Assessment of prediction tasks and time window selection in temporal modeling of electronic health record data: A systematic review

Sarah Pungitore, MS (corresponding author)

Applied Mathematics Program

College of Science

617 N Santa Rita Ave

Tucson, AZ 85721, United States

spungitore@email.arizona.edu

(520)-621-6559

https://orcid.org/0000-0001-5782-6741

Vignesh Subbian, PhD

Associate Professor

Department of Biomedical Engineering

Department of Systems and Industrial Engineering

The University of Arizona

Tucson, AZ 85721-0020, United States

https://orcid.org/0000-0001-9974-8382

Keywords: machine learning, electronic health records, temporal data, systematic review

Word Count: 3990 (excluding tables and captions); 4543 (including tables and captions)

Statement on Competing Interests: All authors declare that they have no competing interests.

1. ABSTRACT

Temporal electronic health record (EHR) data are often preferred for clinical prediction tasks because they offer more complete representations of a patient's pathophysiology than static data. A challenge when working with temporal EHR data is problem formulation, which includes defining the time windows of interest and the prediction task. Our objective was to conduct a systematic review that assessed the definition and reporting of concepts relevant to temporal clinical prediction tasks. We searched PubMed® and IEEE Explore® databases for studies from January 1, 2010 applying machine learning models to EHR data for patient outcome prediction. Publications applying time-series methods were selected for further review. We identified 92 studies and summarized them by clinical context and definition and reporting of the prediction problem. For the time windows of interest, 12 studies did not discuss window lengths, 57 used a single set of window lengths, and 23 evaluated the relationship between window length and model performance. We also found that 72 studies had appropriate reporting of the prediction task. However, evaluation of prediction problem formulation for temporal EHR data was complicated by heterogeneity in assessing and reporting these concepts. Even among studies modeling similar clinical outcomes, there were variations in terminology used to describe the prediction problem, rationale for window lengths, and determination of the outcome of interest. As temporal modeling using EHR data expands, minimal reporting standards should include time-series specific concerns to promote rigor and reproducibility in future studies and facilitate model implementation in clinical settings.

2. INTRODUCTION

Electronic health record (EHR) systems have been implemented in the majority of hospitals and physician offices in the United States, with an average of 91% of non-federal acute care hospitals having adopted a certified EHR in 2019 [1,2]. EHR systems provide a rich source of timestamped patient data on vital signs, medications, laboratory measurements, clinical observations, and procedures [1]. Along with typical clinical and administrative uses, EHR data has been leveraged for several secondary purposes, including clinical research and disease surveillance [3].

Effective secondary use of EHR data for clinical research often relies on understanding and processing temporal patient data. Although temporal data can be aggregated using descriptive statistics, aggregation limits the usefulness of clinical time-series data by over-simplifying or generalizing sequential patterns [4]. In particular, prediction of clinical outcomes has shifted from using traditional statistical methods and static, cross-sectional data to machine

learning algorithms designed to model outcomes based on a sequence of inputs [5]. For example, temporal clinical data have been used to develop and evaluate machine learning methods to predict acute conditions, such as in-hospital mortality [6] and sepsis [7], and chronic conditions, such as onset of cardiovascular disease [8] and renal function deterioration [9].

In response to the upward trend in using machine learning algorithms for prediction with EHR data, several reviews have provided in-depth analyses of the challenges of processing temporal EHR data and applying models for prediction [1,10-13]. These reviews focused on issues in working with temporal data, including irregularities in data sampling (e.g., worsening patients tend to have a higher frequency of recorded clinical measurements), sparsity and missing data (e.g., incomplete data from patients visiting multiple hospital systems), and patient heterogeneity (e.g., presence of sub-cohorts within a large, comprehensive EHR dataset) [13]. However, few studies have looked at the difficulties inherent in problem formulation with temporal EHR data. Notably, a case study evaluating processing strategies for temporal EHR data [5] assessed the relationship between model performance and indexing methods for prediction of hospital mortality and hypokalemia. The investigators found using the outcome as the reference point for time-series indexing, while prevalent across the selected studies reviewed, provided misleading results and limited the clinical utility of models. However, no systematic reviews, to our knowledge, have evaluated components of time-series problem formulation using EHR data and how well they are reported across studies.

The major components of problem formulation for patient-level time-series prediction are defining the outcome and the time windows of interest [14]. For the outcome, researchers select the event of interest and how it is measured in both cases (those positive for the event) and controls (those negative for the event) during processing. When determining the time windows of interest, researchers select the length of the observation and prediction windows, how they are related to each other, and how the model will update predictions for future patients.

The observation window is defined as the period of time where observations are used to update model predictions [14]. The prediction window, however, has two different meanings depending on the context. First, the prediction window can be the gap between the end of the observation window and the time of prediction [15]. The other definition is the “time-at-risk” where the prediction window is the period of time directly following the observation window

when a prediction can be made such that there is no gap between observation and prediction [14,16]. To standardize the definitions across studies, we define prediction window as an umbrella term for the gap and time-at-risk windows where the gap window is the period of time between observation and prediction and the time-at-risk window is the period of time when a prediction is made. These definitions are illustrated in Figure 1.

This review aims to study how prediction problems are defined and formulated when developing models from temporal clinical data. Specifically, we sought to answer the following:

1. How do researchers establish observation and prediction window lengths and relationships to each other?
2. How is the prediction task determined based on the time windows and outcome of interest?
3. How consistently are these concepts reported across studies?

Fig. 1 Observation and prediction window diagram for the various definitions of the prediction window. For the gap window, a patient is marked as a case when the outcome occurs after a period of time (or "gap") after the observation window. Any patient with outcomes that occur within the gap window or after the label is applied are marked as controls. For the time-at-risk window, a patient is marked as a case if the outcome occurs at any point in the time-at-risk window (which may have a fixed length or vary depending on the amount of data present and the prediction task). All other patients are labeled as controls. Stars represent a hypothetical occurrence of the outcome.

3. MATERIALS AND METHODS

3.1. Search Strategy and Study Selection

We conducted a systematic review of studies based on temporal clinical data modeling using 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [17]. Our search involved two literature databases (PubMed® and IEEE Explore® databases) and a combination of keywords associated with

EHR data ("electronic health records" or "electronic medical records"), time-series analysis ("time-series" or "neural networks"), machine learning ("machine learning" or "deep learning") but excluded imaging ("image" or "imaging") and natural language processing (NLP) modeling ("NLP" or "natural language processing" or "text processing") since their temporal data processing pipelines differ from those for structured, temporal EHR data. The search was restricted to studies published in English from January 1, 2010. After excluding duplicates, we screened titles and abstracts to determine potential studies for review. Studies were included if they applied neural network models for patient-level outcome prediction using temporal EHR data. Studies were excluded based on the following criteria: (1) studies using non-clinical applications, signal processing studies, population-based studies, and other systematic reviews; (2) lack of structured EHR data (e.g., studies using free text such as clinical notes and imaging data); (3) temporal data used to develop static models, or models that were trained on aggregated temporal EHR data. After screening titles and abstracts, we read the full text of the selected articles and applied the same inclusion and exclusion criteria as above to assess the relevancy of the studies. Two reviewers (SP and VS) screened the full text articles. In addition to the literature search, additional studies meeting the inclusion and exclusion criteria were included after a grey literature search using Google Scholar. The process for study selection is summarized in Figure 2.

