Journal of Biomedical Informatics 144 (2023) 104430

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Deep learning prediction models based on EHR trajectories: A systematic review

Ali Amirahmadi [a,*], Mattias Ohlsson [a,b], Kobra Etminani [a]

[a] Center for Applied Intelligent Systems Research, Halmstad University, Sweden
[b] Computational Biology & Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, Sweden

A R T I C L E I N F O

Keywords:
Disease prediction
EHR trajectories
Systematic review
Electronic health records
Deep learning

A B S T R A C T

Background : Electronic health records (EHRs) are generated at an ever-increasing rate. EHR trajectories, the temporal aspect of health records, facilitate predicting patients' future health-related risks. It enables healthcare systems to increase the quality of care through early identification and primary prevention. Deep learning techniques have shown great capacity for analyzing complex data and have been successful for prediction tasks using complex EHR trajectories. This systematic review aims to analyze recent studies to identify challenges, knowledge gaps, and ongoing research directions.

Methods: For this systematic review, we searched Scopus, PubMed, IEEE Xplore, and ACM databases from Jan 2016 to April 2022 using search terms centered around EHR, deep learning, and trajectories. Then the selected papers were analyzed according to publication characteristics, objectives, and their solutions regarding existing challenges, such as the model's capacity to deal with intricate data dependencies, data insufficiency, and explainability.

Results : After removing duplicates and out-of-scope papers, 63 papers were selected, which showed rapid growth in the number of research in recent years. Predicting all diseases in the next visit and the onset of cardiovascular diseases were the most common targets. Different contextual and non-contextual representation learning methods are employed to retrieve important information from the sequence of EHR trajectories. Recurrent neural networks and the time-aware attention mechanism for modeling long-term dependencies, self-attentions, convolutional neural networks, graphs for representing inner visit relations, and attention scores for explainability were frequently used among the reviewed publications.

Conclusions: This systematic review demonstrated how recent breakthroughs in deep learning methods have facilitated the modeling of EHR trajectories. Research on improving the ability of graph neural networks, attention mechanisms, and cross-modal learning to analyze intricate dependencies among EHRs has shown good progress. There is a need to increase the number of publicly available EHR trajectory datasets to allow for easier comparison among different models. Also, very few developed models can handle all aspects of EHR trajectory data.

1. Introduction

Healthcare systems usually record patients' histories in the form of electronic health records (EHRs). Generated data during routine delivery of health care, accumulate into EHRs longitudinally, and forge EHR trajectories [1]. Typically EHRs cover various dimensions of patients' health status, such as demographics, diagnoses, medications, procedures, vital signs, test results, medical imagery, discharge summaries, physician notes, and nursing notes in a sequence of visits (Fig. 1). Analyzing and discovering unknown patterns in health trajectories that lead toward adverse outcomes facilities early identification and primary prevention.

Recent progress in machine learning and its subdomains, particularly deep learning (DL) methods, has provided new and highly capable tools for pattern discovery. Traditional machine learning techniques need help to effectively extract and organize discriminative features from data, often necessitating pre-processing and data transformation pipelines and significant input from human experts. In contrast, deep learning methods leverage multiple layers of neural processing to automatically learn abstract data representations without expert intervention. As a result, deep learning methods are particularly well-suited for modeling the complex medical histories of electronic health records (EHRs) [2,3].

* Corresponding author.
E-mail address: ali.amirahmadi@hh.se (A. Amirahmadi).

https://doi.org/10.1016/j.jbi.2023.104430
Received 7 October 2022; Received in revised form 8 June 2023; Accepted 17 June 2023
Available online 26 June 2023

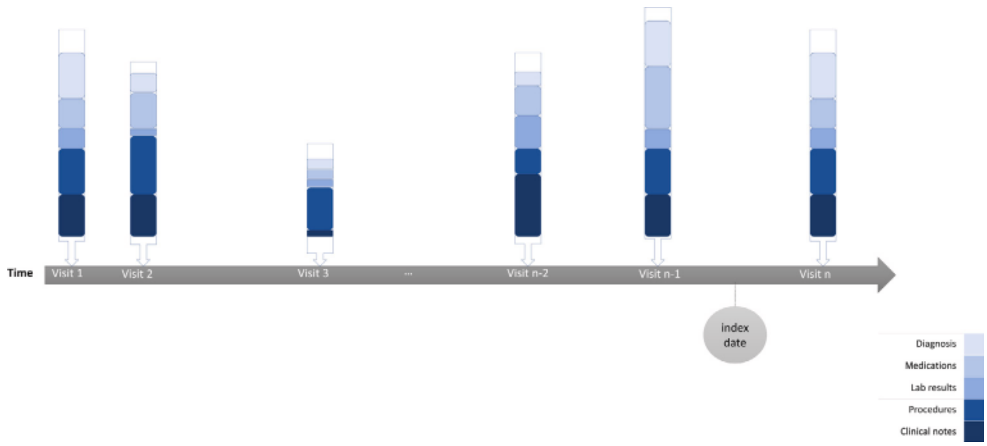

Fig. 1. A schematic example of an EHR for a patient with *n* visits prior to some index date and a possible future visit. Each visit generates various amounts of the healthcare data, such as diagnoses, prescribed medications, laboratory results, the indication of procedures used, and clinical notes. The amount of information can vary between visits, as indicated by the height of the colored bars. Visits are not regular in time.

This emerging research area of using DL methods on EHR data needs to address various challenges related to data and modeling. Such challenges include representing high dimensional, sparse, and heterogeneous data from different sources with multiple data modalities into effective feature spaces, dealing with data insufficiency, and model explainability. To fully take advantage of the longitudinal nature of EHR trajectories, the DL methods should be capable of modeling time series, and this requirement adds additional challenges.

