Journal of Biomedical Informatics 126 (2022) 103980

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Methodological Review

Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies

Feng Xie [a,b,1], Han Yuan [b,1], Yilin Ning [b], Marcus Eng Hock Ong [a,c], Mengling Feng [d], Wynne Hsu [e,f], Bibhas Chakraborty [a,b,g,h], Nan Liu [a,b,f,i,*]

[a] Programme in Health Services and Systems Research, Duke-NUS Medical School, Singapore
[b] Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore
[c] Department of Emergency Medicine, Singapore General Hospital, Singapore
[d] Saw Swee Hock School of Public Health, National University of Singapore, Singapore
[e] School of Computing, National University of Singapore, Singapore
[f] Institute of Data Science, National University of Singapore, Singapore
[g] Department of Statistics and Data Science, National University of Singapore, Singapore
[h] Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, United States
[i] SingHealth AI Health Program, Singapore Health Services, Singapore

ARTICLE INFO

Keywords:
Systematic review
Electronic health records
Temporal data
Representation
Deep learning

ABSTRACT

Objective: Temporal electronic health records (EHRs) contain a wealth of information for secondary uses, such as clinical events prediction and chronic disease management. However, challenges exist for temporal data representation. We therefore sought to identify these challenges and evaluate novel methodologies for addressing them through a systematic examination of deep learning solutions.
Methods: We searched five databases (PubMed, Embase, the Institute of Electrical and Electronics Engineers [IEEE] Xplore Digital Library, the Association for Computing Machinery [ACM] Digital Library, and Web of Science) complemented with hand-searching in several prestigious computer science conference proceedings. We sought articles that reported deep learning methodologies on temporal data representation in structured EHR data from January 1, 2010, to August 30, 2020. We summarized and analyzed the selected articles from three perspectives: nature of time series, methodology, and model implementation.
Results: We included 98 articles related to temporal data representation using deep learning. Four major challenges were identified, including data irregularity, heterogeneity, sparsity, and model opacity. We then studied how deep learning techniques were applied to address these challenges. Finally, we discuss some open challenges arising from deep learning.
Conclusion: Temporal EHR data present several major challenges for clinical prediction modeling and data utilization. To some extent, current deep learning solutions can address these challenges. Future studies may consider designing comprehensive and integrated solutions. Moreover, researchers should incorporate clinical domain knowledge into study designs and enhance model interpretability to facilitate clinical implementation.

1. Introduction

An electronic health record (EHR) [1] collects patients' health information in structured and unstructured digital formats. While the primary objective of an EHR is to improve the efficiency of healthcare systems, it also contains valuable information for secondary uses [2].

Abbreviations: AI, Artificial Intelligence; AKI, Acute Kidney Injury; BERT, Bidirectional Encoder Representations; CNN, Convolutional Neural Network; DAE, Denoising Autoencoder; EHR, Electronic Health Record; GAN, Generative Adversarial Network; GCN, Graph Convolutional Network; GNN, Graph Neural Network; GRU, Gated Recurrent Unit; ICU, Intentive Care Unit; LSTM, Long Short-Term Memory; MIMIC, Medical Information Mart for Intensive Care; MLP, Multi-Layer Perceptron; RBM, Restricted Boltzmann Machine; RETAIN, REverse Time AttentIoN; RNN, Recurrent Neural Network; SAE, Stacked Autoencoder; TANN, Time-Aware Neural Network; VAE, Variational Autoencoder.

* Corresponding author at: Programme in Health Services and Systems Research, Duke-NUS Medical School, 8 College Road, 169857, Singapore.
E-mail address: liu.nan@duke-nus.edu.sg (N. Liu).

[1] These authors contributed equally.

https://doi.org/10.1016/j.jbi.2021.103980
Received 23 July 2021; Received in revised form 7 November 2021; Accepted 20 December 2021
Available online 30 December 2021

EHR contains two types of data: structured data such as diagnoses, procedures, medication prescriptions, vital signs, and lab tests, and unstructured data such as clinical notes, physiological signals, and medical images. Most structured EHR data are documented with timestamps by tracking repeated measurements of a patient's conditions over time. Compared to static data, temporal data provides longitudinal information on a patient's medical history, where hidden patterns (e.g., disease progression or changing variables over time) could be exploited. The growing amount of temporal EHR data presents an opportunity to develop more comprehensive and usable models for risk stratification, disease prognosis, or chronic disease management such as chronic kidney disease prediction [3–4] and adverse drug event detection [5–6].

Although researchers have demonstrated that incorporating temporal EHR data into predictive models can improve discriminative performance [3,7–8], such information is not often fully utilized due to its temporal nature [9]. Most conventional regression and machine learning methods are unable to efficiently extract the temporal pattern from data that contains multiple sets of repeated variables. Some traditional approaches rely on extracting a single value aggregated from the time series, such as mean, median, or other aggregated statistics [10]. It resulted in the loss of potentially valuable sequential information due to the inability to exploit the temporal dynamics of the data [11]. Therefore, better account for the temporality of time series clinical data becomes an important research question.

Temporal EHR data with complex structures and unevenly distributed clinical events present multiple technical challenges, including data irregularity, heterogeneity, sparsity, and model opacity, among others. In view of the limitations of standard learning algorithms in dealing with these challenges, the state-of-the-art deep learning-based methods, such as recurrent neural networks (RNNs) [12–13], long short-term memory (LSTM) [14–16], and gated recurrent unit (GRU) [17], have been proposed for temporal EHR data representation. These sequential deep learning architectures are potentially suitable for dealing with the temporal nature of the EHR. With their ability of learning, flexibility, and generalizability by complex nonlinearity, deep learning algorithms have demonstrated superiority when modeling temporal EHR data in many applications [18–21].

Several recent reviews have summarized the use of deep learning for analyzing general EHR data [22–25]. However, none provides a systematic and in-depth summary of the technical challenges and deep learning solutions for handling temporal EHR data. This review sought to consolidate the recent development of novel deep learning methods for representing temporal data and evaluate selected studies from the perspective of primary challenges and the methodologies that address them. We systematically explored the primary issues involved in analyzing temporal EHR data and thoroughly investigated the state-of-the-art deep learning solutions. Moreover, we identified that there are still open challenges such as usability and transferability, which suggest potential topics for further research.

2. Methods

2.1. Search strategy and data sources

We performed a systematic review of methodological studies on the use of deep learning techniques for temporally structured EHR representations. We conducted the literature search in five databases: PubMed, Embase, the Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library, the Association for Computing Machinery (ACM) Digital Library, and Web of Science. We also searched relevant articles in several prestigious computer science conference proceedings not included in the databases above. The detailed search strategy is presented in the Appendix. We restricted our search to papers published between January 1, 2010, and August 30, 2020. We anticipated that only a few relevant articles would be published before 2010 since deep learning for EHR has been a relatively new development in the last decade.

2.2. Inclusion and exclusion criteria

We followed the PRISMA [26] guidelines to report our systematic review. We included all methodological papers published in English, which employed deep learning for handling temporal EHR data. Review articles, duplicate records, and studies not relevant to EHR or deep learning were excluded. We further excluded pure application papers that did not propose novel methods to address challenges and papers that only dealt with static data (i.e., non-temporal data) or unstructured data (for example, free texts, physiological signals, and medical images). Two reviewers (FX and HY) independently screened all studies and, if ambiguous, discussed with NL to reach a consensus on paper selection.

2.3. Data extraction, synthesis, and analysis

First, we categorized included papers according to the technical challenges that they attempted to address. We identified four main technical challenges in temporal EHR data analysis: data irregularity, data sparsity, data heterogeneity, and model opacity. Second, we evaluated these papers in detail from three aspects: nature of time series, methodology, and model implementation. With regard to the nature of time series, we extracted information, including the time series components and the method of sparse code representation. In terms of methodology, we extracted the method's name, the technical challenges it addressed, and the architecture of deep neural networks. For model implementation, we collected information on the clinical application, EHR datasets used (e.g., Medical Information Mart for Intensive Care [MIMIC] [27]), the main evaluation metrics (e.g., area under the receiver operating characteristic curve [AUROC]), and their main comparators. Finally, we consolidated all extracted information for subsequent analysis and investigation.

3. Results

3.1. Selection process and results overview

Our initial search yielded 1421 papers, of which 495 duplicates were removed, while 926 records went through title and abstract screening. Then, 780 records were excluded as they were not relevant to EHR (n = 246), did not utilize deep learning methods (n = 185), did not involve temporal data representation (n = 61), were only applications using existing methods (n = 23), were review articles (n = 22), or were based on unstructured data (n = 243). As a result, we included 146 articles for full-text review. A total of 98 papers were eventually included, as shown in Table 1. Fig. 1 illustrates the PRISMA diagram on literature selection.

Fig. 2 summarizes the statistics for the included papers. Between 2010 and 2020, the volume of articles has increased significantly. Among the included articles, data irregularity (n = 37) was the most frequently studied challenge, diagnosis (n = 61) was the most commonly used temporal variable, and RNN (n = 72, including LSTM and GRU) was the most widely adopted deep learning architecture. Of the 98 studies, the majority (n = 88) used encounters (e.g., episodes, visits, admissions) as the time step, while others chose fixed time windows (e.g., one hour, day, or month) as the time series.

3.2. Challenges and deep learning solutions

The following section summarizes the four major challenges (i.e., data irregularity, sparsity, heterogeneity, and model opacity) posed by the temporal EHR data and examine their corresponding deep learning solutions.

3.2.1. Data irregularity in temporal EHR

Irregular data is pervasive in temporal EHR [41,123], where the time

Table 1
Characteristics of the included papers in this review.