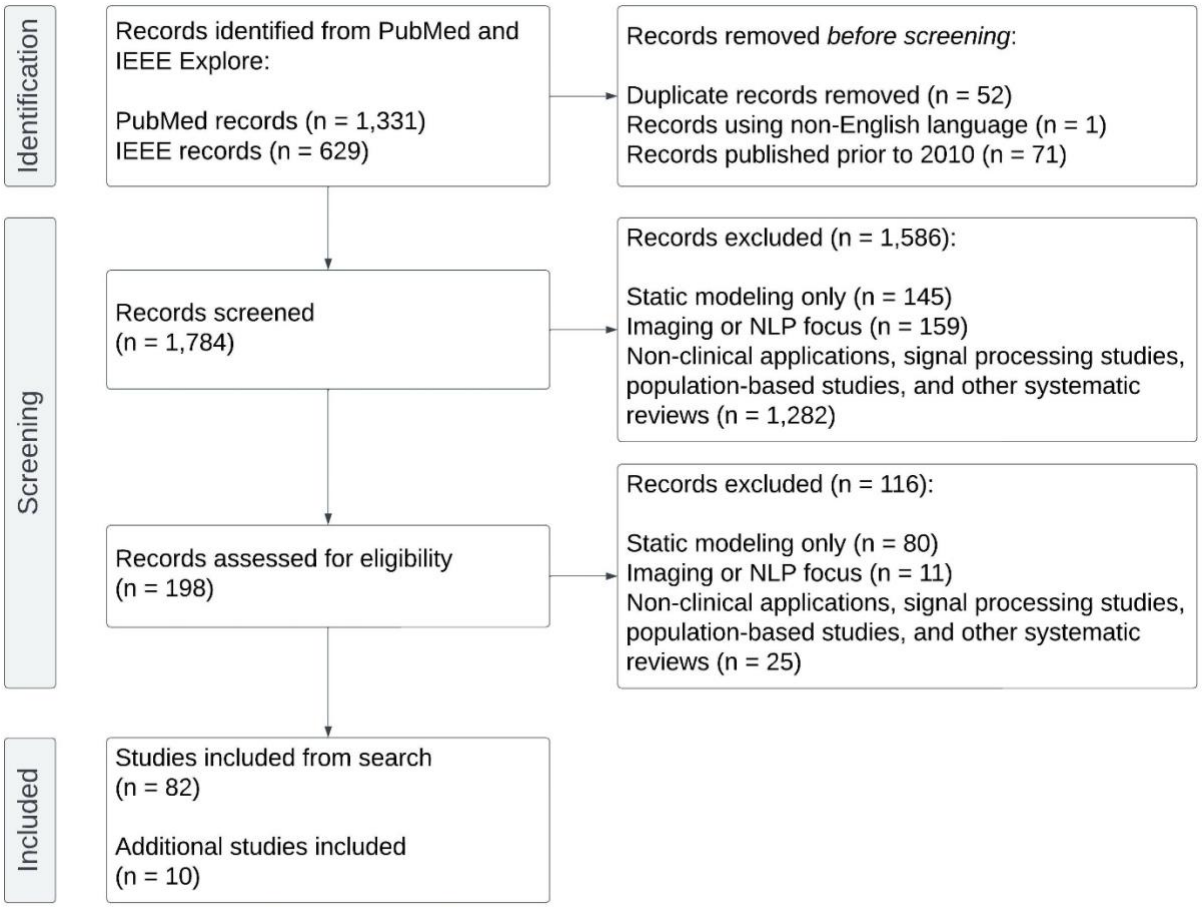

Fig. 2 Study selection flowchart using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [17].

3.2. Data Extraction

We extracted the following characteristics from each study:

- Clinical context: clinical setting; clinical outcome(s) being predicted.
- Data source and sample lengths: source of EHR data; number of observations used for modeling; time of data collection (see supplementary material A).
- Model selection and evaluation: basic model architectures; comparisons to baseline (either static or temporal) models; measures for model evaluation (see supplementary material B).
- Time window selection and evaluation: definition of observation and prediction windows; selection of window lengths; evaluation of models trained with different window lengths.

- Prediction task: research question of interest; prediction frequency.

3.3. Quality Assessment

Quality assessment was performed using two standard frameworks for evaluating minimal reporting in machine learning research, Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [18] and Minimum Information about Clinical Artificial Intelligence Modeling (MI-CLAIM) [19]. For both frameworks, there was a particular focus on reporting of the methods (study design, data processing, and modeling). In addition to TRIPOD and MI-CLAIM, we assessed the extent to which time-series specific concerns (the time windows of interest and prediction task) were reported in the methods to maintain consistency with TRIPOD and MI-CLAIM standards. Appropriate reporting of the time windows of interest was assessed separately from appropriate reporting of the prediction task. Studies were considered to have appropriate reporting of the time windows of interest if the type, length, and relationship of the windows were discussed in the methods. Studies were considered to have appropriate reporting of the prediction task if researchers discussed the outcome of interest and how often predictions were updated in the methods.

4. RESULTS

4.1. Clinical Context

After selecting eligible studies (n = 82) and studies from alternate searches (n = 10), there were 92 studies included for review. Around 96% of the studies could be categorized into prediction related to acute (n = 60, 65%) or chronic (n = 28, 31%) conditions. The remaining 4% modeled prediction of all-cause hospital readmission (n = 4) [20-23]. The most common acute conditions were prediction of sepsis or bacteremia (n = 17) [7,15,24-38], in-hospital mortality (n = 14) [39-52], multi-task learning or general clinical outcome prediction (n = 12, which usually included sepsis and in-hospital mortality predictions along with multiple diagnoses predictions) [6,53-63], and acute kidney injury (n = 5) [64-68]. Some chronic diseases considered for modeling were onset of heart failure (n = 5) [69-73], cancer (n = 4) [10,74-77], cardiovascular disease (n = 4) [8,78-82], arthritis (n = 2) [81,82], and cognitive impairment (n = 2) [83,84].