This systematic review aims to analyze the recent studies in DL methods for predicting patient future health-related risks by analyzing EHR trajectories, to identify challenges, knowledge gaps, and ongoing research directions. Furthermore, this review focuses on end-to-end DL challenges for the predication tasks, in comparison with other reviews on the same topic [4–7]. In the rest of the paper, we will identify and analyze the most critical challenges and the proposed solutions. To this end, we devised a search and selection strategy, then evaluated eligible papers to explore ongoing research challenges and directions and discussed solutions and the knowledge gaps.

Statement of significance

Problem. Identifying challenges, knowledge gaps, and ongoing research directions in DL methods for predicting patients' future health-related risks using EHR trajectories.

What is Already Known. Recent developments in DL methods and the availability of large EHR trajectory data provide promising opportunities for knowledge discovery and the development of prediction models. Analyzing and discovering unknown patterns in EHR trajectories that lead toward adverse outcomes, facilities early identification, and primary prevention. RNNs and data concatenation are frequently used for EHR trajectories representation learning and data aggregation, respectively.

What This Paper Adds. This review showed that more recent studies have used self-supervised learning, language models, and multimodal embedding methods for representation learning and to aggregate heterogeneous EHR data sources. Furthermore, graph neural networks and transformers have become popular because of their capability for efficient representation learning and model explainability. In addition, this review critically assessed the advantages and disadvantages of the methods used.

2. Materials and methods

Papers suitable for this review should be characterized by using DL methods for patients' risk prediction using EHR trajectory data. To this purpose, we searched the papers' titles, abstracts, and keywords in 4 databases (Scopus, PubMed, IEEEExplore, and the Association for Computing Machinery (ACM) Digital Library) using the following strategy on April 2022:

(("patient health trajectory") OR ("EHR trajectory") OR ("patient care pathway") OR ("patient health pathway") OR ("patient trajectory prediction") OR ("disease prediction")) AND (("deep learning") OR ("neural network")) AND (("electronic health record") OR ("electronic health data") OR ("EHR") OR ("electronic medical record") OR ("healthcare informatics"))

We limited our research only to papers published after 2016, written in English, and published in peer-reviewed journals or conference proceedings. This resulted in 130 papers. In addition, 21 related papers were added via manual search through citations and authors' suggestions. At first, we removed duplicates and review papers and then reviewed the papers' abstracts to check their relevancy. The three primary eligibility criteria were the utilization of DL methods, the historical sequential aspect of EHRs (trajectories), and performing prediction on patients' health status and diseases. After applying mentioned filters, 63 papers remained. Fig. 2a shows the PRISMA flowchart for including papers in this review.

We extracted data from the selected papers by reviewing their primary challenges and proposed solutions, suggested architecture, contributions, baseline models used for comparing results, input data and prediction target, evaluation metrics, data source, and dataset and code availability for reproducibility (supplementary material).

3. Results

3.1. Publication characteristics

Yearly publication numbers show an increasing volume of studies on the topic. Starting with only a few publications in 2016 and 2017, the number of publications increased about three times, from 6 to 17 during 2018 to 2019, and have stayed on that level through 2020–2021 (Fig. 2b). Reasons for this growth could be the increasing availability of EHR trajectory data and the continued improvement of DL methods to address the data complexities. In addition, for selected papers, the venues with the most publications were ACM SIGKDD international conference on knowledge discovery and data mining with 5 publications and Nature and BMC medical informatics and decision making with 4, Journal of biomedical informatics, EEE Journal of Biomedical and Health Informatics with 3 publications.

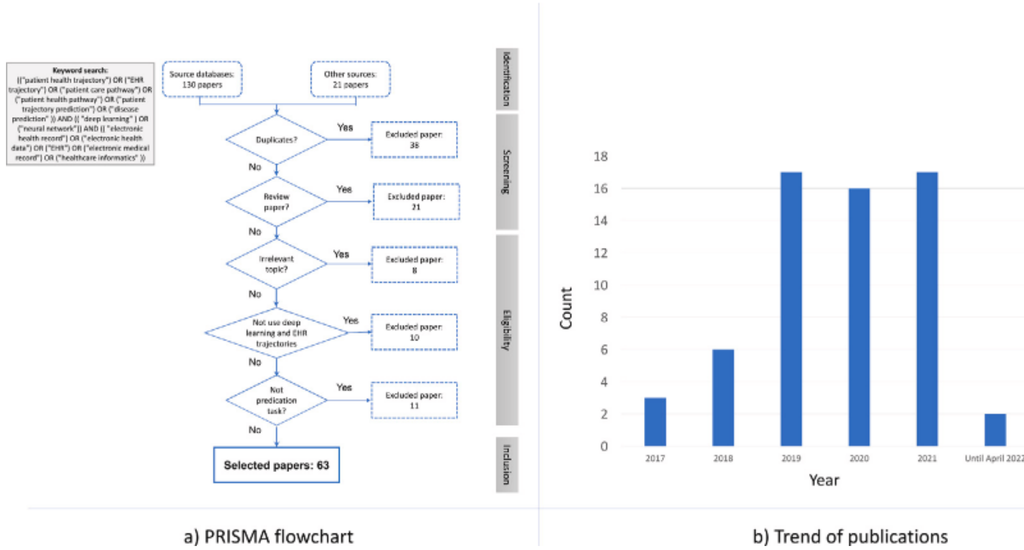

Fig. 2. An overview of Reviewed papers characteristics. (a) PRISMA flowchart selection procedure. (b) The number of publications by year for the 63 selected papers.

Table 1
The prediction task, as divided into "diseases" and "other" prediction tasks, and the studied outcomes found in reviewed papers. Only outcomes with more than four reviewed papers are shown. Additional information about DL types, EHR data sources, and other specifications of each reviewed paper can be found in Supplementary Table 1.