Year	Paper	Challenge					Deep learning solution		Clinical application	Data source
		Data irregularity	Data sparsity	Data heterogeneity	Model opacity	Others [a]	Method name	Main architecture		
2013	Lasko et al. [28]	✔	✔	✔				Autoencoder	Phenotyping	Vanderbilt University
2015	Esteban et al. [29]		✔					MLP	Clinical events prediction	Charité University Hospital of Berlin
2015	Mehrabi et al. [30]	✔						RBM	Diagnosis association discovery	Rochester Epidemiology Project
2015	Tran et al. [31]	✔	✔	✔			EMR-driven nonnegative restricted Boltzmann machines (eNRBM)	RBM	Suicide risk stratification	Barwon Health
2016	Choi et al. [32]	✔					Doctor AI	RNN	Diagnosis and medication prediction	Sutter Health Palo Alto Medical Foundation
2016	Miotto et al. [33]		✔				Deep Patient	Autoencoder (DAE)	Multiple diseases prediction	Mount Sinai data warehouse
2016	Zhu et al. [34]			✔				CNN, Word2vec	Phenotyping	A real clinical EHRs dataset
2016	Choi et al. [35]				✔		REverse Time AttentIoN (RETAIN)	RNN	Heart failure prediction	Sutter Health
2017	Baytas et al. [14]	✔					Time-Aware LSTM (T-LSTM)	RNN (LSTM), Autoencoder	Parkinson's disease progression prediction	An artificially generated EHR dataset; Parkinson's Progression Markers Initiative (PPMI) Dataset
2017	Che et al. [36]					✔	ehrGAN	CNN	Heart failure and diabetes classification and data generation	A health insurance company
2017	Che et al. [37]				✔		GRU-D	RNN (GRU)	Multiple clinical tasks	Gesture phase segmentation dataset; PhysioNet Challenge 2012 dataset; MIMIC-III
2017	Feng et al. [38]		✔				MG-CNN	CNN	Costs and length of stay prediction	The Hospital Quality Monitoring System (HQMS) database
2017	Mei et al. [39]	✔	✔				Deep Diabetologist	RNN	Personalized hypoglycemia medication prediction	A real clinical database from a city in China
2017	Nguyen et al. [40]	✔	✔				Deep net for medical Record (Deepr)	CNN	Unplanned readmission prediction	A large private hospital chain in Australia
2017	Pham et al. [41]	✔					DeepCare	RNN (LSTM)	Diagnoses prediction and Intervention recommendation	A large regional Australian hospital
2017	Sha et al. [42]				✔		GRU-based RNN with hierarchical attention (GRNN-HA)	RNN (GRU)	Mortality prediction	MIMIC-III
2017	Stojanovic et al. [43]		✔				Disease+procedure2vec (dp2v)	Skip-gram	Healthcare quality prediction	The State Inpatient Database (SID)
2017	Suo et al. [44]			✔	✔			RNN (GRU)	Diagnosis prediction	Study of Osteoporotic Fractures Dataset; BloodTest dataset from University Hospital of Catanzaro, Italy
2017	Suo et al. [45]			✔				CNN	Multiple disease prediction	A real clinical database
2017	Zheng et al. [46]	✔						RNN (GRU)	Severity scores prediction	Alzheimer's Disease Neuroimaging

(continued on next page)

Table 1 (*continued*)

Year	Paper	Challenge					Deep learning solution		Clinical application	Data source
		Data irregularity	Data sparsity	Data heterogeneity	Model opacity	Others [a]	Method name	Main architecture		
										Initiative (ADNI) dataset; National University Hospita (NUH)
2017	Yang et al. [20]	✔		✔				RNN (LSTM, GRU)	Therapy decisions prediction	A real clinical database
2017	Yang et al. [47]				✔		Predictive Task Guided Tensor Decomposition (TaGiTeD)	Tensor Optimization	Hospitalization and medical expense prediction	Two real clinical databases
2018	Bai et al. [48]	✔			✔		Timeline	RNN	Disease progression prediction	The Surveillance, Epidemiology, and End Results (SEER) Medicare Linked Database
2018	Cheung et al. [49]			✔	✔		Attention-based cross-modal convolutional neural network (AXCNN)	CNN	Readmission prediction	A large hospital system in Arizona
2018	Le et al. [50]	✔					Dual memory neural computer (DMNC)	RNN (LSTM)	Disease progression and drug prescription prediction	MIMIC-III
2018	Lee et al. [51]				✔		Medical context attention-based RNN (MCA-RNN)	RNN	Diagnosis prediction	National Patient Sample (NPS) dataset
2018	Lei et al. [52]			✔			RNN-DAE	RNN, Autoencoder (DAE)	Mortality prediction	Shanghai Shuguang Hospital
2018	Lin et al. [53]	✔						CNN, RNN (LSTM)	Sepsis prediction	Christiana Care Health System
2018	Ma et al. [54]				✔		Knowledge-based attention model (KAME)	RNN (GRU)	Diagnosis prediction	Medicaid dataset; the Diabetes dataset; MIMIC-III
2018	Nguyen et al. [55]	✔					Recurrent Sequence of Sets (Resset)	RNN	Diabetes and mental health prediction	A large regional Australian hospital
2018	Park et al. [56]				✔		Frequency-Aware Attention based LSTM (FA-Attn-LSTM)	RNN (LSTM)	Cardiovascular disease risk prediction	Asan Medical Center dataset (Seoul)
2018	Park et al. [57]				✔		EHR History-based prediction using Attention Network (EHAN)	RNN	Medical code prediction	Seoul National University Bundang Hospital (SNUBH)
2018	Rajkomar et al. [58]			✔				RNN (LSTM), TANN, MLP	Mortality, readmission, length of stay, diagnoses prediction	University of California, San Francisco; University of Chicago Medicine
2018	Suo et al. [59]			✔				CNN	Phenotyping	A real clinical database
2018	Suresh et al. [60]			✔				RNN (LSTM), Autoencoder	Mortality prediction	MIMIC-III
2018	Wu et al. [61]			✔				RNN (LSTM)	Asthma phenotyping	Mayo Clinic
2018	Xiao et al. [62]				✔		TopicRNN	RNN (GRU)	Readmissions prediction	A Congestive Heart Failure (CHF) cohort
2018	Yang et al. [63]	✔					Time-aware subGroup Basis Approach with Forecasted events (TGBA-F)	RNN (LSTM)	Septic shock prediction	Christiana Care Health System
2018	Zhang et al. [64]				✔		Patient2Vec	RNN (GRU)	Hospitalizations prediction	University of Virginia Health System
2018	Huang et al. [65]			✔			Stacked denoising auto-encoder (SDAE)	Autoencoder (DAE)	Acute coronary syndrome risk prediction	China People's Liberation Army (PLA) General Hospital
2018	Choi et al. [66]		✔				Multilevel Medical Embedding (MiME)	RNN (GRU)	Multiple disease prediction	Sutter Health
2019	An et al. [67]				✔		DeepRisk	RNN (LSTM)		

(*continued on next page*)

Table 1 (*continued*)

Year	Paper	Challenge					Deep learning solution		Clinical application	Data source
		Data irregularity	Data sparsity	Data heterogeneity	Model opacity	Others [a]	Method name	Main architecture		
									Cardiovascular diseases prediction	Central South University in China
2019	Ashfaq et al. [68]			✔				RNN (LSTM)	Readmission prediction	Southern Sweden
2019	Fiorini et al. [69]	✔					Tangle	RNN (LSTM)	Diabetes therapy initiation prediction	Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) electronic databases of Australia
2019	Guo et al. [70]				✔		CrossOver Attention Model (COAM)	RNN	Multiple disease prediction	A real clinical database in China; MIMIC-III
2019	Jun et al. [71]	✔						Autoencoder (VAE)	Mortality prediction	MIMIC-III
2019	Kwon et al. [72]				✔		RetainVis	RNN	Risk prediction model visualization	Health Insurance Review and Assessment Service (HIRA)
2019	Lee et al. [73]	✔					Recent context-aware LSTM	RNN (LSTM)	Clinical events prediction	MIMIC-III
2019	Li et al. [74]	✔					Variable sensitive GRU (VS-GRU)	RNN (GRU)	Mortality and disease prediction	MIMIC-III; PhysioNet
2019	Lin et al. [75]	✔						RNN (LSTM)	Unplanned ICU readmission prediction	MIMIC-III
2019	Liu et al. [76]	✔						RNN (GRU)	Mortality and ICU admission prediction	MIMIC-III
2019	Liu et al. [77]	✔		✔				RNN (LSTM)	Sepsis prediction	The PhysioNet Computing in Cardiology Challenge 2019
2019	Macias et al. [78]	✔						RNN (LSTM)	Sepsis prediction	The PhysioNet Computing in Cardiology Challenge 2019
2019	Peng et al. [79]		✔				Temporal Self-Attention Network (TeSAN)	RNN (GRU)	Mortality prediction	MIMIC-III; Centers for Medicare & Medicaid Services (CMS)
2019	Ruan et al. [80]	✔					RNN-DAE	RNN (GRU)	Mortality, comorbidity prediction and phenotyping	Shanghai Shuguang Hospital in China
2019	Wang et al. [81]	✔			✔		MCPL-based FT-LSTM	RNN (LSTM)	Clinical events prediction	MIMIC-III
2019	Wang et al. [82]		✔				CompNet	CNN, GCN	Medication prediction	MIMIC-III
2019	Wang et al. [83]		✔				Patient2vec	RNN	Diagnosis prediction	MIMIC-III
2019	Wang et al. [84]	✔					Multilevel Representation Model (MRM)	RNN (LSTM)	Mortality and potassium ion concentration abnormality prediction	MIMIC-III
2019	Xiang et al. [85]		✔					RNN (LSTM)	Concept similarity analysis and disease prediction	Cerner Health Facts database
2019	Xu et al. [86]				✔			RNN	Adverse cardiovascular events prediction	A real clinical database
2019	Yang et al. [87]				✔		Grouped Correlational Generative Adversarial Networks (GcGAN)	GAN	Data generation evaluated by treatment recommendation	Pediatric department of a hospital
2019	Zhang et al. [88]	✔							Missing data Imputation	UC Irvine Machine Learning

(*continued on next page*)

Table 1 (*continued*)