4.2. Observation and Prediction Windows

4.2.1. Window Definition

In all studies that explicitly mentioned the observation window, the definition was consistent with prior literature. However, the definition of prediction window (if specified) varied between studies. Of the studies where the definition(s) of the prediction window were either assumed or explicitly stated (n = 77, 84%), 21 (27%) defined the prediction window as a gap of time, 48 (62%) defined the prediction window as time-at-risk, and eight (10%) used both definitions for prediction (Table 1). However, 51 (55%) studies had appropriate reporting of both the length and type of the prediction window(s) in the methods while 34 (37%) did not have appropriate reporting of the type only, two (2%) did not have appropriate reporting of the length only, and 5 (5%) did not have appropriate reporting of both length and type. In studies where we found appropriate reporting of the relationship between observation and prediction window(s) (n = 70, 76%), the windows did not overlap and there were clearly defined cut-offs for when the windows of interest began and ended. Different terms were used to describe the prediction window, including "hold-off window" [27], "prediction horizon" [30], "future of any horizon" [59], and "predictive time horizon" [85].

Table 1. List of reviewed studies by prediction window definition. Percentages are calculated out of a total of 92 studies.

Prediction Window Definition	Number (%) of Studies	References
Gap window	21 (23%)	[15,29,31,34-39,47,69,71,74,75,81,86-91]
Time-at-risk window	48 (52%)	[6-8,20-23,30,40-42,45,48,49,51-55,57-59,61-64,66,68,73,76-78,80,82,84,85,92-103]
Both definitions used	8 (9%)	[25,27,43,50,60,65,83,104]
Unclear prediction window definition	15 (16%)	[24,26,28,32,33,44,46,56,67,70,72,79,105-107]

4.2.2. Window Selection

Of the 92 studies, 12 (13%) did not explicitly mention the lengths of the time windows used, 57 (62%) only considered a single observation and prediction window length, 22 (24%) selected several window lengths for evaluation, and one (1%) performed an optimization algorithm to determine optimal window lengths (Table

2). For studies that assessed different observation and prediction window lengths, the optimal window length was selected after training separate models on each window length combination, comparing model performance, and then selecting the window lengths based on the model with the best performance. For example, in a study predicting sepsis [27], investigators first fixed the length of the gap window and varied the observation window from two to 24 hours and then fixed the length of the observation window and varied the gap window from five to 24 hours. They found that the method varying the observation window (and setting the gap window to 0 hours) produced better model performance with a 24 hour observation window as the optimal window length.

Table 2. Summary of evaluation methods for observation and prediction window lengths. Percentages are calculated out of a total of 92 studies.

Method of Evaluation	**Number (%) of Studies**	**References**
Observation and prediction window lengths were not directly discussed	12 (13%)	[7,21,28,33,44,46,70,78,96,100,102,106]
A single observation window length and prediction window length were used in all analyses	57 (62%)	[6,8,20,22-24,26,30,35-38,40-43,45,48,49,51-53,55,56,59-63,65-67,73-77,79-86,88,89,91-95,97,101,103,104,108]
Different observation and/or prediction window lengths were established and model performance was compared for each window length	22 (24%)	[15,25,27,29,31,34,39,47,50,54,64,68,69,71,72,87,90,98,99,105,107]
Optimization algorithm was used to determine the optimal window lengths along with testing different window lengths	1 (1%)	[87]

For window selection, researchers employed two strategies: right and left alignment. These terms were explicitly mentioned in some studies [25,27] and we have broadly applied these to all studies in this review to further characterize window selection. Briefly, right alignment refers to setting the prediction window to a certain length and leaving the observation window variable while left alignment refers to setting the observation window to a certain length and leaving the prediction window variable (Figure 3). Of the reviewed studies, 43 (47%) used right alignment only, nine (10%) used left alignment only, and 28 (30%)

used both such that both window lengths were defined as part of the prediction problem (Table 3). For the remaining studies (n = 12, 13%), it was unclear whether the windows were right or left aligned since the window lengths were not provided in the main text.

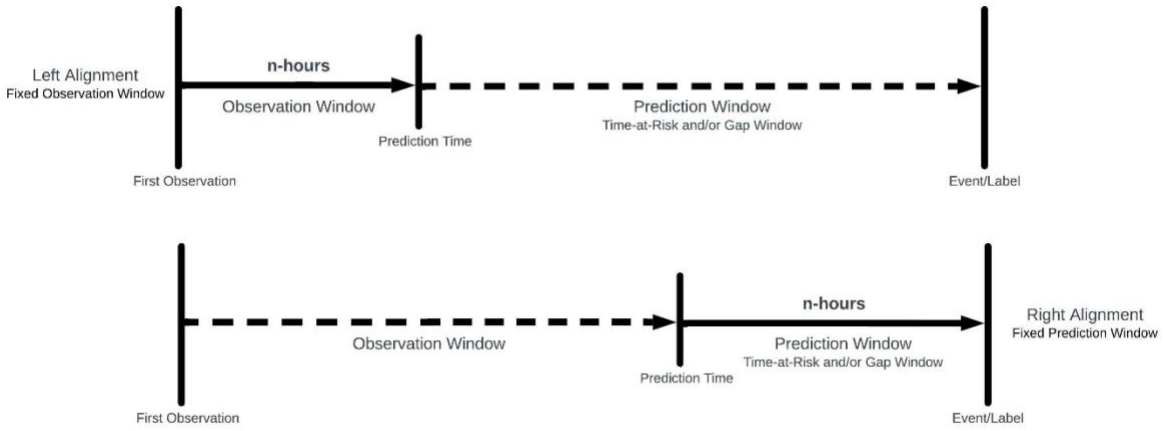

Fig. 3 Visualization of the definitions for left and right alignment. Left alignment refers to a fixed observation window of n-hours with a variable prediction window. Right alignment refers to a fixed prediction window (either time-at-risk window or gap window) of n-hours with a variable observation window. The bolded timeframes are those with pre-specified lengths and the dashed timeframes are variable depending on a patient's length of stay, health record history, or onset time of the outcome.

Table 3. List of reviewed studies using right alignment and/or left alignment. Percentages are calculated out of a total of 92 studies.