Prediction task	Outcome	# of papers	Papers
Diseases	Multiple diseases	19	[8–26]
	Cardiovascular disease	17	[15,18,19,27–40]
	Diabetes	7	[30,35,36,41–44]
	Kidney diseases	7	[27,29,36,45–48]
	Chronic obstructive pulmonary disease	4	[27,29,35,44]
	Alzheimer's diseases	4	[49–52]
Other	Mortality prediction	6	[10,19,26,49,53,54]
	Readmission prediction	6	[10,19,26,49,53,55]

3.2. Clinical outcomes

There are different objectives addressed among the reviewed papers. Table 1 shows the most frequent prediction tasks, divided into two types: Disease-specific predictions and other prediction tasks. Predicting all of the diagnoses in the next visit and predicting cardiovascular diseases were the most frequently studied clinical outcomes. Diabetes and kidney disease predictions were also common. For other tasks, mortality and readmission prediction were also frequently studied in the reviewed papers.

3.3. Challenges and solutions

Predicting patients' future health-related risks using EHR trajectories and DL is full of challenges, ranging from data preprocessing to understanding the final model using explainable AI methods. We divided the predictive modeling workflow using EHR trajectory data into the following steps and extracted associated challenges that reviewed papers addressed and their proposed solutions: (i) Data preprocessing, (ii) Data aggregation, (iii) Representation learning (temporal coherence and inner visits dependencies), (iv) Data insufficiency, and (v) Interpretability and explainability.

These modeling steps are not mutually exclusive and overlap in many concrete applications. In the coming sections, challenges and solutions associated with each modeling step will be presented. This review did not look into the challenges connected to model evaluation, as this has been addressed in previous work [6]. Study characteristics for the 63 reviewed papers used during the analysis, and partially reported in the following sub-sections, are summarized in Supplementary Table 1.

3.3.1. Data preprocessing

Data preprocessing is one of the most time-consuming steps in data-driven modeling [56]. For patient trajectory modeling, common preprocessing steps include format standardization, data reduction, missing data imputation, and reducing the high granularity of medical codes. Format standardization is important when EHRs are recorded in various hospital arrangements. OpenEHR [57] and FHIR (Fast Healthcare Interoperability Resources) as standard organizer platforms [58] were used on a few occasions to facilitate interoperability [46,59].

The density of information in the EHR's trajectory is not uniform in time for different patients (EHR data continuity), leading to many alternatives to define the extent of a visit (bucketing). Some studies use time windows, such as six hours [45], while others have no time window and treat visits without any grouping of information (e.g. [60]).

To cope with missing values difficulties, data imputation was utilized with different methods [48]. Bi-Directional generative adversarial networks (Bi-GAN) for its ability to impute missing elements and predict time-dependent varying-length data in a single model [61,62] and dropout technique that is equivalent to random remove of some visits and codes to make the model more robust to missing data [29] were two prominent ones. In addition, some research showed that distinguishing the presence or absence of features with a new binary column is beneficial [45,63]

Removing rare codes and reducing medical codes' granularity were usually performed to decrease data sparsity and protect patients' privacy at the preprocessing step. Reducing medical codes granularity is often a necessity which can be resolved by using just the three first digits of the ICD codes[1] [8,64,65] or grouping them with Clinical

[1] https://www.cdc.gov/nchs/icd/icd10cm.htm.

Classification Software (CCS)[2] [21,26] (see Section 3.3.2 for further details on reducing medical code granularity).

3.3.2. Data aggregation

EHRs consist of heterogeneous data sources with multiple data modalities such as medical codes, clinical text, lab values, and medical images. Developing an effective DL model strongly depends on efficient feature extraction and feature engineering techniques from all available sources and aggregating them into one data structure. Fig. 3 summarized the schematic flow of this aggregation process found in the reviewed papers.

Historical records of patients' diagnoses, medications, and procedures are present in the form of medical codes, usually with high granularity and in the format of ICD codes, ATC codes[3] and CPT codes,[4] respectively. While some researchers start with reducing granularity (e.g. [20,21]), others convert them to their original text name at the first step [66]. Different topic modeling techniques such as Latent Dirichlet Allocation (LDA) [42,64,65,67] and Non-negative Matrix Factorization (NMF) [11,12,68], name entity recognition methods [66,69] and text embedding with pre-trained networks [48,66] have been applied to extract features from unstructured text (such as clinical notes). Usually, the reduced granularity codes or their corresponding text name, along with clinical text features, were converted to multi-hot encoding. Treating each code as a word and each visits a sentence, and then encoding codes to one-hots was the other way used by some transformer-based methods [70]. Structured data such as laboratory test results, patient demographics, and vital signs are often aggregated directly. In a few studies, continuous data are discretized before aggregation [10,25,42,48,50,55,60]. The time between visits is often treated as additional structured data (see e.g. [10,40,71])

In the next step, most reviewed papers either concatenated all features to a single embedding layer (see e.g. [33,50,65]) or each modality was fed to a separate embedding layer and then concatenated or projected into a shared subspace in a later step (see e.g. [9,65,71]). Moreover, recent papers aggregated data from different modalities by training the NN on cross-modal alignment [72,73]. One article used a 2D matrix representation of all features such that similar features appeared close in the matrix geometry [43]. Another approach is to represent data as a graph and model data using graph neural network (GNN) methods. Examples include graphs by considering dependencies between diagnosis, medications, and test results in a single visit [49,53], diagnosis-patient-lab test result graphs [36], and patients-diagnosis graphs [22,74].

3.3.3. Representation learning

Since different representations can entangle or hide more or less different explanatory factors of variations behind data, the success of ML algorithms generally depends on data representation learning [2]. Developing an effective representation method that maps the aggregated health data (in specific EHRs) to a coherent space and reveals dependencies among different EHR entities will be crucial to building an accurate predictive model. Aggregated data often have high dimensionality and sparsity and contain many latent relations among different features. In the rest of this section, first, we will discuss temporal coherence and inner visit dependencies as the most important types of relations in EHRs and then analyze two major perspectives to represent EHRs.