Year	Paper	Challenge					Deep learning solution		Clinical application	Data source
		Data irregularity	Data sparsity	Data heterogeneity	Model opacity	Others [a]	Method name	Main architecture		
								RNN (LSTM), Autoencoder (DAE)		Repository; a clinical vital signs dataset
2019	Zhang et al. [89]				✔		Interpretable clinical knowledge guided risk prediction model (KNOWRISK)	RNN (LSTM)	Heart failure prediction	MIMIC-III
2019	Zhang et al. [90]	✔						RNN (LSTM)	Septic shock prediction	Christiana Care Health System
2019	Zhang et al. [91]					✔	MetaPred	CNN, RNN (LSTM)	Mild Cognitive Impairment (MCI), Alzheimer, Parkinson's disease prediction	Research data warehouse from Oregon Health & Science University Hospital
2020	Afshar et al. [92]			✔			Temporal and static tensor factorization (TASTE)	Tensor Factorization	Heart failure phenotyping	Sutter Palo Alto Medical Foundation; CMS
2020	An et al. [93]			✔			Relation augmented hierarchical multi-task learning framework (RAHM)	RNN (LSTM)	Medication stocking prediction	MIMIC-III
2020	Barbieri et al. [94]	✔						RNN	ICU readmission prediction	MIMIC-III
2020	Chu et al. [95]			✔			Deep adversarial learning model and a multi-task learning model (DAL-EP and MTL-EP)	RNN	Heart failure endpoint prediction	Chinese PLA General Hospital
2020	Duan et al. [96]	✔						RNN	Clinical events prediction	Chinese PLA General Hospital
2020	Gao et al. [97]			✔			CrOss-Modal PseudO-SiamEse network (COMPOSE)	CNN	Patient-trial matching	ClinicalTrials.gov
2020	Gao et al. [98]			✔			StageNet	RNN (LSTM)	Mortality prediction	MIMIC-III; End-Stage Renal Disease (ESRD) dataset
2020	Jin et al. [99]				✔		CarePre	RNN	Diagnosis prediction	MIMIC-III
2020	Jun et al. [100]	✔						RNN	Mortality prediction	MIMIC-III; PhysioNet
2020	Landi et al. [101]			✔			ConvAE	CNN, Autoencoder	Disease prediction	Mount Sinai Health System
2020	Lauritsen et al. [102]	✔						CNN, RNN (LSTM)	Sepsis prediction	Multiple Danish hospitals
2020	Li et al. [103]		✔				Graph Neural Network-Based Diagnosis Prediction (GNDP)	CNN	Diagnosis prediction	MIMIC-III; a real clinical database
2020	Li et al. [104]	✔			✔		BERT for EHR (BEHRT)	Transformer	Disease prediction	Clinical Practice Research Datalink (CPRD)
2020	Li et al. [105]		✔				Cross-field categorical attributes embedding (CCAE)	RNN	Clinical endpoint prediction	SEER Research
2020	Liu et al. [106]	✔	✔				Medi-Care AI	RNN	Medication prediction	MIMIC-III; PhysioNet
2020	Liu et al. [107]			✔			Hybrid method of RNN and GNN (RGNN)	RNN (LSTM), GNN	Prescription prediction	MIMIC-III
2020	Luo et al. [108]	✔			✔		Hierarchical Time-Aware Attention Networks (HiTANet)	Transformer	Disease prediction	A real clinical database
2020	Panigutti et al. [109]				✔		DoctorXAI	RNN	Next visit prediction	MIMIC-III
2020	Qiao et al. [110]			✔			Multi-modal Clinical Data based Hierarchical Multi-label model (MHM)	RNN (GRU)	Diagnosis prediction	MIMIC-II; MIMIC-III
2020	Rongali et al. [111]			✔			CLOUT	RNN (LSTM)	Mortality prediction	MIMIC-III
2020	Song et al. [112]		✔				Local-Global Memory Neural Network (LGMNN)	RNN (LSTM)	Medication prediction	MIMIC-III; The Second Affiliated Hospital of Zhejiang

(*continued on next page*)

Table 1 (*continued*)

Year	Paper	Challenge					Deep learning solution		Clinical application	Data source
		Data irregularity	Data sparsity	Data heterogeneity	Model opacity	Others [a]	Method name	Main architecture		
										University Medical College
2020	Su et al. [113]	✔					Graph-Attention Augmented Temporal Neural Network (GATE)	RNN (GRU)	Medication prediction	MIMIC-III
2020	Wang et al. [114]				✔		Feature rearrangement based convolutional layer (FReaConv)	CNN	Heart failure and mortality prediction	Shanghai Shuguang Hospital
2020	Xiang et al. [115]				✔		Time-Sensitive, Attentive Neural Network (TSANN)	RNN (LSTM)	Asthma exacerbation prediction	Cerner Health Facts database
2020	Yin et al. [116]			✔			Time-Aware Multi-modal auto-Encoder (TAME)	RNN (LSTM), Autoencoder	Sepsis phenotyping	IEEE ICHI Data Analytics Challenge on Missing data Imputation (DACMI); MIMIC-III
2020	Yu et al. [117]				✔			RNN (LSTM)	Mortality prediction	MIMIC
2020	Yu et al. [118]				✔			RNN (LSTM, GRU)	Mortality prediction	MIMIC-III
2020	Zeng et al. [119]	✔		✔	✔		Multilevel Self-Attention Model (MSAM)	Autoencoder (SAE)	Disease and medical cost prediction	MIMIC-III; PFK
2020	Zhang et al. [120]		✔				Hierarchical Attention Propagation (HAP)	RNN	Procedure and diagnosis prediction	ACTFAST; MIMIC-III
2020	Zheng et al. [121]				✔		TRACER	RNN	AKI and mortality prediction	National University Hospital in Singapore; MIMIC-III
2020	Park et al. [18]				✔			RNN (GRU)	Bacteremia prediction	Asan Medical Center
2020	Thorsen-Meyer et al. [122]				✔			RNN (LSTM)	Mortality prediction	Four ICUs in the Capital Region of Denmark

[a] Others includes two papers that solved the problem of model development using limited data. Specific details are provided in the discussion.

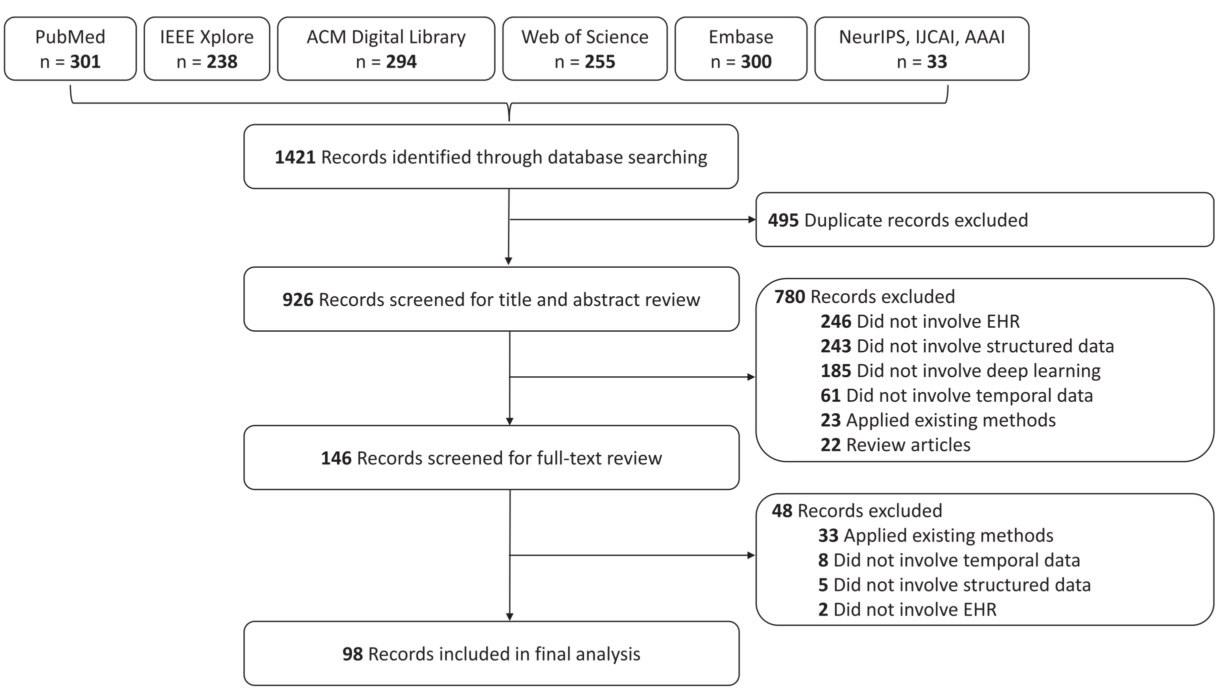

Fig. 1. Literature selection flow of deep learning models in temporal EHR data.

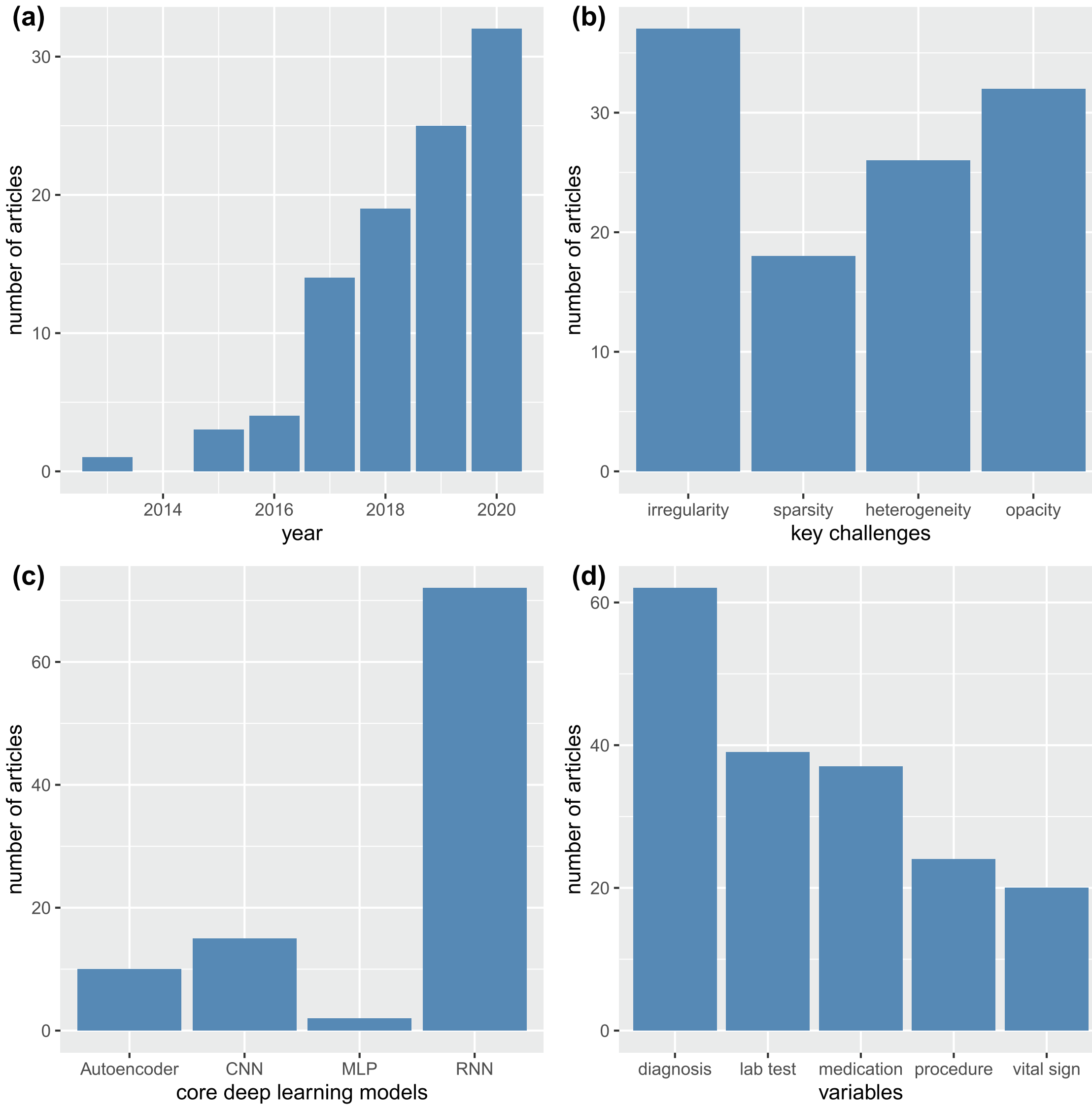

Fig. 2. Summary statistics for all included papers: (a) histogram of yearly publications from 2010 to 2020; (b) histogram by key challenges; (c) histogram by core deep learning models; and (d) histogram by included variables.