Alignment Type	Number of Studies (%)	References
Right alignment	43 (47%)	[7,15,20-22,24,31,35-38,43,47-49,51,53-55,59,61,63,66-68,72,73,76,78,80,84,88,91,92,94,96,97,100-104,108]
Left alignment	9 (10%)	[6,33,40-42,45,62,83,98]
Both alignment types used	28 (30%)	[8,23,25,27,29,30,34,39,50,56-58,60,64,69,71,74,75,77,81,82,85-87,89,90,93,99]

Unclear alignment type	12 (13%)	[26,28,44,46,52,65,70,79,95,105-107]

When selecting the length of the observation and prediction windows, investigators cited clinical relevance or optimal model performance when rationale was provided [43,93]. For chronic diseases, there was less variability in how windows were selected among studies studying the same clinical context. Notably, for different studies predicting cancer onset, the observation and prediction windows were all set to three years and one year, respectively, although multiple types of cancers were considered [74,75,77]. On the other hand, for acute conditions, the length of the observation and prediction windows was dependent on both the clinical context and the individual study. For example, in the 17 studies looking at sepsis or bacteremia, the observation window lengths tested ranged from three hours [26] to 48 or 72 hours [34]. The prediction window lengths also ranged from six hours [24,32,35,36,38,108] to 24 hours [15] although it was not always possible to determine whether the prediction window was referring to a gap window or time-at-risk window.

4.2.3. Effect of Window Selection on Model Performance

Model performance was assessed in 23 (25%) studies after changing the length of the observation and/or prediction windows (Table 2). Of the studies that looked at changes in the observation window (13%), an increase in the observation window corresponded with an increase in model performance as measured by AUC (Table 4). For studies using alternative metrics, an increase in the length of the observation window corresponded with an increase in model performance as measured by accuracy [87] and F-1 score [27]. In contrast to observation window experiments, evaluation of the effect of changing the prediction window length on model performance was complicated by inconsistent reporting of prediction window definitions. In studies where it was either stated or assumed that the gap window was changing between experiments, a decrease in the length of the gap window increased model performance as measured by AUC, accuracy [87,90], sensitivity [64,90], or specificity [31,64,90] (Table 4). However, the result of changing the time-at-risk window on model performance was inconclusive. Model performance as measured by AUC either increased, decreased, or did not show a clear pattern [50] when increasing the length of the time-at-risk window (Table 4).

Table 4. Summary of changes in model performance, as measured by AUC, after increasing the length of the observation, gap, and time-at-risk windows.

Evaluation of Performance	Increase in Performance (AUC)	Decrease in Performance (AUC)
Increase observation window length	[15,25,27,34,39,47,54,69,71,72,87,99]	None
Increase in gap window length	None	[15,25,27,29,31,34,39,47,64,69,71,87,90,99]
Increase in time-at-risk window length	[72,98]	[54,68]

4.3. Prediction Task

For reporting of the prediction task, we found the majority (n = 72, 78%) of studies reported the outcome of interest and how often predictions were updated in the methods while 20 (22%) did not. Of the 20 studies that did not appropriately report the prediction task, 14 also did not provide appropriate descriptions of the prediction window type, alignment, or length of the observation window. There were also 17 studies with dynamic predictions (i.e., updating predictions continuously) although there may have been more studies with dynamic predictions despite not directly stating this in the text [7,15,24,31,35,38,39,42,47,49,52,60,65,67,68,94,108]. Out of these studies, the majority (n = 12) used right alignment with the remaining using left alignment (n = 1), both (n = 2), or an unclear alignment (n = 2).

5. DISCUSSION

In this review, we assessed studies using temporal EHR data to make patient-level predictions for factors related to problem formulation to determine how the time windows of interest and prediction task were defined and reported. Temporal machine learning models have been used to model a variety of clinical use cases, including acute and chronic conditions. Although selecting different observation and prediction windows can impact the clinical utility and rigor of temporal models, nearly half of the studies reviewed did not clearly report the length or type of time windows used despite most studies sufficiently outlining the prediction task. The discrepancy in reporting of the time windows of interest demonstrates a need for improvements in evaluating model rigor, reproducibility, and applicability in a clinical setting.

5.1. How to formulate the prediction problem?

5.1.1. General Considerations

Prior to selecting the time windows of interest and the prediction task, researchers should first consider the underlying goal of the model. If the model is to be used for research related to model development and/or evaluation, selection of the prediction task and time windows of interest are primarily limited by data availability and computational power. However, if the goal is implementation in a clinical setting such that the model will be interacting with clinicians, there are additional considerations, including how often the prediction will be made, whether to right or left-align the time windows of interest, and how the outcome will be defined in the context of the clinical setting.

Time window lengths are important considerations for model implementation. One consistent attribute across studies that looked at multiple observation and prediction window lengths was that increasing the length of the observation window and decreasing the length of the gap window improved model performance, likely because a larger observation window lends more data for model training while a small gap window reduces the period between the data used for training and the prediction time. For researchers looking to select the best window lengths, optimal model performance should be achieved when using a large observation window and a small gap window. However, researchers must also consider how the gap window length affects the clinical utility of their model. For example, a model may be of low clinical utility if the time between observation and prediction is not long enough for selecting an appropriate intervention or course of action. Additionally, since the relationship between the time-at-risk window and model performance was inconclusive, assessing different time-at-risk window lengths based on the clinical context may be appropriate. To demonstrate these concepts, we will outline how researchers may approach modeling of acute conditions (e.g., sepsis) and chronic conditions (e.g., chronic heart failure).

5.1.2. Example of Prediction Problem Formulation for Acute Conditions

For implementation, it is easiest to formulate the prediction problem by thinking about the outcome and working backwards to the observation window. However, it should be noted that while it is easier to consider

the prediction problem in reverse, models should be able to provide predictions without knowledge of the outcome since the outcome is unknown at the time of prediction. Thus, for prediction of sepsis, the first step is defining sepsis to split the data into cases and controls. However, the definition of sepsis varied between studies and a thorough review of current definitions or expert review are likely needed to ensure the condition is defined appropriately. For example, in one study [7], sepsis was defined as "two or more SIRS [Systemic Inflammatory Response Syndrome] criteria, a blood culture order, and at least one element of end-organ failure" while in another study [30], sepsis was identified using at least two points on the Sequential Organ Failure Assessment (SOFA) along with suspected infection.

After defining the outcome for each patient, the length of the gap window should be determined to ensure that model predictions provide adequate time to apply interventions. For predicting sepsis, the next step is to assess the interventions available and the optimal time to apply these interventions. While early application of interventions such as antibiotics and fluid resuscitation are associated with better patient outcomes [109], there are few guidelines about how early in advance is sufficient, which is likely why prediction window lengths varied between studies in this review. Nevertheless, a critical review of the evidence base and consultation with clinical experts should be performed to determine the length of the gap window. The next decision is whether the prediction should be precise (a narrow time-at-risk window) or general (a broad time-at-risk window). Based on the studies reviewed, there does not appear to be a consensus about which strategy is more effective for implementation although most of the studies predicting onset of sepsis used a narrow time-at-risk window presumably to have more precise timing for interventions. Thus, a narrow time-at-risk window suggests more patient-focused prediction and a broad time-at-risk window suggests prediction for resource allocation.