Temporal coherence. EHR trajectory data's long- and short-term trends contain important information for many prediction tasks considered in this review. Two main approaches were used to model longitudinal dependencies: recurrent neural networks (RNNs) and attention mechanisms and, in a few cases, convolutional neural networks [12,44].

Contrary to feed-forward neural networks and traditional time series models like Markov chains which are memoryless [75], RNNs preserve the information from previous steps by self-loops. While bidirectional GRU [17,41,65,76], LSTM [10,77] and T-LSTM [65,78] were the most successful ones, simple RNNs outperform a wide range of neural networks, like intersection RNN [79], the neural Turing machine [80], memory-augmented neural network [81] and the differentiable Neural Computer [82], for predicting acute kidney injuries within 48-hours [45].

Time-attentions [83] and self-attention [84] mechanisms were used to determine the importance of each visit [27,40,41] and the relations of each entity in different visits for deriving final output [21,27,64], respectively.

Confounding interactions between disease progression and medications could also cause changes in patients' health trajectories. Predicting diseases and medicines with a multi-task network [14] and adding time interval information to the LSTM gates update equations [9] are suggested for this problem.

The distribution of visits for EHR trajectory data (denoted time irregularity) can vary to a large degree among the patients within a population. Some trajectories contain visits regularly divided in time, while others are distributed irregularly. In many prediction tasks, the time intervals between visits are used as an additional source of information. A common approach is to add this information directly as input to the models [21,27,40,71] or implicitly in the form of patients age [55,70]. Modifying the operation of the LSTM or GRU gates to capture time irregularity was also proposed [8,9,19].

Inner visits dependencies. Exploring relations among medical codes within a visit could reveal important features for the prediction tasks. The reviewed papers showed an advantage of modeling such inner visit relations between medical codes, compared to treating them as unordered bag-of-words [18,49,53]. CNNs [11,35,60] and GNNs are widely employed to handle this challenge. For example, 2D CNN showed increased performance when 1D EHR code vectors were collected into a 2D matrix, where distances within the matrix reflect correlations among the codes [43].

Graph-based structure representation methods encode inner visits dependencies by manually defining and building a graph of relations, like the ontology graph of IDC codes [85] and diagnosis-patient-lab Multipartite graph [36], or self-attentions. Self-attentions key-query-value mechanism learns a trainable 2D matrix that encodes each input pairs relations like a weighted graph's adjacency matrix [19,21,30,53,64]. Two additional good examples of combining graphs and self-attention can be found in [53] using graph convolutional transformers and in [49] with methods based on variational autoencoders.

The review has found two major approaches to represent EHR trajectories, fully supervised or pre-training & fine-tuning, see Fig. 4.

Fully supervised approach. Fully supervised approaches try to build the representation and learn the features during training of the main DL model in an end-to-end manner, making them more robust to new entities and patterns. Additionally, these methods mostly learn the features by designing appropriate neural network architecture [86] and need much more labeled data. Most researchers that used this approach built the representation only by considering co-occurrence information of medical codes and ignoring the time coherence aspects. In other words, these types of methods assume only the distributional semantics [87]. Hence codes which occur more often at the same visit are more similar. Consequently, these methods are non-contextual [88,89]. Representing

[2] https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.
[3] https://www.who.int/standards/classifications/other-classifications/the-anatomical-therapeutic-chemical-classification-system-with-defined-daily-doses.
[4] https://www.ama-assn.org/amaone/cpt-current-procedural-terminology.

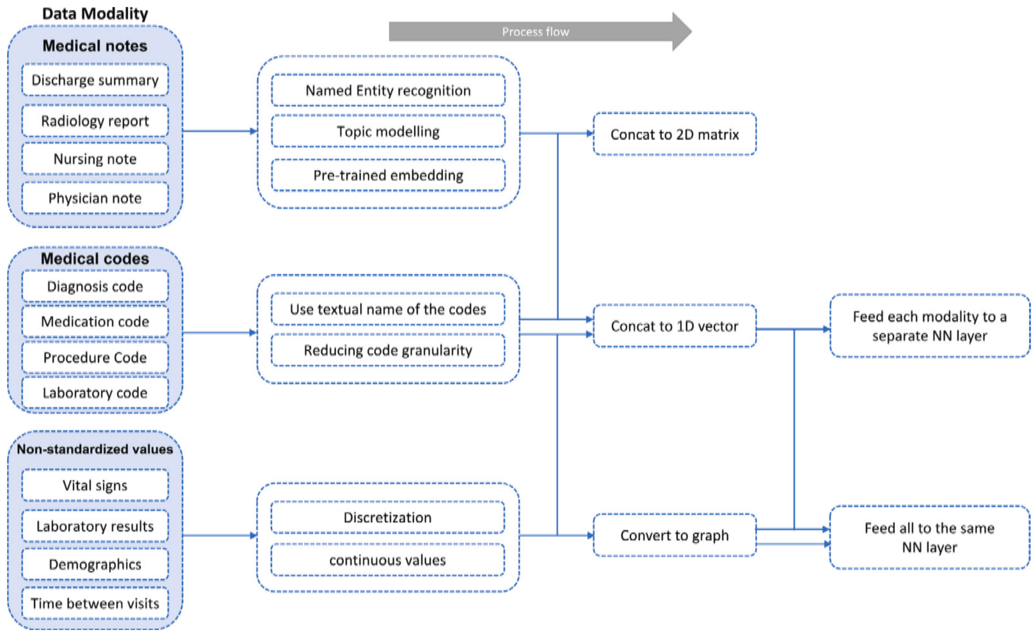

Fig. 3. A summary of different methods used by reviewed papers to aggregate information from different EHR modalities. Medical codes and clinical text integrate into multi-hot encoding and then concatenate with other non-standardized values toward the embedding step. The other approach was converting the EHR's different data types to a graph of relations.