intervals between various encounters vary, resulting in challenges of modeling the whole time series. These irregular time intervals may contain valuable hidden information. For example, shorter time intervals may imply more frequent examinations, indicating a patient's worsening condition. To utilize the latent information, researchers usually extract a series of time intervals [3,48], represented as follows for the i-th patient.

$$\boldsymbol{\varphi}_{i,(a-1,a)} = |t_a - t_{a-1}|, a = 1, \cdots, T \quad (1)$$

Deep learning could naturally capture this long-term sequential effect, and two groups of deep learning solutions have been proposed. One group of methods directly model time series, taking irregular time intervals as the input variables, where customized neural network architectures ingeniously fuse each irregular time point [33,40,81,85]. Although the time lapse between successive elements in patients' records may vary from days to years, novel deep learning approaches can fit the unequally distributed data with their inherent temporal structure (e.g., gate architecture of the LSTM) based on the time distribution and its interval $\boldsymbol{\varphi}_{i,(a-1,a)}$ directly, such as T-LSTM [81]. Moreover, a variety of integrated data processing systems have been proposed based on diverse deep learning structures. For example, Deep Patient [33] utilized a three-layer stack of denoising autoencoders to capture hierarchical regularities and dependency in the temporal coding data. REverse Time AttentIoN (RETAIN) [35] was developed based on a two-level neural attention model to detect influential past visits. Deepr [40] ingeniously transformed a record into a sequence of discrete elements separated by coded time gaps and hospital transfers. Xiang et al. [85] applied dynamic input windows to acquire time-sensitive coding information.

Another group of approaches attempts to transform irregular data into regular ones by determining a fixed interval and then treating the time points without data as missing [124]. While the irregular data could naturally be transformed into regular data with the same time intervals, this strategy may result in many missing values [63,74,125] since there are many time intervals without measurements, necessitating imputation. In this situation, a masking vector $u \in \{0, 1\}$ [12] is usually used to represent its missing status. The missing values would reduce statistical power and cause bias in estimating mass parameters in deep learning methods [126].

Most researchers used carry-forward imputation to address the missing value issue in temporal EHR [127], where the last observed values were used for all subsequent missing observation points. However, this solution is likely to introduce bias, assuming that the value remains unchanged from the last observation. Traditional imputation methods, such as median imputation and multiple imputation, may not effectively capture hidden patterns in temporal data, calling for deep learning approaches with temporal representations. Recently, Macias et al. [78] proposed a novel imputation method in an intensive care unit (ICU) setting by exploiting temporal dependencies through autoencoder-represented information. Zhang et al. [88] imputed missing values of multivariate time series by a denoising autoencoder. Based on GRU, Che and colleagues [37] designed GRU-D to utilize informative missingness with prior-based regularization. Furthermore, Jun and colleagues [71] developed a general framework to incorporate effective missing data imputation with a variational autoencoder.

3.2.2. Data sparsity in temporal EHR

The term “data sparsity” indicates that the data contains a large number of zero entries, a problem that is relatively common in temporal EHRs. There are two main reasons why a temporal EHR is susceptible to the data sparsity issue. First, the data is mainly missing due to the lack of visits on the time scale, as healthy patients generally visit the hospital less frequently, leaving very little information recorded in the EHR system [128]. In this case, the expected values or previous values could be imputed if there is no actual measurement at the time point. This problem may be addressed by the methods presented in Section 3.2.1. Second, the standard one-hot encoding [129] would result in many zero values due to the wide range of medical ontologies. This section focuses on data sparsity from the second reason, which is the most prevalent reason and contributes the most to the issue.

Medical ontologies are structured terminologies that link related concepts, such as diagnoses, medications, and treatments. They are the most commonly investigated temporal variables (used in 78 of the 98 papers included). Most medical concepts are mapped to a corresponding coding system, such as the International Classification of Diseases (ICD) [130], Current Procedural Terminology (CPT) [131], and Medication Reference Terminology (MED-RT) [132]. One-hot encoding [129] is the most straightforward technique, converting these categorical variables into several binary columns, where 'one' indicates the presence of the medical code. However, it is not an ideal method for encoding high-dimensional categorical variables. On the other hand, conventional medical ontologies such as SNOMED [133], Charlson Comorbidity Index [134], RxNorm [135], and Logical Observation Identifiers Names and Codes (LOINC) [136] employ structured hierarchical approaches to convert medical concepts into lower-dimensional representations but are unable to extract abundant semantic relations inherent in the temporal data [38,79,137].

Therefore, embedding has become a mainstream method [34] for learning latent representations from high-dimensional sparse medical codes. Embedding originated from Word2Vec [138], an unsupervised feature extraction method for natural language processing (NLP), which converts words into numerical embeddings by mapping each word token into a high-dimensional vector space. Choi et al. [32,139] have recently proposed to learn distributed representations of sparse medical codes (e.g., diagnoses, medications, and procedure codes) using Word2Vec and applied them to several clinical prediction tasks. Subsequently, Med2-Vec [140] was proposed to extend the original Word2Vec with a multilayer perceptron to learn succinct codes and visit-level representations.

Med2Vec was further extended to integrate embedding systems with different deep learning architectures for sparse temporal EHR data. Lu and colleagues [141] proposed utilizing hyperbolic embeddings of medical concepts instead of traditional Euclidean space geometry. Moreover, the cross-field categorical attributes embedding (CCAE) [105] was developed to learn a vectorized representation for cancer patients at attribute-level by orders, where strong semantic coupling among categorical variables was exploited effectively. Esteban et al. [29] employed the Markov model to learn personalized Markov embeddings. MC2Vec [137] was then designed to capture the proximity relationships between medical concepts through a two-step optimization framework that recursively refines the embedding for superior output. Patient2vec [83] introduced the RNN model to learn sequential context-aware features of visits and the correlations between physical symptoms and associated treatments. Zhang et al. [120] developed hierarchical attention propagation (HAP), a hierarchical propagating attention across the entire ontology structure, where a medical code adaptively learns its embedding from all other codes in the hierarchy instead of only its ancestors.

3.2.3. Data heterogeneity in temporal EHR

The heterogeneity of data is another challenge that undermines the quality of analysis. Specifically, a comprehensive EHR dataset may include records from patients with diverse conditions and disparate outcomes. Heterogeneous patients are typically referred to as sub-cohorts (sub-phenotypes) within a population. Patients in the same sub-cohort would be more closely related in terms of their medical circumstances than patients in other sub-cohorts [142]. A successful division of patient sub-cohorts may improve the accuracy of downstream analyses such as cohort analysis, case-based reasoning, treatment comparison, and personalized medicine [140]. Another form of data heterogeneity is the variety of clinical outcomes, including different disease stages or conditions, as well as their complex interactions [143]. Although several conventional methods have been proposed to handle heterogeneous clinical data, deep learning has shown superiority over them.

Patient sub-cohorts are identified by phenotyping techniques that require the patients in a specific cohort to satisfy complex criteria [144]. Conventionally, patient sub-cohorts are determined based on patient similarity through statistical distances like Euclidean distance [59] or machine learning methods such as k-means [34]. However, when dealing with high-dimensional and multimodal longitudinal data, these traditional methods cannot identify complicated patient phenotypes and cannot retain most long-term temporal information [34]. In contrast, a deep learning structure such as RNN could capture these complex temporal dynamics in the longitudinal EHR for evaluating patient similarity. The typical deep neural network architectures include CNN, RNN (LSTM, GRU), and Autoencoder (See Table 1). For instance, Wu et al. [61] utilized the RNN model to improve asthma phenotyping using ICU time sequence data. The proper phenotyping of patients has demonstrated its value in medical decision making [61].

Diverse clinical outcomes, such as different disease conditions and stages and complex interrelationships among the outcomes, make it challenging to model the projection from input variables to outputs, which renders the traditional single-task approach (e.g., binary logistic regression) ineffective. While some statistical methods, such as multinomial logistic regression, can make a multi-label prediction, deep learning promises more. The shallow layers in deep neural networks can be used for more than one task, aiming to learn feature representations for multiple prediction tasks simultaneously. As reported in [44,86,118], several multitask frameworks were proposed based on RNN, which shared the same shallow networks but with a task-specific layer to monitor a specific disease or outcome. Shared layers (unified feature representation across multiple tasks) could save computing resources and provide a model that performs better than single-task-oriented models [143]. Considering heterogeneous patient profile and diverse outcomes allows researchers to model more sophisticated correlations between input variables and output labels. Suresh et al. [60] applied a two-step procedure in their projects: the first step being unsupervised clustering through a sequence-to-sequence autoencoder, and the second step being the multitask prediction.

3.2.4. Opacity in modeling temporal EHR

While deep learning provides diversified solutions to deal with temporal EHR data, its black box nature presents another significant challenge. Due to the depth of neural network layers and the complexity of each module, understanding sophisticated deep learning models remains elusive, particularly when dealing with temporal data. Many researchers have attempted to explain deep learning models with post hoc explanations (e.g., Doctor XAI) [109,145], while others advocate that the models themselves should be interpretable [146–147].

Two groups of approaches have been proposed to interpret black box deep learning models: mimic learning and attention mechanisms. Mimic learning simulates deep learning models through an inherently transparent model such as a simple decision tree [148]. Gradient boosting trees (GBT) [37] was also used to imitate the process of GRU and achieved superior performance and good interpretability when extracting the importance of features.

The concept of attention mechanism [149], derived initially from NLP and used in machine translation to adjust weights of different words, has been popular in research on temporal EHR [150]. Researchers have used attention to determine which time points in the patients' medical history are more predictive of outcomes [44,57,64,86,118,121,151]. Attention can also provide insight into the importance of different visits or variables for aiding medical decision making, where larger attention represents greater importance.