Based on the time-at-risk window, one can assign the length of the observation window and the prediction task. For the study looking at a broad time-at-risk window [25], the observation window was set between four and 48 hours (with optimal model performance at 48 hours) and the prediction task was defined as prediction of sepsis onset within the remainder of that visit. For studies looking at a narrow time-at-risk window, the observation window was either variable or unknown and predictions were made continuously

such that for each hour, prediction of sepsis during the hour immediately following the gap window was assessed. Both strategies have potential for implementation since neither require knowledge of sepsis onset when the prediction is made.

5.1.3. Example of Prediction Problem Formulation for Chronic Conditions

Here, we consider the prediction problem for modeling of chronic heart failure (HF) onset. Similar to acute conditions, we start with defining the time windows and prediction task from the onset of disease. Since all studies reviewed predicting onset of chronic conditions used visits as the modeling unit rather than observations from a single visit, HF onset was defined as the visit when disease codes for HF were reported [69-72]. To avoid an imbalanced set for model training, several studies sampled controls by matching visit times aligned to HF onset (or index visit) for cases [69-71].

We then define the gap and time-at-risk windows. Among HF studies, the gap window was defined either as a minimum of three months [69,71,72] or as the time between the previous visit and the index visit [70,73] although justification was not provided. Thus, while interventions exist for HF [71], there does not appear to be a consensus about the amount of time required for these interventions to be effective. However, there was a consensus about the length of the time-at-risk window in that all the HF studies were only interested in prediction of HF at the next visit rather than within a specified period of time, likely to provide more targeted therapies. Finally, setting the observation window can be more difficult with chronic conditions than with acute ones because a single visit will have a defined start point at admission but data availability (or the completeness of patient records over many years) is necessary to consider when defining the observation window. For HF onset prediction, there were varying observation window lengths and definitions with some studies using a number of visits [72,73] and others using a period of time, such as two years [71] or 10 months [69], and then aggregating visit information for each time step.

5.2. What are best practices for reporting prediction problems?

Discussion of time windows and prediction tasks varied across studies and clinical contexts. All the studies we reviewed had appropriate reporting of modeling methods based on TRIPOD and MI-CLAIM standards and most

had appropriate reporting of participant selection, data source(s), and data processing methods. Although most studies had appropriate reporting of the prediction task, we found nearly half of the studies did not provide a clear description of the length and/or type of prediction window(s) used for modeling. There were also studies that did not define the window lengths or prediction task in the methods but rather as a part of the results when describing the outcome for each prediction task. Since these inconsistencies make it difficult to assess and reproduce study design, we propose reporting standards consider specific concerns when modeling time-series data and appropriate additions for minimal reporting of time-series modeling to improve rigor and reproducibility.

Regardless of reporting standards, we suggest all studies clearly describe the prediction problem, including how the time windows of interest will be defined, whether the windows overlap, and how predictions will be made, as part of the methods. If possible, we also highly recommend that studies include a visualization that demonstrates the relationship of the time windows of interest (between cases and controls if applicable), at which point the prediction will be made, and at which point the outcome is being assessed. For several studies reviewed, the description provided in the text could have been interpreted differently based on the varying definitions of prediction window and on nuances in the use of words such as "in" and "within." However, figures (when provided) clarified the prediction problem significantly to the point where we considered previously inappropriate reporting of the prediction problem appropriate.

6. CONCLUSION

We reviewed 92 studies applying predictive models on EHR data and assessed the prediction problem, including time window selection and definition of the prediction task. Previous studies have demonstrated significant promise for use of temporal EHR data in predicting clinical outcomes. Here, we found unique challenges in selection of the prediction task and time windows of interest that were complicated by a lack of standardized reporting for temporal EHR applications. Based on these observations, we offered guidelines for time-series problem formulation and reporting that can be expanded upon and integrated into frameworks such as TRIPOD [18] and MI-CLAIM [19]. Given the rapid growth of temporal modeling using EHR data, such standardized reporting would promote rigor and reproducibility in future studies and facilitate the process of implementing algorithms in clinical settings.

7. DECLARATIONS

7.1. Ethical Approval

Not applicable

7.2. Competing Interests

All authors declare that they have no competing interests.

7.3. Authors' Contributions

SP, VS: conceptualized idea; developed search criteria; assessed records for eligibility; edited and approved final manuscript

SP: screened articles; drafted manuscript

7.4. Funding

This work was supported in part by the National Science Foundation under grant #1838745 and the National Institute of General Medical Sciences of the National Institutes of Health under grant GM132008.

7.5. Availability of Data and Materials

Not applicable

8. REFERENCES

1. Shickel B., Tighe P.J., Bihorac A., Rashidi P.: Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE Journal of Biomedical and Health Informatics. 22, 1589-1604 (2018)
2. ONC Data Brief: Use of Certified Health IT and Methods to Enable Interoperability by U.S. Non-Federal Acute Care Hospitals, 2019 (The Office of the National Coordinator for Health Information Technology) 54, (2021).
3. Shah S.M., Khan R.A.: Secondary Use of Electronic Health Record: Opportunities and Challenges. IEEE Access. 8, 136947-136965 (2020)
4. Zhao J., Papapetrou P., Asker L., Boström H.: Learning from heterogeneous temporal data in electronic health records. Journal of Biomedical Informatics. 65, 105-119 (2017)
5. Sherman E., Gurm H., Balis U., Owens S., Wiens J.: Leveraging Clinical Time-Series Data for Prediction: A Cautionary Tale. AMIA Annual Symposium proceedings AMIA Symposium. 2017, 1571-1580 (2018)
6. Harutyunyan H., Khachatrian H., Kale D.C., Ver Steeg G., Galstyan A.: Multitask learning and benchmarking with clinical time series data. Sci Data. 6, 96 (2019)