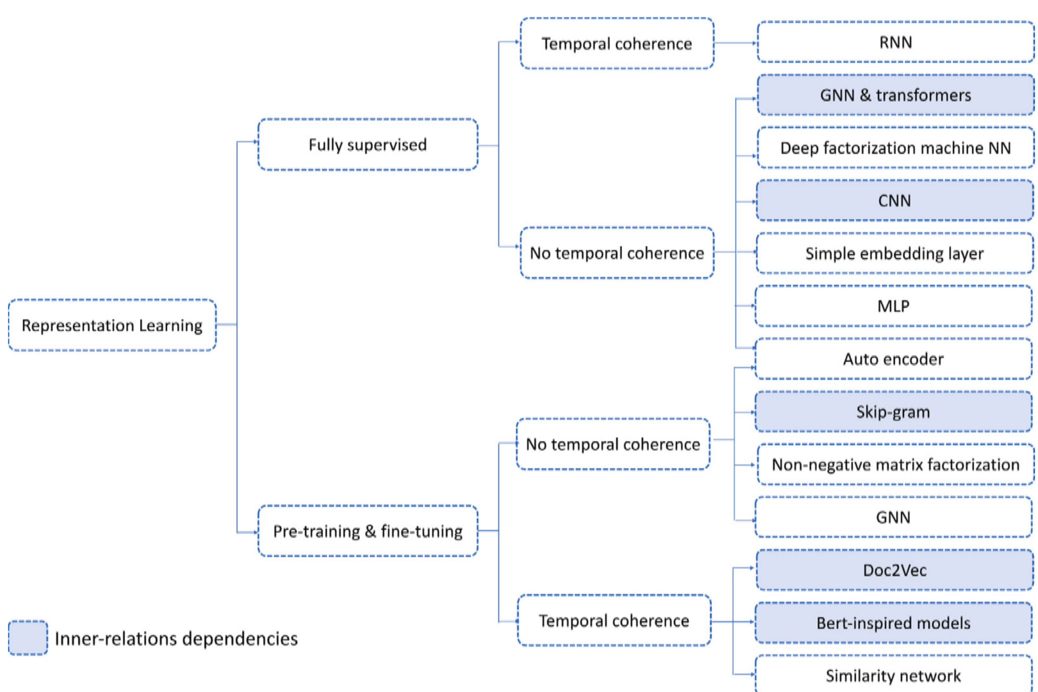

Fig. 4. Representation learning methods classify as either fully supervised or pre-training & fine-tuning. Fully supervised methods use an end-to-end DL model to represent data, while pre-training & fine-tuning methods train a separate DL model for data representation. Considering the temporal coherence variations in data in the learned representation usually need to use the contextual representation and the sequential information rather than only co-occurrence information.

the graphical relations among medical codes with GNN and transformers [53], embedding, and perceptrons layers [10,16,45,65], CNN [43] and autoencoder architecture [47] were the most common applied DL models following the fully supervised approach.

Pre-training & fine-tuning approaches. Unlike the fully supervised approaches, these methods build the representations by a separate model on a specific unsupervised or self-supervised task via training the model on designed targets and then using this representation for adapting to various downstream tasks. Once trained, these models convert raw EHR data into a new representation. Some of these methods only used the co-occurrence information for the non-contextual embedding of medical codes and clinical text, including Skip-gram [16,31,34,90], NMF [11,12,68], stacked denoising autoencoder [42,91], variational autoencoder [49], GNNs [36,74], and transformers and the self-attention mechanism [49,85].

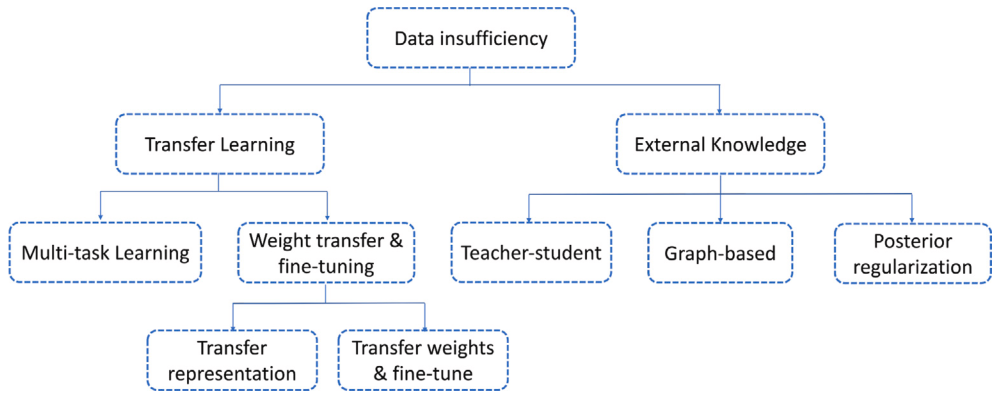

Fig. 5. To enhance the capability of DL models to deal with data insufficiency, reviewed papers used transfer learning approaches and added external knowledge sources to the model. Multi-task learning and transferring learned DL weights were used for the former, while teacher–student, the graph of knowledge, and posterior regularization were used in the latter approach.

On the other hand, some researchers incorporate time-ordered and temporal coherence information of EHRs in addition to their co-occurrences to build a more effective contextual embedding [88,89]. Doc2vec [55,92], siamese-based and triplet networks [44,46,93] and Bert-inspired models [48,64,70,94] were among such methods.

3.3.4. *Data insufficiency*

The available EHR data might not always be sufficient to build accurate and generalizable prediction models, here denoted *data insufficiency*. There are many reasons for insufficient data, such as complicated labeling processes, and rare or newly emergent diseases. Two main approaches to circumvent data insufficiency was found in the reviewed papers: transfer learning and incorporating external knowledge sources, see Fig. 5.

In the transfer learning approach, which was the most common one, researchers used pre-trained models coming from some similar prediction task, e.g. multi-task learning, or using similar datasets [51]. In the second approach, researchers incorporated external medical knowledge into the DL models, such as using medical ontologies (ICD and ATC ontology trees) to improve representation learning [15,85]. Here one can increase attention weights to the ancestors in an ICD ontology tree to deal with less frequent diseases [15]. Graph of knowledge [22,25], posterior regularization [29,95] and learning using privileged information [96] were the other solutions.

