainen, E. *et al.* Transformers for cardiac patient mortality risk prediction from heterogeneous electronic health records. *Sci. Rep.* **13**, 3517 (2023).