There are two major attention approaches, one of which treats all temporal variables at the same level, while the other takes data hierarchy into account. As shown in [49,51,152–153], neural networks with the attention layer were proposed to calculate the weights without a data hierarchy. The crossover attention model (COAM) [70] was designed by leveraging the correlation between diagnosis and treatment information through the crossover attention mechanism. Park et al. [56] further improved the attention mechanism by adding feature occurrence frequency to capture critical temporal variables that appeared infrequently. Bai et al. [48] enhanced the attention by learning time decay factors, making it possible to interpret the chronic disease progression and understand how the risks of future visits change over time.

Furthermore, some models incorporated heterogeneous data such as treatments, medications, procedures, and diagnoses. Specifically, DeepRisk [67] integrated multiple time-ordered clinical data as a whole by handling correlation among predictors via a single deep neural network and three attention-based LSTMs. KAME [54] was further developed as a knowledge-based attention mechanism, which used medical domain knowledge, and computed the attention based on a directed acyclic graph of various medical concepts.

Another major attention approach exploited the hierarchical structure of temporal data, such as the data from both admission and medical event levels. During admission, many events occur, which are usually recorded by medical codes like ICD. This hierarchical structure with multiple attention levels was intended to integrate local and global time information and enhance model performance and interpretation. Several studies [35,42,108,115,119] have applied a two-level neural attention model within RNN, where the first level pertains to medical events while the second level pertains to visits. A good example is RETAIN [35], which integrates two levels of attention to make use of time information in feature aggregation and explain the critical medical event in the input sequence. To improve the interpretability of RETAIN, Kwon et al. [72] developed RetainVis, an interactive visualization tool. Another attempt was to simultaneously apply graph-level attention (based on knowledge graph) with other attentions [89]. Overall, the hierarchical attention could digest the sequence information correctly in temporal EHR and provide an insightful interpretation of the importance of each variable or timepoint.

4. Discussion

This review summarized the challenges related to temporal EHR data and discussed how deep learning solutions could help to overcome them. While temporal EHR data is valuable for biomedical informatics research, its complex structure poses challenges to standard learning algorithms. Deep learning models have shown the ability to present temporal data in a novel manner while retaining sequential information efficiently. We identified four major challenges through a systematic literature review, including data irregularity, sparsity, heterogeneity, and model opacity. During the last decade, numerous deep learning-related techniques have been proposed, and this number continues to grow rapidly over time, demonstrating the importance and potential of deep learning in temporal EHR data analysis. While these deep learning techniques have shown promising results, several challenges remain, including the need for high-quality data and the issue regarding their applicability to clinical practice. Ideally, future studies could consider designing a comprehensive system that combines solutions to all challenges.

Despite various attempts to address data irregularity, heterogeneity and sparsity, there is still a great need to improve the data itself. For deep learning algorithms to be successful, large-scale EHR datasets are always required. The most commonly used dataset in our included papers was MIMIC (n = 37/98) [27], a well-organized and freely accessible critical care database developed at the Beth Israel Deaconess Medical Center. The majority of studies analyzed only one dataset, and only ten utilized two or more datasets, raising questions about the transferability and generalizability of the models. Therefore, we recommend the development of more large-scale EHR databases that are freely accessible worldwide, providing the opportunity for multicenter validation of current models. Aside from data size, the quality of a dataset is another critical factor affecting model performance; improvements in data collection and processing, such as the correction of outliers due to mistyping or misalignment, may be considered.

Although the availability of large, labelled data is always desirable, situations with limited data are common in medical settings because of the costs of labelling and the sensitive nature of data sharing. Many approaches have been proposed to resolve this issue, including data augmentation and optimal use of data. Generative Adversarial Networks (GANs) provide a compelling solution to amplify temporal data, and Che et al. [36] demonstrated that the newly generated data is of adequate quality. On the other hand, Zhang et al. [91] made use of limited data through the novel representation framework MetaPred. Another potential solution would be transfer learning, which allows us to transfer knowledge between multiple hospitals or EHRs, and combine various sources to extract knowledge, referred to as multi-source transfer learning [154]. General transfer learning consists of a two-stage paradigm [155], where the leading deep learning network is generally trained on a large-scale, publicly available benchmark dataset. Next, the pre-trained network is further conditioned on the specific local data with limited samples. Transfer learning has the potential to relieve the data shortage in healthcare and improve model generalizability [156].

There has been considerable discussion of the opacity issue for temporal deep learning models, especially for medical applications with high-stakes decisions. Several attention mechanisms have been proposed to address these black box models [35,42,108,115,119]. Grad-CAM [157] is a widely adopted algorithm developed initially to provide visual explanations for CNN by highlighting the important regions, and was recently extended to the medical field [158]. However, these post-hoc interpretability approaches may lead to explanations from certain artifacts learned by the model rather than actual knowledge derived from the data [159]. This limitation raises concerns about model usability in real healthcare settings. As a comparison, ante-hoc interpretable models are preferred by doctors and nurses in clinical practice because they can understand them naturally and inherently [146,160–161]. Ustun and Rudin recently developed the Risk-calibrated Supersparse Linear Integer Model (RiskSLIM) [162] and further improved it through the optimization of risk scores [163]. Besides that, Xie et al. provided practical solutions, AutoScore [10] and its extensions [164–165], leveraging interpretable machine learning for clinical score [166] generation. These intrinsically interpretable methods have considerable potential for integrating deep learning techniques to facilitate model validation in real-world settings.

In this review, RNN-based architectures (n = 72, including LSTM and

GRU) were most commonly used, since RNNs are well suited to handle sequence data with short-term or long-term temporal correlation, in accordance with a previous review [167]. While CNN (n = 15) can also handle temporal input, it is more effective in capturing spatial correlation in images [167]. In contrast, RNN-based models are designed to accept a variety of time-varying inputs, making them the preferred methods for modeling temporal EHRs. Hybrid models that combine RNN and CNN have also been designed and implemented [53,91,102]. Transformer [168] and MLP-Mixer [169] have recently emerged as popular alternative frameworks. Transformer computes temporal representations entirely based on self-attention without the use of sequence-based RNNs or convolution. It has shown great potential in temporal EHR representation [104,108]. Furthermore, MLP-Mixer [169] was later developed with a more straightforward structure, requiring no convolutions. With one MLP for per-location features and another for spatial information, MLP-Mixer appears to be a conceptually and technically succinct alternative for processing temporal EHR data.

This study has several limitations. First, we sought to understand the current state of the literature from a methodological perspective. We did not attempt to summarize all clinical applications, report specific models and training details for each paper or compare the performance of deep learning solutions. Second, considering the heterogeneity of data preprocessing, parameter tuning approaches, and clinical tasks among the included studies, we could not recommend the overall best deep learning methods for temporal EHR data analysis. Third, this review focused exclusively on deep learning methods for analyzing structured temporal EHR data. It will be beneficial in the future to investigate techniques that deal with both structured and unstructured data (e.g., clinical notes, medical images, and physiological signals). Lastly, the exclusion of preprints in our analysis may have overlooked some new evidence but was able to ensure the inclusion of only peer-reviewed scientific results.

5. Conclusion

We comprehensively reviewed the primary issues in analyzing temporal EHR data and presented state-of-the-art deep learning solutions. Various significant challenges arising from the temporal representation of EHR data were addressed to some extent by current solutions. Future research may focus on model transferability, incorporating clinical domain knowledge into study design, and enhancing model interpretability to facilitate clinical implementation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2021.103980.

References

[1] K. Hayrinen, K. Saranto, P. Nykanen, Definition, structure, content, use and impacts of electronic health records: A review of the research literature, Int. J. Med. Inf. 77 (5) (2008) 291–304.

[2] Critical Data M. *Secondary analysis of electronic health records.* Springer Nature; 2016.

[3] A. Singh, G. Nadkarni, O. Gottesman, S.B. Ellis, E.P. Bottinger, J.V. Guttag, Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration, J. Biomed. Inform. 53 (2015) 220–228.

[4] X. Ma, T. Imai, E. Shinohara, S. Kasai, K. Kato, R. Kagawa, K. Ohe, EHR2CCAS: A framework for mapping EHR to disease knowledge presenting causal chain of disorders–chronic kidney disease example, J. Biomed. Inform. 115 (2021) 103692, https://doi.org/10.1016/j.jbi.2021.103692.

[5] J. Zhao, A. Henriksson, M. Kvist, L. Asker, H. Bostrom, Handling Temporality of Clinical Events for Drug Safety Surveillance, AMIA Annu Symp Proc. 2015 (2015) 1371–1380.

[6] Zhao YX, Yuan H, Wu Y. Prediction of Adverse Drug Reaction using Machine Learning and Deep Learning Based on an Imbalanced Electronic Medical Records Dataset. 2021.

[7] J. Zhao, Learning predictive models from electronic health records, Department of Computer and Systems Sciences, Stockholm University, 2017.

[8] A. Gupta, T. Liu, C. Crick, Utilizing time series data embedded in electronic health records to develop continuous mortality risk prediction models using hidden Markov models: A sepsis case study, Stat. Methods Med. Res. 29 (11) (2020) 3409–3423.

[9] B.A. Goldstein, A.M. Navar, M.J. Pencina, J.P. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, J. Am. Med. Inform Assoc. 24 (1) (2017) 198–208.

[10] F. Xie, B. Chakraborty, M.E.H. Ong, B.A. Goldstein, N. Liu, AutoScore: A Machine Learning-Based Automatic Clinical Score Generator and Its Application to Mortality Prediction Using Electronic Health Records, JMIR Med. Inform. 8 (10) (2020) e21798, https://doi.org/10.2196/21798.

[11] J. Zhao, P. Papapetrou, L. Asker, H. Boström, Learning from heterogeneous temporal data in electronic health records, J. Biomed. Inform. 65 (2017) 105–119.

[12] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, Recurrent Neural Networks for Multivariate Time Series with Missing Values, Sci. Rep. 8 (1) (2018) 6085.

[13] H. Hewamalage, C. Bergmeir, K. Bandara, Recurrent Neural Networks for Time Series Forecasting: Current status and future directions, Int. J. Forecast. 37 (1) (2021) 388–427.

[14] Baytas IM, Xiao C, Zhang X, Wang F, Jain AK, Zhou J. Patient Subtyping via Time-Aware LSTM Networks. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017; Halifax, NS, Canada.

[15] G. Maragatham, S. Devi, LSTM Model for Prediction of Heart Failure in Big Data, J. Med. Syst. 43 (5) (2019) 111.

[16] W. Lu, L. Ma, H. Chen, X. Jiang, M. Gong, A Clinical Prediction Model in Health Time Series Data Based on Long Short-Term Memory Network Optimized by Fruit Fly Optimization Algorithm, IEEE Access 8 (2020) 136014–136023.