3.3.5. *Interpretability and explainability*

The ability of machine learning models to explain the decisions or to provide feedback that can improve the interpretability aspect of the model will be essential for acceptance among the intended users. Some attempts were made among the reviewed papers to provide such feedback. Gradient-weighted class activation mapping (Grad-Cam) [40,97], visualizing attention weights, and visualization of embedded feature spaces were three major approaches to cope with model explainability. The attention mechanism used in several of the reviewed papers to model the longitudinal dependencies for the trajectory data provides inherent interpretability. Hence, time attention weights were used to show the importance of each visit [14,27] and self-attentions weights could show correlations between different inputs features for the final decision [19,30,53,64]. In this context, BertViz [98] was a common tool to visualize self-attentions. Furthermore, obtained clusters of medical codes in embedded feature space were commonly interpreted as models explainability [36,49,51,70].

4. Discussion

This systematic review provides an overview of current DL prediction models for EHR trajectory data, their prediction tasks and targeted diseases, and technical challenges. We analyzed 63 selected papers that meet our criteria over 121 initial gathered papers. Given the publication statistics since 2016 and the overall growing interest in using AI in healthcare, we expect this topic to attract many researchers in the coming years.

Advances in DL methods and availability of EHR trajectories facilitate early identification and primary prevention of serious health conditions such as cardiovascular diseases, which are the principal cause of death for people of most ethnicity in the U.S. [99] and are associated with very high costs [100]. Similarly, predicting many possible diagnoses in the next visit, or particular ones such as diabetes and kidney diseases, have attracted a lot of focus, see Table 1.

4.1. *Data preprocessing*

Preparing appropriate data is the backbone of successful patient trajectory modeling. Although DL methods generally need fewer feature engineering and preprocessing steps, some data preparations must be considered. Most of the EHR data preprocessing for DL methods focuses on handling EHR data quality and continuity. They are mainly addressed by format standardization, missing data imputation, reducing data granularity, and irregular data bucketing.

Format standardization and interoperability become important during the deployment and implementation of methods and are rarely analyzed by researchers during model development. However, unlike FHIR, OpenEHR covers semantic aspects, leading to researchers trying to improve the semantic consistency of FHIR for unharmonized data using DL techniques for tokenizing and mapping to new embeddings [10,46].

In data imputation, Bidirectional RNNs can utilize the information from both past and future to recover the missed values. However, relatively poor performance may occur for a variable that has been missed for a while (i.e., the last observation happened a long time ago). On the other hand, GAN methods can achieve reasonable accuracy because of their adversarial structure but may be difficult to train [101]. Even though the dropout technique does not impute missing values, it can enhance the model's capacity to tolerate missing values and noise. Some researchers used an absent/present flag instead of imputing missing values, which can lead to the model being biased [102].

Keeping the high granularity of medical codes comes with both advantages and disadvantages. High granularity can improve the performance by utilizing all of the information [103]. On the other hand, in situations of limited training data samples, the high input sparsity can lead to the curse of dimensionality problems [104]. In addition, discretizing continuous variables of lab tests (e.g., to normal, high, and low) may be beneficial, helping the model to be more robust against data shifts arising from differences between, e.g., labs, formats, or devices.

Data continuity is the next concern tackled in preprocessing. Here a common procedure is to merge time events into visits (bucketing). Also, the time between the generated visits can be used as an additional information source. One issue may arise when the data density of different patients is very different, and the model's performance gets biased toward patients with many records. In these cases, pre-training & fine-tuning approaches (Section 3.3.3), like Bert, can help. As an example, using single-visit patients in the pre-training phase and six months data window instead of the entire patient's history for the fine-tuning was advantageous [64].

4.2. Data aggregation

Incorporating complementary information from both structured and unstructured EHR data sources was another challenge that was studied by many reviewed papers to improve the model performance [28,64,66,69,105].

For extracting features from the clinical text, NER, topic modeling, and embedding with pre-train networks techniques have been used; see Fig. 3. NER methods do not consider similarities between documents and need additional preprocessing to handle this. Topic modeling methods such as LDA and NMF use the associations between documents. However, the number of topics must be defined beforehand by the user. The built LDA model is usable for new documents, but the results are not deterministic, and multiple attempts on the same dataset can produce inconsistent outcomes. In contrast, NMF is more reliable due to its deterministic nature and computational efficiency [106,107]. Embedding textual data with pre-trained networks, whether contextualized or non-contextualized, typically requires less processing due to the effective mapping of data to more suitable spaces, making it especially useful in situations with limited data availability. Contextualized representation learning networks perform better when trained on domain-specific datasets, such as ClinicalBert, while non-contextualized representation learning methods are more effective for general text datasets [108]. Recent studies have demonstrated the superiority of ClinicalBert as a contextualized text embedding method over BioWord2Vec, a non-contextualized text embedding method [66,109,110]. Converting medical codes to their original name facilitates the utilization of mentioned text embedding techniques.

Finally, most reviewed papers used either joint or coordinated representation learning approaches to aggregate data from various sources with different modalities. In the former approach, data from different modalities are projected onto a shared subspace, while in the latter approach, the coordination between the different modalities is learned. The first approach can fuse several modalities, while it is hard for the second approach to coordinate more than two modalities. Furthermore, the second approach is more capable of learning each modality individually and preserving the exclusive characteristics of each data source [111].

4.3. Representation learning

Natural language processing (NLP) and EHR data modeling have much in common, and many of the reviewed papers tried to transfer NLP's recent achievements to the EHR domain. EHR representation learning techniques generally aim to benefit from the longitudinal information flow across a sequence of visits and the inner relations within a single visit [18]. This is similar to NLP methods that model the relations between different words within a sentence and between sentences. To enhance the robustness of the representation, many studies benefited from multi-task learning [43,45]. Furthermore, the irregularity in time was sometimes added as a unique feature of EHR trajectories by modifying the temporal coherence features extractor of the underlying method.