7. Bedoya A.D., Futoma J., Clement M.E., et al: Machine learning for early detection of sepsis: an internal and temporal validation study. JAMIA Open. 3, 252-260 (2020)
8. Zhao J., Feng Q., Wu P., et al: Learning from Longitudinal Data in Electronic Health Record and Genetic Data to Improve Cardiovascular Event Prediction. Sci Rep. 9, 717 (2019)
9. Singh A., Nadkarni G., Gottesman O., Ellis S.B., Bottinger E.P., Guttag J.V.: Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration. Journal of Biomedical Informatics. 53, 220-228 (2015)
10. Ayala Solares J.R., Diletta Raimondi F.E., Zhu Y., et al: Deep learning for electronic health records: A comparative review of multiple deep neural architectures. Journal of Biomedical Informatics. 101, 103337 (2020)
11. Si Y., Du J., Li Z., et al: Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. Journal of Biomedical Informatics. 115, 103671 (2021)
12. Xiao C., Choi E., Sun J.: Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. Journal of the American Medical Informatics Association. (2018)
13. Xie F., Yuan H., Ning Y., et al: Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies. Journal of Biomedical Informatics. 126, 103980 (2022)
14. OHDSI: 13. pp. OHDSI, (2019).
15. Lauritsen S.M., Kalor M.E., Kongsgaard E.L., et al: Early detection of sepsis utilizing deep learning on electronic health record event sequences. Artif Intell Med. 104, 101820 (2020)
16. Reps J.M., Schuemie M.J., Suchard M.A., Ryan P.B., Rijnbeek P.R.: Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. Journal of the American Medical Informatics Association. 25, 969-975 (2018)
17. Liberati A., Altman D.G., Tetzlaff J., et al: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of Clinical Epidemiology. 62, e1-e34 (2009)
18. Collins Gs Fau - Reitsma J.B., Reitsma Jb Fau - Altman D.G., Altman Dg Fau - Moons K.G.M., Moons K.G.: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. (2015)
19. Norgeot B., Quer G., Beaulieu-Jones B.K., et al: Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nature Medicine. 26, 1320-1324 (2020)
20. Ashfaq A., Sant'Anna A., Lingman M., Nowaczyk S.: Readmission prediction using deep learning on electronic health records. J Biomed Inform. 97, 103256 (2019)
21. Barbieri S., Kemp J., Perez-Concha O., et al: Benchmarking Deep Learning Architectures for Predicting Readmission to the ICU and Describing Patients-at-Risk. Sci Rep. 10, 1111 (2020)

22. Lin Y.W., Zhou Y., Faghri F., Shaw M.J., Campbell R.H.: Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. PLoS One. 14, e0218942 (2019)
23. Reddy B.K., Delen D.: Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology. Comput Biol Med. 101, 199-209 (2018)
24. Chang Y., Rubin J., Boverman G., et al: A Multi-Task Imputation and Classification Neural Architecture for Early Prediction of Sepsis from Multivariate Clinical Time Series. 2019:Page 1-Page 4.
25. Khoshnevisan F., Ivy J., Capan M., Arnold R., Huddleston J., Chi M.: Recent Temporal Pattern Mining for Septic Shock Early Prediction. 2018:229-240.
26. Li Q., Huang L.F., Zhong J., Li L., Li Q., Hu J.: Data-driven Discovery of a Sepsis Patients Severity Prediction in the ICU via Pre-training BiLSTM Networks. 2019:668-673.
27. Lin C., Zhang Y., Ivy J., et al: Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM. 2018:219-228.
28. Nonaka N., Seita J.: Demographic Information Initialized Stacked Gated Recurrent Unit for an Early Prediction of Sepsis. 2019:1-4.
29. Park H.J., Jung D.Y., Ji W., Choi C.M.: Detection of Bacteremia in Surgical In-Patients Using Recurrent Neural Network Based on Time Series Records: Development and Validation Study. J Med Internet Res. 22, e19512 (2020)
30. Persson I., Ostling A., Arlbrandt M., Soderberg J., Becedas D.: A Machine Learning Sepsis Prediction Algorithm for Intended Intensive Care Unit Use (NAVOY Sepsis): Proof-of-Concept Study. JMIR Form Res. 5, e28000 (2021)
31. Rafiei A., Rezaee A., Hajati F., Gheisari S., Golzan M.: SSP: Early prediction of sepsis using fully connected LSTM-CNN model. Comput Biol Med. 128, 104110 (2021)
32. Reyna M.A., Josef C.S., Jeter R., et al: Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. Critical Care Medicine. 48, (2020)
33. Saqib M., Sha Y., Wang M.D.: Early Prediction of Sepsis in EMR Records Using Traditional ML Techniques and Deep Learning LSTM Networks. Annu Int Conf IEEE Eng Med Biol Soc. 2018, 4038-4041 (2018)
34. Van Steenkiste T., Ruyssinck J., De Baets L., et al: Accurate prediction of blood culture outcome in the intensive care unit using long short-term memory neural networks. Artif Intell Med. 97, 38-43 (2019)
35. Vicar T., Novotna P., Hejc J., Ronzhina M., Smisek R.: Sepsis Detection in Sparse Clinical Data Using Long Short-Term Memory Network with Dice Loss. 2019:Page 1-Page 4.
36. Wickramaratne S.D., Mahmud M.S.: Bi-Directional Gated Recurrent Unit Based Ensemble Model for the Early Detection of Sepsis. 2020:70-73.
37. Zhang D., Yin C., Hunold K.M., Jiang X., Caterino J.M., Zhang P.: An interpretable deep-learning model for early prediction of sepsis in the emergency department. Patterns (N Y). 2, 100196 (2021)

38. He Z., Du L., Zhang P., Zhao R., Chen X., Fang Z.: Early Sepsis Prediction Using Ensemble Learning With Deep Features and Artificial Features Extracted From Clinical Electronic Health Records. Crit Care Med. 48, e1337-e1342 (2020)
39. Aczon M.D., Ledbetter D.R., Laksana E., Ho L.V., Wetzel R.C.: Continuous Prediction of Mortality in the PICU: A Recurrent Neural Network Model in a Single-Center Dataset. Pediatr Crit Care Med. 22, 519-529 (2021)
40. Deasy J., Lio P., Ercole A.: Dynamic survival prediction in intensive care units from heterogeneous time series without the need for variable selection or curation. Sci Rep. 10, 22129 (2020)
41. Gandin I., Scagnetto A., Romani S., Barbati G.: Interpretability of time-series deep learning models: A study in cardiovascular patients admitted to Intensive care unit. J Biomed Inform. 121, 103876 (2021)
42. Gupta A., Liu T., Crick C.: Utilizing time series data embedded in electronic health records to develop continuous mortality risk prediction models using hidden Markov models: A sepsis case study. Statistical Methods in Medical Research. 29, 3409-3423 (2020)
43. Harrison E., Chang M., Hao Y., Flower A.: Using machine learning to predict near-term mortality in cirrhosis patients hospitalized at the University of Virginia health system. 2018:112-117.
44. Liu L., Liu Z., Wu H., et al: Multi-task Learning via Adaptation to Similar Tasks for Mortality Prediction of Diverse Rare Diseases. AMIA Annu Symp Proc. 2020, 763-772 (2020)
45. Maheshwari S., Agarwal A., Shukla A., Tiwari R.: A comprehensive evaluation for the prediction of mortality in intensive care units with LSTM networks: patients with cardiovascular disease. Biomed Tech (Berl). 65, 435-446 (2020)
46. Sha Y., Wang M.D.: Interpretable Predictions of Clinical Outcomes with An Attention-based Recurrent Neural Network. presented at: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics; 2017; Boston, Massachusetts, USA.
47. Shickel B., Loftus T.J., Adhikari L., Ozrazgat-Baslanti T., Bihorac A., Rashidi P.: DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning. Sci Rep. 9, 1879 (2019)
48. Tan Q., Ma A.J., Deng H., et al: A Hybrid Residual Network and Long Short-Term Memory Method for Peptic Ulcer Bleeding Mortality Prediction. AMIA Annu Symp Proc. 2018, 998-1007 (2018)
49. Thorsen-Meyer H.C., Nielsen A.B., Nielsen A.P., et al: Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. Lancet Digit Health. 2, e179-e191 (2020)
50. Wang Y., Zhu Y., Lou G., Zhang P., Chen J., Li J.: A maintenance hemodialysis mortality prediction model based on anomaly detection using longitudinal hemodialysis data. J Biomed Inform. 123, 103930 (2021)
51. Yu K., Zhang M., Cui T., Hauskrecht M.: Monitoring ICU Mortality Risk with A Long Short-Term Memory Recurrent Neural Network. Pac Symp Biocomput. 25, 103-114 (2020)