[17] Khoshnevisan F, Ivy J, Capan M, Arnold R, Huddleston J, Chi M. Recent Temporal Pattern Mining for Septic Shock Early Prediction. Paper presented at: 2018 IEEE International Conference on Healthcare Informatics (ICHI); 4-7 June 2018, 2018.

[18] H.J. Park, D.Y. Jung, W. Ji, C.-M. Choi, 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 (8) (2020) e19512, https://doi.org/10.2196/19512.

[19] B.K. Reddy, D. Delen, Predicting hospital readmission for lupus patients: An RNN-LSTM-based deep-learning methodology, Comput. Biol. Med. 101 (2018) 199–209.

[20] Yang Y, Fasching PA, Tresp V. Predictive Modeling of Therapy Decisions in Metastatic Breast Cancer with Recurrent Neural Network Encoder and Multinomial Hierarchical Regression Decoder. Paper presented at: 2017 IEEE International Conference on Healthcare Informatics (ICHI); 23-26 Aug. 2017, 2017.

[21] Hung C, Chen W, Lai P, Lin C, Lee C. Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database. Paper presented at: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 11-15 July 2017, 2017.

[22] C. Xiao, E. Choi, J. Sun, Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review, J. Am. Med. Inform Assoc. 25 (10) (2018) 1419–1428.

[23] B. Shickel, P.J. Tighe, A. Bihorac, P. Rashidi, Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis, IEEE J. Biomed. Health Inform. 22 (5) (2018) 1589–1604.

[24] J.R. Ayala Solares, F.E. Diletta Raimondi, Y. Zhu, F. Rahimian, D. Canoy, J. Tran, A.C. Pinho Gomes, A.H. Payberah, M. Zottoli, M. Nazarzadeh, N. Conrad, K. Rahimi, G. Salimi-Khorshidi, Deep learning for electronic health records: A comparative review of multiple deep neural architectures, J. Biomed. Inform. 101 (2020) 103337, https://doi.org/10.1016/j.jbi.2019.103337.

[25] Y. Si, J. Du, Z. Li, X. Jiang, T. Miller, F. Wang, W. Jim Zheng, K. Roberts, Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review, J. Biomed. Inform. 115 (2021) 103671, https://doi.org/10.1016/j.jbi.2020.103671.

[26] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.

[27] A.E.W. Johnson, T.J. Pollard, L.u. Shen, L.-W. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R.G. Mark, MIMIC-III, a freely accessible critical care database, Sci Data. 3 (1) (2016), https://doi.org/10.1038/sdata.2016.35.

[28] T.A. Lasko, J.C. Denny, M.A. Levy, J. Devaney, Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data, PLoS ONE 8 (6) (2013).

[29] Esteban C, Schmidt D, Krompaß D, Tresp V. Predicting Sequences of Clinical Events by Using a Personalized Temporal Latent Embedding Model. Paper presented at: 2015 International Conference on Healthcare Informatics; 21-23 Oct. 2015, 2015.

[30] Mehrabi S, Sohn S, Li D, et al. Temporal Pattern and Association Discovery of Diagnosis Codes Using Deep Learning. Paper presented at: 2015 International Conference on Healthcare Informatics; 21-23 Oct. 2015, 2015.
[31] T. Tran, T.D. Nguyen, D. Phung, S. Venkatesh, Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM), J. Biomed. Inform. 54 (2015) 96–105.
[32] E. Choi, M.T. Bahadori, A. Schuetz, W.F. Stewart, J. Sun, Doctor AI: Predicting Clinical Events via Recurrent Neural Networks, in: JMLR Workshop Conf Proc, 2016, pp. 301–318.
[33] R. Miotto, L. Li, B.A. Kidd, J.T. Dudley, Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records, Sci Rep. 6 (2016) 26094.
[34] Zhu Z, Yin C, Qian B, Cheng Y, Wei J, Wang F. Measuring Patient Similarities via a Deep Architecture with Medical Concept Embedding. Paper presented at: 2016 IEEE 16th International Conference on Data Mining (ICDM); 12-15 Dec. 2016, 2016.
[35] E. Choi, M.T. Bahadori, J. Sun, J. Kulas, A. Schuetz, W. Stewart, Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. Paper presented at: Advances in Neural Information Processing Systems2016, 2016.
[36] Che Z, Cheng Y, Zhai S, Sun Z, Liu Y. Boosting Deep Learning Risk Prediction with Generative Adversarial Networks for Electronic Health Records. Paper presented at: 2017 IEEE International Conference on Data Mining (ICDM); 18-21 Nov. 2017, 2017.
[37] Che Z, Liu Y. Deep Learning Solutions to Computational Phenotyping in Health Care. Paper presented at: 2017 IEEE International Conference on Data Mining Workshops (ICDMW); 18-21 Nov. 2017, 2017.
[38] Feng Y, Min X, Chen N, et al. Patient outcome prediction via convolutional neural networks based on multi-granularity medical concept embedding. Paper presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 13-16 Nov. 2017, 2017.
[39] J. Mei, S. Zhao, F. Jin, et al., Deep Diabetologist: Learning to Prescribe Hypoglycemic Medications with Recurrent Neural Networks, Stud Health Technol Inform. 245 (2017) 1277.
[40] P. Nguyen, T. Tran, N. Wickramasinghe, S. Venkatesh, Deepr: A Convolutional Net for Medical Records, IEEE J. Biomed. Health. Inf. 21 (1) (2017) 22–30.
[41] T. Pham, T. Tran, D. Phung, S. Venkatesh, Predicting healthcare trajectories from medical records: A deep learning approach, J. Biomed. Inform. 69 (2017) 218–229.
[42] Y. Sha, M.D. Wang, Interpretable Predictions of Clinical Outcomes with An Attention-based Recurrent Neural Network, Acm bcb. 2017 (2017) 233–240.
[43] J. Stojanovic, D. Gligorijevic, V. Radosavljevic, N. Djuric, M. Grbovic, Z. Obradovic, Modeling Healthcare Quality via Compact Representations of Electronic Health Records, IEEE/ACM Trans. Comput. Biol. Bioinformat. 14 (3) (2017) 545–554.
[44] Q. Suo, F. Ma, G. Canino, et al., A Multi-Task Framework for Monitoring Health Conditions via Attention-based Recurrent Neural Networks, AMIA Annu. Symp. Proc. 2017 (2017) 1665–1674.
[45] Suo Q, Ma F, Yuan Y, et al. Personalized disease prediction using a CNN-based similarity learning method. Paper presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 13-16 Nov. 2017, 2017.
[46] Zheng K, Wang W, Gao J, Ngiam KY, Ooi BC, Yip WLJ. Capturing Feature-Level Irregularity in Disease Progression Modeling. Proceedings of the 2017 ACM on Conference on Information and Knowledge Management; 2017; Singapore, Singapore.
[47] K. Yang, X. Li, H. Liu, et al., TaGiTeD: Predictive task guided tensor decomposition for representation learning from electronic health records. Paper presented at: Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
[48] T. Bai, B.L. Egleston, S. Zhang, S. Vucetic, Interpretable Representation Learning for Healthcare via Capturing Disease Progression through Time, Kdd. 2018 (2018) 43–51.
[49] Cheung BLP, Dahl D. Deep learning from electronic medical records using attention-based cross-modal convolutional neural networks. Paper presented at: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 4-7 March 2018, 2018.
[50] Le H, Tran T, Venkatesh S. Dual Memory Neural Computer for Asynchronous Two-view Sequential Learning. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2018; London, United Kingdom.
[51] Lee W, Park S, Joo W, Moon I. Diagnosis Prediction via Medical Context Attention Networks Using Deep Generative Modeling. Paper presented at: 2018 IEEE International Conference on Data Mining (ICDM); 17-20 Nov. 2018, 2018.
[52] Lei L, Zhou Y, Zhai J, et al. An Effective Patient Representation Learning for Time-series Prediction Tasks Based on EHRs. Paper presented at: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 3-6 Dec. 2018, 2018.
[53] Lin C, Zhang Y, Ivy J, et al. Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM. Paper presented at: 2018 IEEE International Conference on Healthcare Informatics (ICHI); 4-7 June 2018, 2018.
[54] Ma F, You Q, Xiao H, Chitta R, Zhou J, Gao J. KAME: Knowledge-based Attention Model for Diagnosis Prediction in Healthcare. Proceedings of the 27th ACM International Conference on Information and Knowledge Management; 2018; Torino, Italy.
[55] Nguyen P, Tran T, Venkatesh S. Resset: A Recurrent Model for Sequence of Sets with Applications to Electronic Medical Records. Paper presented at: 2018 International Joint Conference on Neural Networks (IJCNN); 8-13 July 2018, 2018.
[56] Park HD, Han Y, Choi JH. Frequency-Aware Attention based LSTM Networks for Cardiovascular Disease. Paper presented at: 2018 International Conference on Information and Communication Technology Convergence (ICTC); 17-19 Oct. 2018, 2018.
[57] Park S, Kim YJ, Kim JW, Park JJ, Ryu B, Ha J. [Regular Paper] Interpretable Prediction of Vascular Diseases from Electronic Health Records via Deep Attention Networks. Paper presented at: 2018 IEEE 18th International Conference on Bioinformatics and Bioengineering (BIBE); 29-31 Oct. 2018, 2018.
[58] A. Rajkomar, E. Oren, K. Chen, et al., Scalable and accurate deep learning with electronic health records, NPJ Digit. Med. 1 (2018) 18.
[59] Q. Suo, F. Ma, Y. Yuan, et al., Deep Patient Similarity Learning for Personalized Healthcare, IEEE Trans. Nanobiosci. 17 (3) (2018) 219–227.
[60] Suresh H, Gong JJ, Guttag JV. Learning Tasks for Multitask Learning: Heterogenous Patient Populations in the ICU. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2018; London, United Kingdom.
[61] S. Wu, S. Liu, S. Sohn, et al., Modeling asynchronous event sequences with RNNs, J. Biomed. Inform. 83 (2018) 167–177.
[62] C. Xiao, T. Ma, A.B. Dieng, D.M. Blei, F. Wang, Readmission prediction via deep contextual embedding of clinical concepts, PLoS ONE 13 (4) (2018), e0195024.
[63] Yang X, Zhang Y, Chi M. Time-aware Subgroup Matrix Decomposition: Imputing Missing Data Using Forecasting Events. Paper presented at: 2018 IEEE International Conference on Big Data (Big Data); 10-13 Dec. 2018, 2018.
[64] J. Zhang, K. Kowsari, J.H. Harrison, J.M. Lobo, L.E. Barnes, Patient2Vec: A Personalized Interpretable Deep Representation of the Longitudinal Electronic Health Record, IEEE Access 6 (2018) 65333–65346.
[65] Z. Huang, W. Dong, H. Duan, J. Liu, A Regularized Deep Learning Approach for Clinical Risk Prediction of Acute Coronary Syndrome Using Electronic Health Records, IEEE Trans. Biomed. Eng. 65 (5) (2018) 956–968.
[66] Edward Choi, Xiao Cao, F. WalterStewart, Sun. Jimeng, MiME: multilevel medical embedding of electronic health records for predictive healthcare, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18), Curran Associates Inc., Red Hook, NY, USA, 2018, pp. 4552–4562.
[67] Y. An, N. Huang, X. Chen, F. Wu, J. Wang, High-risk Prediction of Cardiovascular Diseases via Attention-based Deep Neural Networks, IEEE/ACM Trans. Comput. Biol. Bioinform. (2019).
[68] A. Ashfaq, A. Sant'Anna, M. Lingman, S. Nowaczyk, Readmission prediction using deep learning on electronic health records, J. Biomed. Inform. 97 (2019), 103256.
[69] S. Fiorini, F. Hajati, A. Barla, F. Girosi, Predicting diabetes second-line therapy initiation in the Australian population via time span-guided neural attention network, PLoS ONE 14 (10) (2019), e0211844.
[70] W. Guo, W. Ge, L. Cui, H. Li, L. Kong, An Interpretable Disease Onset Predictive Model Using Crossover Attention Mechanism From Electronic Health Records, IEEE Access 7 (2019) 134236–134244.
[71] Jun E, Mulyadi AW, Suk H. Stochastic Imputation and Uncertainty-Aware Attention to EHR for Mortality Prediction. Paper presented at: 2019 International Joint Conference on Neural Networks (IJCNN); 14-19 July 2019, 2019.
[72] B.C. Kwon, M. Choi, J.T. Kim, et al., RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records, IEEE Trans. Visual Comput. Graphics 25 (1) (2019) 299–309.
[73] Lee JM, Hauskrecht M. Recent Context-aware LSTM for Clinical Event Time-series Prediction. *Artif Intell Med Conf Artif Intell Med (2005-)*. 2019;11526:13-23.
[74] Q.T. Li, Y. Xu, VS-GRU: A Variable Sensitive Gated Recurrent Neural Network for Multivariate Time Series with Massive Missing Values, Appl. Sci.-Basel 9 (15) (2019).
[75] Y.W. Lin, Y.Q. Zhou, F. Faghri, M.J. Shawl, R.H. Campbell, Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long shortterm memory, PLoS ONE 14 (7) (2019).
[76] L. Liu, H. Li, Z. Hu, et al., Learning Hierarchical Representations of Electronic Health Records for Clinical Outcome Prediction, AMIA Annu Symp Proc. 2019 (2019) 597–606.
[77] Liu L, Wu H, Wang Z, Liu Z, Zhang M. Early Prediction of Sepsis From Clinical Data via Heterogeneous Event Aggregation. Paper presented at: 2019 Computing in Cardiology (CinC); 8-11 Sept. 2019, 2019.
[78] Macias E, Boquet G, Serrano J, Vicario J, Ibeas J, Morel A. Novel Imputing Method and Deep Learning Techniques for Early Prediction of Sepsis in Intensive Care Units. Paper presented at: 2019 Computing in Cardiology (CinC); 8-11 Sept. 2019, 2019.
[79] Peng X, Long G, Shen T, Wang S, Jiang J, Blumenstein M. Temporal Self-Attention Network for Medical Concept Embedding. Paper presented at: 2019 IEEE International Conference on Data Mining (ICDM); 8-11 Nov. 2019, 2019.
[80] T. Ruan, L. Lei, Y. Zhou, et al., Representation learning for clinical time series prediction tasks in electronic health records, BMC Med. Inform. Decis. Mak. 19 (Suppl 8) (2019) 259.
[81] L. Wang, H. Wang, Y. Song, Q. Wang, MCPL-Based FT-LSTM: Medical Representation Learning-Based Clinical Prediction Model for Time Series Events, IEEE Access 7 (2019) 70253–70264.
[82] Wang S, Ren P, Chen Z, Ren Z, Ma J, Rijke Md. Order-free Medicine Combination Prediction with Graph Convolutional Reinforcement Learning. Proceedings of the 28th ACM International Conference on Information and Knowledge Management; 2019; Beijing, China.