Fig. 4 summarizes the capability of the used methods for addressing temporal coherence and inner relations dependencies. While the fully supervised approach is more straightforward to implement, the pre-training & fine-tuning approach can deal more with data insufficiency and explainability problems. For aggregating data from different sources, projecting modalities onto a shared subspace fit better for the former approach. It is worth noting that the combination of these two approaches, Bert-inspired models plus RNNs, have achieved state-of-the-art results in some prediction tasks [30]. Future research may benefit from disentangling representation learning methods to avoid uninformative features. In addition, the mode collapse problem may be a potential threat to self-supervised methods and should be studied further.

GNNs represent one of the most interesting modeling directions and have been used in different domains [112] such as computer vision [113] and NLP [114]. Latent associations between various EHRs modalities and visits make GNNs a strong candidate for EHR representation learning. Reflecting on the physicians' decision process (inner visits relations) and integrating distinct modalities and prior medical knowledge are some of the efforts using this approach [22,36,53].

4.4. Data insufficiency

The problem of data insufficiency can happen for many different reasons, e.g. health privacy regulatory restrictions, and rare and newly emergent diseases. The reviewed papers tackled this problem by transfer learning and incorporating available rich medical external knowledge into the models [15,51,85]. Table 2 summarizes both methods' advantages and disadvantages. An interesting alternative is to develop synthetic health data. Synthea [115] is a good example of an approach to generate synthetic patient data, although it has some limitations, such as the ability to generate multi-modal continuous variables [116].

4.5. Interpretability and explainability

Trustworthy AI [117] in the healthcare domain represents many important aspects, such as explainability, confidence, privacy, and fairness. Visualizing embedded feature space and attention weights have been used frequently to make DL models explainable. These methods study the similarity of codes in embedded spaces, features' importance, and their correlations, respectively.

In EHR's trajectory prediction, we want to understand how a set of diseases and interventions interact to reach a specific outcome. Grad-Cam and time attention methods show the importance of each visit for the predicted outcome. These methods are highly sensitive to the specifics of data pre-processing and visit embedding techniques, and cannot indicate important parts of the visit or how diseases and interventions interact to increase or reduce the likelihood of the outcome. On the other hand, visualizing self-attentions weights demonstrate how diseases and interventions interact in a patient's EHR trajectory but do not easily connect to the predicted outcome, as illustrated in Fig. 6. Lastly, visualizing the embedded spaces only indicate the effectiveness of learned data representation and, as for self-attentions, does not reveal the impact on predicted outcomes.

An additional challenge arises when we want to compare models for their explainability ability, where many papers claim their models are explainable by reviewing only a few examples. Hence, benchmark datasets are necessary to evaluate and judge the power of the explainability of models. Also, many of these explainability approaches were local, while global approaches could provide researchers a deeper understanding of the model and risk health's risk factors [118]. Models' confidence level is the other requirement for trustworthiness which informs us about the uncertainty of the decision. Over/under-confident models can mislead the physicians' final decision and, consequently, their trust. Thus it seems essential to consider the model's confidence beside other evaluation measures [119,120]. Moreover, evidential DL

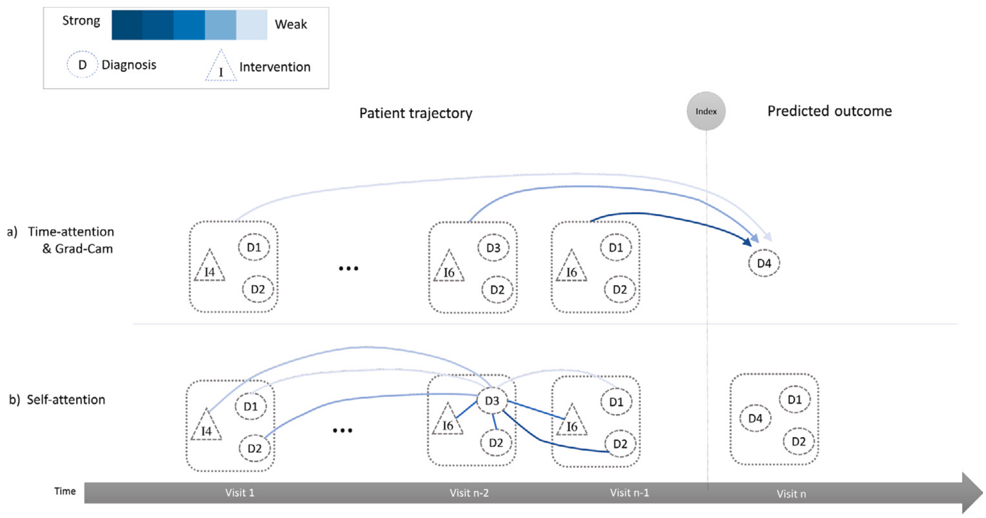

Fig. 6. Explainability of DL models for predicting diagnosis in the next visit based on their EHR trajectory. (a) Time-attention and Grad-Cam methods can highlight the importance of each visit. These methods are sensitive to the visit embedding techniques and cannot show which part of the visit (e.g., which diseases) that is most important. (b) Visualizing Self-attentions helps us to see how various diseases and interventions interact to predict the outcome but cannot show the direct importance of each entity for the predicted target.

Table 2
Advantages and disadvantages for different used methods to handle data insufficiency.

Method	Advantage	Disadvantage
Transfer learning	Can be used for a range of similar tasks with limited labeled training data samples	Usually needs lots of data and extensive networks to learn basic features well (for the general task)
	Speeds up the training process in downstream tasks by starting from better initial conditions	Needs similar enough labeled targets compared to the downstream fine-tuning task
External knowledge	Increases the performance of the model by incorporating more information	Challenging to design and implement
	More explainable	Can add human biases to the model

methods have been employed to estimate the models' decision uncertainty [121,122]. Finally, the work by Suriyakumar et al. [54] showed that protecting patients' privacy, modeling fairness for minorities, and robustness against dataset shift by differentially private learning methods [123] is still challenging.