52. Yu R Fau - Zheng Y., Zheng Y Fau - Zhang R., Zhang R Fau - Jiang Y., Jiang Y Fau - Poon C.C.Y., Poon C.C.Y.: Using a Multi-Task Recurrent Neural Network With Attention Mechanisms to Predict Hospital Mortality of Patients. (2020)
53. Choi E., Bahadori M.T., Schuetz A., Stewart W.F., Sun J.: Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR Workshop Conf Proc. 56, 301-318 (2016)
54. Chu J., Dong W., Huang Z.: Endpoint prediction of heart failure using electronic health records. J Biomed Inform. 109, 103518 (2020)
55. Kaji D.A., Zech J.R., Kim J.S., et al: An attention based deep learning model of clinical events in the intensive care unit. PLoS One. 14, e0211057 (2019)
56. Lee J.M., Hauskrecht M.: Recent Context-aware LSTM for Clinical Event Time-series Prediction. Artif Intell Med Conf Artif Intell Med (2005-). 11526, 13-23 (2019)
57. Lee J.M., Hauskrecht M.: Modeling multivariate clinical event time-series with recurrent temporal mechanisms. Artif Intell Med. 112, 102021 (2021)
58. Lei L., Zhou Y., Zhai J., et al: An Effective Patient Representation Learning for Time-series Prediction Tasks Based on EHRs. 2018:885-892.
59. Pham T., Tran T., Phung D., Venkatesh S.: Predicting healthcare trajectories from medical records: A deep learning approach. J Biomed Inform. 69, 218-229 (2017)
60. Rajkomar A., Oren E., Chen K., et al: Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 1, 18 (2018)
61. Rodrigues J.F., Spadon G., Brandoli B., Amer-Yahia S.: Lig-Doctor: Real-World Clinical Prognosis using a Bi-Directional Neural Network. 2020:569-572.
62. Tang F., Xiao C., Wang F., Zhou J.: Predictive modeling in urgent care: a comparative study of machine learning approaches. JAMIA Open. 1, 87-98 (2018)
63. Wang T., Tian Y., Qiu R.G.: Long Short-Term Memory Recurrent Neural Networks for Multiple Diseases Risk Prediction by Leveraging Longitudinal Medical Records. IEEE J Biomed Health Inform. 24, 2337-2346 (2020)
64. Chen Z., Chen M., Sun X., et al: Analysis of the Impact of Medical Features and Risk Prediction of Acute Kidney Injury for Critical Patients Using Temporal Electronic Health Record Data With Attention-Based Neural Network. Front Med (Lausanne). 8, 658665 (2021)
65. Kim K., Yang H., Yi J., et al: Real-Time Clinical Decision Support Based on Recurrent Neural Networks for In-Hospital Acute Kidney Injury: External Validation and Model Interpretation. J Med Internet Res. 23, e24120 (2021)
66. Peng Y.C., Souza N.S.D., Bush B., Brown C., Venkataraman A.: Predicting Acute Kidney Injury via Interpretable Ensemble Learning and Attention Weighted Convoutional-Recurrent Neural Networks. 2021:1-6.
67. Rank N., Pfahringer B., Kempfert J., et al: Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. NPJ Digit Med. 3, 139 (2020)

68. Tomasev N., Glorot X., Rae J.W., et al: A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 572, 116-119 (2019)
69. Maragatham G., Devi S.: LSTM Model for Prediction of Heart Failure in Big Data. J Med Syst. 43, 111 (2019)
70. Rasmy L., Wu Y., Wang N., et al: A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. J Biomed Inform. 84, 11-16 (2018)
71. Chen R., Stewart W.F., Sun J., Ng K., Yan X.: Recurrent Neural Networks for Early Detection of Heart Failure From Longitudinal Electronic Health Record Data: Implications for Temporal Modeling With Respect to Time Before Diagnosis, Data Density, Data Quantity, and Data Type. Circ Cardiovasc Qual Outcomes. 12, e005114 (2019)
72. Duan H., Sun Z., Dong W., He K., Huang Z.: On Clinical Event Prediction in Patient Treatment Trajectory Using Longitudinal Electronic Health Records. IEEE J Biomed Health Inform. 24, 2053-2063 (2020)
73. Jin B., Che C., Liu Z., Zhang S., Yin X., Wei X.: Predicting the Risk of Heart Failure With EHR Sequential Data Modeling. IEEE Access. 6, 9256-9261 (2018)
74. Liang C.W., Yang H.C., Islam M.M., et al: Predicting Hepatocellular Carcinoma With Minimal Features From Electronic Health Records: Development of a Deep Learning Model. JMIR Cancer. 7, e19812 (2021)
75. Wang Y.H., Nguyen P.A., Islam M.M., Li Y.C., Yang H.C.: Development of Deep Learning Algorithm for Detection of Colorectal Cancer in EHR Data. Stud Health Technol Inform. 264, 438-441 (2019)
76. Yang Y., Fasching P.A., Tresp V.: Predictive Modeling of Therapy Decisions in Metastatic Breast Cancer with Recurrent Neural Network Encoder and Multinomial Hierarchical Regression Decoder. 2017:46-55.
77. Yeh M.C., Wang Y.H., Yang H.C., Bai K.J., Wang H.H., Li Y.J.: Artificial Intelligence-Based Prediction of Lung Cancer Risk Using Nonimaging Electronic Medical Records: Deep Learning Approach. J Med Internet Res. 23, e26256 (2021)
78. An Y., Tang K., Wang J.: Time-Aware Multi-Type Data Fusion Representation Learning Framework for Risk Prediction of Cardiovascular Diseases. IEEE/ACM Trans Comput Biol Bioinform. PP, (2021)
79. Guo A., Beheshti R., Khan Y.M., Langabeer J.R., 2nd, Foraker R.E.: Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models. BMC Med Inform Decis Mak. 21, 5 (2021)
80. Park S., Kim Y.J., Kim J.W., Park J.J., Ryu B., Ha J.-W.: Interpretable Prediction of Vascular Diseases from Electronic Health Records via Deep Attention Networks. presented at: 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE); 2018;
81. Ningrum D.N.A., Kung W.M., Tzeng I.S., et al: A Deep Learning Model to Predict Knee Osteoarthritis Based on Nonimage Longitudinal Medical Record. J Multidiscip Healthc. 14, 2477-2485 (2021)
82. Norgeot B., Glicksberg B.S., Trupin L., et al: Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. JAMA Netw Open. 2, e190606 (2019)