[83] Wang W, Guo C, Xu J, Liu A. Bi-Dimensional Representation of Patients for Diagnosis Prediction. Paper presented at: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC); 15-19 Jul 2019, 2019.

[84] Wang Z, Li H, Liu L, Wu H, Zhang M. Predictive Multi-level Patient Representations from Electronic Health Records. Paper presented at: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 18-21 Nov. 2019, 2019.

[85] Y. Xiang, J. Xu, Y. Si, et al., Time-sensitive clinical concept embeddings learned from large electronic health records, BMC Med. Inform. Decis. Mak. 19 (Suppl 2) (2019) 58.

[86] Xu E, Zhao S, Mei J, Xia E, Yu Y, Huang S. Multiple MACE Risk Prediction using Multi-Task Recurrent Neural Network with Attention. Paper presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); 10-13 June 2019, 2019.

[87] Yang F, Yu Z, Liang Y, et al. Grouped Correlational Generative Adversarial Networks for Discrete Electronic Health Records. Paper presented at: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 18-21 Nov. 2019, 2019.

[88] Zhang J, Yin P. Multivariate Time Series Missing Data Imputation Using Recurrent Denoising Autoencoder. Paper presented at: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 18-21 Nov. 2019, 2019.

[89] Zhang X, Qian B, Li Y, Yin C, Wang X, Zheng Q. KnowRisk: An Interpretable Knowledge-Guided Model for Disease Risk Prediction. Paper presented at: 2019 IEEE International Conference on Data Mining (ICDM); 8-11 Nov. 2019, 2019.

[90] Zhang Y, Yang X, Ivy J, Chi M. Time-aware Adversarial Networks for Adapting Disease Progression Modeling. Paper presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); 10-13 June 2019, 2019.

[91] Zhang XS, Tang F, Dodge HH, Zhou J, Wang F. MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2019; Anchorage, AK, USA.

[92] Afshar A, Perros I, Park H, et al. TASTE: temporal and static tensor factorization for phenotyping electronic health records. Proceedings of the ACM Conference on Health, Inference, and Learning; 2020; Toronto, Ontario, Canada.

[93] Y. An, Y. Mao, L. Zhang, et al., RAHM: Relation augmented hierarchical multi-task learning framework for reasonable medication stocking, J. Biomed. Inform. 108 (2020), 103502.

[94] S. Barbieri, J. Kemp, O. Perez-Concha, et al., Benchmarking Deep Learning Architectures for Predicting Readmission to the ICU and Describing Patients-at-Risk, Sci. Rep. 10 (1) (2020) 1111.

[95] J. Chu, W. Dong, Z. Huang, Endpoint prediction of heart failure using electronic health records, J. Biomed. Inform. 109 (2020), 103518.

[96] H. Duan, Z. Sun, W. Dong, K. He, Z. Huang, On Clinical Event Prediction in Patient Treatment Trajectory Using Longitudinal Electronic Health Records, IEEE J. Biomed. Health Inform. 24 (7) (2020) 2053–2063.

[97] Gao J, Xiao C, Glass LM, Sun J. COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2020; Virtual Event, CA, USA.

[98] Gao J, Xiao C, Wang Y, Tang W, Glass LM, Sun J. StageNet: Stage-Aware Neural Networks for Health Risk Prediction. Proceedings of The Web Conference 2020; 2020; Taipei, Taiwan.

[99] Z. Jin, S. Cui, S. Guo, D. Gotz, J. Sun, C.N. CarePre, An Intelligent Clinical Decision Assistance System, ACM Trans. Comput. Healthcare 1 (1) (2020) Article 6.

[100] E. Jun, A.W. Mulyadi, J. Choi, H.I. Suk, Uncertainty-Gated Stochastic Sequential Model for EHR Mortality Prediction, IEEE Trans. Neural Netw. Learn Syst. (2020).

[101] I. Landi, B.S. Glicksberg, H.C. Lee, et al., Deep representation learning of electronic health records to unlock patient stratification at scale, NPJ. Digit. Med. 3 (2020) 96.

[102] S.M. Lauritsen, M.E. Kalør, E.L. Kongsgaard, et al., Early detection of sepsis utilizing deep learning on electronic health record event sequences, Artif Intell. Med. 104 (2020), 101820.

[103] Y. Li, B. Qian, X. Zhang, H. Liu, Graph Neural Network-Based Diagnosis Prediction, Big Data (2020).

[104] Y. Li, S. Rao, J.R.A. Solares, et al., BEHRT: Transformer for Electronic Health Records, Sci. Rep. 10 (1) (2020) 7155.

[105] Y. Li, Z. Zhu, H. Wu, S. Ding, Y. Zhao, CCAE: Cross-field categorical attributes embedding for cancer clinical endpoint prediction, Artif. Intell Med. 107 (2020), 101915.

[106] D. Liu, Y.L. Wu, X. Li, L. Qi, Medi-Care AI: Predicting medications from billing codes via robust recurrent neural networks, Neural Netw. 124 (2020) 109–116.

[107] Liu SC, Li T, Ding HY, et al. A hybrid method of recurrent neural network and graph neural network for next-period prescription prediction. *International Journal of Machine Learning and Cybernetics.*.

[108] Luo J, Ye M, Xiao C, Ma F. HiTANet: Hierarchical Time-Aware Attention Networks for Risk Prediction on Electronic Health Records. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2020; Virtual Event, CA, USA.

[109] Panigutti C, Perotti A, Pedreschi D. Doctor XAI: an ontology-based approach to black-box sequential data classification explanations. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency; 2020; Barcelona, Spain.

[110] Z. Qiao, Z. Zhang, X. Wu, S. Ge, W. Fan, MHM: Multi-modal Clinical Data based Hierarchical Multi-label Diagnosis Prediction. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval; 2020; Virtual Event, China, Virtual Event, China, 2020.

[111] S. Rongali, A.J. Rose, D.D. McManus, et al., Learning Latent Space Representations to Predict Patient Outcomes: Model Development and Validation, J. Med. Internet Res. 22 (3) (2020), e16374.

[112] J. Song, Y. Wang, S. Tang, et al., Local-Global Memory Neural Network for Medication Prediction, IEEE Trans. Neural Netw. Learn Syst. (2020).

[113] C. Su, S. Gao, S. Li, GATE: Graph-Attention Augmented Temporal Neural Network for Medication Recommendation, IEEE Access 8 (2020) 125447–125458.

[114] Z. Wang, Y. Zhu, D. Li, Y. Yin, J. Zhang, Feature rearrangement based deep learning system for predicting heart failure mortality, Comput Meth. Programs Biomed. 191 (2020), 105383.

[115] Y. Xiang, H. Ji, Y. Zhou, 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 (7) (2020), e16981.

[116] Yin C, Liu R, Zhang D, Zhang P. Identifying Sepsis Subphenotypes via Time-Aware Multi-Modal Auto-Encoder. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2020; Virtual Event, CA, USA.

[117] K. Yu, M. Zhang, T. Cui, M. Hauskrecht, Monitoring ICU Mortality Risk with A Long Short-Term Memory Recurrent Neural Network, in: Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing, 2020, pp. 103–114.

[118] R. Yu, Y. Zheng, R. Zhang, Y. Jiang, C.C.Y. Poon, Using a Multi-Task Recurrent Neural Network With Attention Mechanisms to Predict Hospital Mortality of Patients, IEEE J. Biomed. Health Inform. 24 (2) (2020) 486–492.

[119] X. Zeng, Y. Feng, S. Moosavinasab, D. Lin, S. Lin, C. Liu, Multilevel Self-Attention Model and its Use on Medical Risk Prediction, Pac. Symp. Biocomput. 25 (2020) 115–126.

[120] Zhang M, King CR, Avidan M, Chen Y. Hierarchical Attention Propagation for Healthcare Representation Learning. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2020; Virtual Event, CA, USA.

[121] Zheng K, Cai S, Chua HR, Wang W, Ngiam KY, Ooi BC. TRACER: A Framework for Facilitating Accurate and Interpretable Analytics for High Stakes Applications. Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data; 2020; Portland, OR, USA.

[122] H.C. Thorsen-Meyer, A.B. Nielsen, A.P. Nielsen, 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 Digital Health 2 (4) (2020) E179–E191.

[123] N.Y. Jane, K.H. Nehemiah, K. Arputharaj, A Temporal Mining Framework for Classifying Un-Evenly Spaced Clinical Data: An Approach for Building Effective Clinical Decision-Making System, Appl. Clin. Inform. 7 (1) (2016) 1–21.

[124] Sarafrazi S, Choudhari RS, Mehta C, et al. Cracking the "Sepsis" Code: Assessing Time Series Nature of EHR Data, and Using Deep Learning for Early Sepsis Prediction. Paper presented at: 2019 Computing in Cardiology (CinC); 8-11 Sept. 2019, 2019.

[125] Strauman AS, Bianchi FM, Mikalsen KØ, Kampffmeyer M, Soguero-Ruiz C, Jenssen R. Classification of postoperative surgical site infections from blood measurements with missing data using recurrent neural networks. Paper presented at: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 4-7 March 2018, 2018.

[126] T. Köse, S. Özgür, E. Coşgun, A. Keskinoğlu, P. Keskinoğlu, Effect of Missing Data Imputation on Deep Learning Prediction Performance for Vesicoureteral Reflux and Recurrent Urinary Tract Infection Clinical Study, Biomed Res. Int. 2020 (2020) 1895076.

[127] J. Shao, D.C. Jordan, Y.L. Pritchett, Baseline observation carry forward: reasoning, properties, and practical issues, J. Biopharm. Stat. 19 (4) (2009) 672–684.

[128] Y. Cheng, F. Wang, P. Zhang, J. Hu, Risk prediction with electronic health records: A deep learning approach. Paper presented at: Proceedings of the 2016 SIAM International Conference on Data Mining, 2016.

[129] A.V. Uriarte-Arcia, I. Lopez-Yanez, C. Yanez-Marquez, One-hot vector hybrid associative classifier for medical data classification, PLoS ONE 9 (4) (2014), e95715.

[130] P. Krawczyk, L. Swiecicki, ICD-11 vs. ICD-10 - a review of updates and novelties introduced in the latest version of the WHO International Classification of Diseases, Psychiatr. Pol. 54 (1) (2020) 7–20.

[131] Current procedural terminology (CPT). *JAMA.* 1970;212(5):873-874.

[132] J.S. Carter, S.H. Brown, M.S. Erlbaum, et al., Initializing the VA medication reference terminology using UMLS metathesaurus co-occurrences, Proc AMIA Symp. 116–120 (2002).

[133] D. Lee, R. Cornet, F. Lau, N. de Keizer, A survey of SNOMED CT implementations, J. Biomed. Inform. 46 (1) (2013) 87–96.

[134] M. Charlson, T.P. Szatrowski, J. Peterson, J. Gold, Validation of a combined comorbidity index, J. Clin. Epidemiol. 47 (11) (1994) 1245–1251.

[135] S. Liu, W. Ma, R. Moore, V. Ganesan, S. Nelson, RxNorm: prescription for electronic drug information exchange, IT Prof. 7 (5) (2005) 17–23.

[136] S.M. Huff, R.A. Rocha, C.J. McDonald, et al., Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary, J. Am. Med. Inform. Assoc. 5 (3) (1998) 276–292.

[137] Peng X, Long G, Pan S, Jiang J, Niu Z. Attentive Dual Embedding for Understanding Medical Concepts in Electronic Health Records. Paper presented at: 2019 International Joint Conference on Neural Networks (IJCNN); 14-19 July 2019, 2019.

[138] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2; 2013; Lake Tahoe, Nevada.

[139] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, J. Am. Med. Inform. Assoc. 24 (2) (2017) 361–370.

[140] E. Choi, M.T. Bahadori, E. Searles, et al., Multi-layer representation learning for medical concepts. Paper presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[141] Lu Q, Silva Nd, Kafle S, et al. Learning Electronic Health Records through Hyperbolic Embedding of Medical Ontologies. Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; 2019; Niagara Falls, NY, USA.

[142] Sun J, Wang F, Hu J, Edabollahi S. Supervised patient similarity measure of heterogeneous patient records. *Acm Sigkdd Explorations Newsletter.* 2012;14(1):16-24 %@ 1931-0145.

[143] H. Harutyunyan, H. Khachatrian, D.C. Kale, G. Ver Steeg, A. Galstyan, Multitask learning and benchmarking with clinical time series data, Sci. Data 6 (1) (2019) 96.

[144] C. Shivade, P. Raghavan, E. Fosler-Lussier, et al., A review of approaches to identifying patient phenotype cohorts using electronic health records, J. Am. Med. Inform. Assoc. 21 (2) (2014) 221–230.

[145] Amarasinghe K, Kenney K, Manic M. Toward Explainable Deep Neural Network Based Anomaly Detection. Paper presented at: 2018 11th International Conference on Human System Interaction (HSI); 4-6 July 2018, 2018.

[146] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nature Machine Intelligence. 1 (5) (2019) 206–215.

[147] Z.C. Lipton, The mythos of model interpretability, Queue 16 (3) (2018) 31–57.

[148] Frosst N, Hinton G. Distilling a neural network into a soft decision tree. *arXiv preprint arXiv:171109784.* 2017.

[149] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:14090473.* 2014.

[150] Luong M-T, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:150804025.* 2015.

[151] P.Y. Shi, F. Hou, X.W. Zheng, F. Yuan, Analysis of electronic health records based on long short-term memory. Concurrency and Computation-Practice & Experience, 2020.

[152] P. Chen, W. Dong, J. Wang, X. Lu, U. Kaymak, Z. Huang, Interpretable clinical prediction via attention-based neural network, BMC Med. Inform. Decis Mak. 20 (Suppl 3) (2020) 131.

[153] D.A. Kaji, J.R. Zech, J.S. Kim, et al., An attention based deep learning model of clinical events in the intensive care unit, PLoS ONE 14 (2) (2019), e0211057.

[154] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, S. Mougiakakou, Multisource Transfer Learning With Convolutional Neural Networks for Lung Pattern Analysis, IEEE J. Biomed. Health Inform. 21 (1) (2017) 76–84.

[155] G.I. Parisi, R. Kemker, J.L. Part, C. Kanan, S. Wermter, Continual lifelong learning with neural networks: A review, Neural Netw. 113 (2019) 54–71.

[156] M. De Bois, M.A. El Yacoubi, M. Ammi, Adversarial multi-source transfer learning in healthcare: Application to glucose prediction for diabetic people, Comput Meth. Programs Biomed. 199 (2021), 105874.

[157] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Paper presented at: Proceedings of the IEEE international conference on computer vision2017.

[158] T. He, J. Guo, N. Chen, et al., MediMLP: Using Grad-CAM to Extract Crucial Variables for Lung Cancer Postoperative Complication Prediction, IEEE J. Biomed. Health Inform. 24 (6) (2020) 1762–1771.

[159] Laugel T, Lesot M-J, Marsala C, Renard X, Detyniecki M. The dangers of post-hoc interpretability: Unjustified counterfactual explanations. *arXiv preprint arXiv:190709294.* 2019.

[160] Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: Fundamental principles and 10 grand challenges. *arXiv preprint arXiv:210311251.* 2021.

[161] Ahmad MA, Eckert C, Teredesai A. Interpretable machine learning in healthcare. Paper presented at: Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics2018.

[162] B. Ustun, C. Rudin, Supersparse linear integer models for optimized medical scoring systems, Machine Learning. 102 (3) (2016) 349–391.

[163] B. Ustun, C. Rudin, Learning Optimized Risk Scores, J. Machine Learning Res. 20 (150) (2019) 1–75.

[164] Yuan H, Xie F, Ong MEH, et al. AutoScore-Imbalance: An interpretable machine learning tool for development of clinical scores with rare events data. *arXiv preprint arXiv:210706039.* 2021.

[165] F. Xie, Y. Ning, H. Yuan, et al., AutoScore-Survival: Developing interpretable machine learning-based time-to-event scores with right-censored survival data, J. Biomed. Inform. 125 (2021), 103959, https://doi.org/10.1016/j.jbi.2021.103959 [published online ahead of print, 2021 Nov 23].

[166] F. Xie, M.E.H. Ong, J.N.M.H. Liew, et al., Development and assessment of an interpretable machine learning triage tool for estimating mortality after emergency admissions, JAMA Netw. Open 4 (8) (2021), e2118467, https://doi.org/10.1001/jamanetworkopen.2021.18467.

[167] Wang S, Cao J, Yu P. Deep Learning for Spatio-Temporal Data Mining: A Survey. IEEE Transactions on Knowledge and Data Engineering. 2020:1-1.

[168] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Paper presented at: Advances in neural information processing systems2017.

[169] Tolstikhin I, Houlsby N, Kolesnikov A, et al. Mlp-mixer: An all-mlp architecture for vision. *arXiv preprint arXiv:210501601.* 2021.