83. Fouladvand S., Mielke M.M., Vassilaki M., Sauver J.S., Petersen R.C., Sohn S.: Deep Learning Prediction of Mild Cognitive Impairment using Electronic Health Records. Proceedings (IEEE Int Conf Bioinformatics Biomed). 2019, 799-806 (2019)
84. Ljubic B., Roychoudhury S., Cao X.H., et al: Influence of medical domain knowledge on deep learning for Alzheimer's disease prediction. Comput Methods Programs Biomed. 197, 105765 (2020)
85. Shah P.K., Ginestra J.C., Ungar L.H., et al: A Simulated Prospective Evaluation of a Deep Learning Model for Real-Time Prediction of Clinical Deterioration Among Ward Patients. Crit Care Med. 49, 1312-1321 (2021)
86. AlSaad R., Malluhi Q., Janahi I., Boughorbel S.: Interpreting patient-Specific risk prediction using contextual decomposition of BiLSTMs: application to children with asthma. BMC Med Inform Decis Mak. 19, 214 (2019)
87. Alshwaheen T.I., Hau Y.W., Ass'Ad N., Abualsamen M.M.: A Novel and Reliable Framework of Patient Deterioration Prediction in Intensive Care Unit Based on Long Short-Term Memory-Recurrent Neural Network. IEEE Access. 9, 3894-3918 (2021)
88. Chen D., Jiang J., Fu S., et al: Early Detection of Post-Surgical Complications using Time-series Electronic Health Records. AMIA Jt Summits Transl Sci Proc. 2021, 152-160 (2021)
89. De Brouwer E., Becker T., Moreau Y., et al: Longitudinal machine learning modeling of MS patient trajectories improves predictions of disability progression. Comput Methods Programs Biomed. 208, 106180 (2021)
90. Krishnamurthy S., Ks K., Dovgan E., et al: Machine Learning Prediction Models for Chronic Kidney Disease Using National Health Insurance Claim Data in Taiwan. Healthcare (Basel). 9, (2021)
91. Wu C.L., Wu M.J., Chen L.C., et al: AEP-DLA: Adverse Event Prediction in Hospitalized Adult Patients Using Deep Learning Algorithms. IEEE Access. 9, 55673-55689 (2021)
92. Cobian A., Abbott M., Sood A., et al: Modeling Asthma Exacerbations from Electronic Health Records. AMIA Jt Summits Transl Sci Proc. 2020, 98-107 (2020)
93. Dong X., Deng J., Rashidian S., et al: Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. J Am Med Inform Assoc. 28, 1683-1693 (2021)
94. Jang D.H., Kim J., Jo Y.H., et al: Developing neural network models for early detection of cardiac arrest in emergency department. Am J Emerg Med. 38, 43-49 (2020)
95. Lam C., Tso C.F., Green-Saxena A., et al: Semisupervised Deep Learning Techniques for Predicting Acute Respiratory Distress Syndrome From Time-Series Clinical Data: Model Development and Validation Study. JMIR Form Res. 5, e28028 (2021)
96. Lee J., Ta C., Kim J.H., Liu C., Weng C.: Severity Prediction for COVID-19 Patients via Recurrent Neural Networks. AMIA Jt Summits Transl Sci Proc. 2021, 374-383 (2021)
97. Mohammadi R., Jain S., Agboola S., Palacholla R., Kamarthi S., Wallace B.C.: Learning to Identify Patients at Risk of Uncontrolled Hypertension Using Electronic Health Records Data. AMIA Jt Summits Transl Sci Proc. 2019, 533-542 (2019)

98. Sankaranarayanan S., Balan J., Walsh J.R., et al: COVID-19 Mortality Prediction From Deep Learning in a Large Multistate Electronic Health Record and Laboratory Information System Data Set: Algorithm Development and Validation. J Med Internet Res. 23, e30157 (2021)
99. Shamout F.E., Zhu T., Sharma P., Watkinson P.J., Clifton D.A.: Deep Interpretable Early Warning System for the Detection of Clinical Deterioration. IEEE J Biomed Health Inform. 24, 437-446 (2020)
100. Tao J., Yuan Z., Sun L., Yu K., Zhang Z.: Fetal birthweight prediction with measured data by a temporal machine learning method. BMC Med Inform Decis Mak. 21, 26 (2021)
101. Teoh D.: Towards stroke prediction using electronic health records. BMC Med Inform Decis Mak. 18, 127 (2018)
102. Wu S., Liu S., Sohn S., et al: Modeling asynchronous event sequences with RNNs. J Biomed Inform. 83, 167-177 (2018)
103. Xiang Y., Ji H., Zhou Y., et al: Asthma Exacerbation Prediction and Risk Factor Analysis Based on a Time-Sensitive, Attentive Neural Network: Retrospective Cohort Study. J Med Internet Res. 22, e16981 (2020)
104. Dong X., Deng J., Hou W., et al: Predicting opioid overdose risk of patients with opioid prescriptions using electronic health records based on temporal deep learning. J Biomed Inform. 116, 103725 (2021)
105. Chen W., Wang S., Long G., Yao L., Sheng Q.Z., Li X.: Dynamic Illness Severity Prediction via Multi-task RNNs for Intensive Care Unit. 2018:917-922.
106. Duan H., Sun Z., Dong W., Huang Z.: Utilizing dynamic treatment information for MACE prediction of acute coronary syndrome. BMC Med Inform Decis Mak. 19, 5 (2019)
107. Ge Y., Wang Q., Wang L., et al: Predicting post-stroke pneumonia using deep neural network approaches. Int J Med Inform. 132, 103986 (2019)
108. Liu L., Wu H., Wang Z., Liu Z., Zhang M.: Early Prediction of Sepsis From Clinical Data via Heterogeneous Event Aggregation. 2019:Page 1-Page 4.
109. Rhodes A., Evans L.E., Alhazzani W., et al: Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. Critical Care Medicine. 45, 486-552 (2017)