In connection with the need to have benchmark datasets for evaluating explainability, it is important to reflect on the publicly available MIMIC datasets. As far as we know, the MIMIC [124] dataset is the only publicly available EHR trajectory dataset recorded from ICU patients. Consequently, most researchers evaluate their models on this benchmark dataset. Since MIMIC has restricted properties, there is a considerable bias risk for model development. Therefore providing more public EHR trajectory datasets is a necessity.

The analyzed challenges regarding prediction models using EHR trajectories were focused on DL methods but may also be relevant to traditional ML approaches. Some differences are worth highlighting; traditional ML methods usually require significant time and domain expertise to engineer useful features from different sources of EHR data, which to a large degree, is avoided for DL methods. On the other hand, ML methods often require smaller data sizes to build prediction models in contrast to some DL methods that need substantially large data volumes to learn good data representations. In addition, many traditional ML models are more interpretable than DL methods.

5. Conclusion

Recent promising progress in DL prediction models and the increasing availability of EHR trajectory data has led many researchers and organizations to employ these potentials with the aim of improving healthcare quality. Analyzing the reviewed papers indicated impressive advances in addressing specific EHR data challenges, including data heterogeneity, local and global dependencies, data insufficiency, and model explainability. There are, however still very few approaches that can handle all aspects of EHR trajectory data in a single model. Moreover, the combination of language models (e.g., Bert-inspired models) and multi-modal models have achieved good performance, and that can address many representation learning and explainability changes.

List of abbreviations

EHR, Electronic Health Record; DL, Deep Learning; NN, Neural Network; RNN, Recurrent Neural Network; LSTM, Long Short Term Memory; GRU, Gated Recurrent Unit; T-LSTM, Time LSTM; FHIR, Fast Healthcare Interoperability Resources; Bi- GAN, Bi-Directional Generative Adversarial Networks; CCS, Clinical Classification Software; CNN, Convolution Neural Network; MLP, Multi Layer Perceptron; NMF, Non-negative Matrix Factorization; LDA, Latent Dirichlet Allocation;NER, Named-Entity Recognition; BERT, Bidirectional Encoder Representations from Transformer; GNN, Graph Neural Network;

Declaration of competing interest

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent- licensing arrangements), or non-financial interest (such as personal or professional relationships, affinations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Funding

This study was part of the AIR Lund (Artificially Intelligent use of Registers at Lund University) research environment and received funding from the Swedish Research Council (VR; grant no. 2019-00198).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbi.2023.104430.

References

[1] Martin R Cowie, Juuso I Blomster, Lesley H Curtis, Sylvie Duclaux, Ian Ford, Fleur Fritz, Samantha Goldman, Salim Janmohamed, Jörg Kreuzer, Mark Leenay, et al., Electronic health records to facilitate clinical research, Clin. Res. Cardiol. 106 (1) (2017) 1–9.

[2] Yoshua Bengio, Aaron Courville, Pascal Vincent, Representation learning: A review and new perspectives, IEEE Trans. Pattern Anal. Mach. Intell. 35 (8) (2013) 1798–1828.

[3] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[4] Cao Xiao, Edward Choi, Jimeng 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.

[5] Md Ekramul Hossain, Arif Khan, Mohammad Ali Moni, Shahadat Uddin, Use of electronic health data for disease prediction: A comprehensive literature review, IEEE/ACM Trans. Comput. Biol. Bioinform. 18 (2) (2019) 745–758.

[6] Yuqi Si, Jingcheng Du, Zhao Li, Xiaoqian Jiang, Timothy Miller, Fei Wang, W Jim Zheng, Kirk Roberts, Deep representation learning of patient data from electronic health records (EHR): a systematic review, 2020, arXiv preprint arXiv:2010.02809.

[7] Divya Jain, Vijendra Singh, Feature selection and classification systems for chronic disease prediction: A review, Egypt. Inform. J. 19 (3) (2018) 179–189.

[8] Lu Men, Noyan Ilk, Xinlin Tang, Yuan Liu, Multi-disease prediction using LSTM recurrent neural networks, Expert Syst. Appl. 177 (2021) 114905.

[9] Trang Pham, Truyen Tran, Dinh Phung, Svetha Venkatesh, Predicting healthcare trajectories from medical records: A deep learning approach, J. Biomed. Inform. 69 (2017) 218–229.

[10] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al., Scalable and accurate deep learning with electronic health records, NPJ Digit. Med. 1 (1) (2018) 1–10.

[11] Aditya Jayasimha, Tushaar Gangavarapu, S Sowmya Kamath, Gokul S Krishnan, Deep neural learning for automated diagnostic code group prediction using unstructured nursing notes, in: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, CoDS COMAD 2020, 2020, pp. 152–160.

[12] Tushaar Gangavarapu, Gokul S Krishnan, Sowmya Kamath, Jayakumar Jeganathan, FarSight: long-term disease prediction using unstructured clinical nursing notes, IEEE Trans. Emerg. Top. Comput. (2020).

[13] Tushaar Gangavarapu, Aditya Jayasimha, Gokul S Krishnan, Sowmya Kamath, TAGS: towards automated classification of unstructured clinical nursing notes, in: International Conference on Applications of Natural Language To Information Systems, Springer, 2019, pp. 195–207.

[14] Yang An, Yakun Mao, Liang Zhang, Bo Jin, Keli Xiao, Xiaopeng Wei, Jun Yan, RAHM: Relation augmented hierarchical multi-task learning framework for reasonable medication stocking, J. Biomed. Inform. 108 (2020) 103502.

[15] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F Stewart, Jimeng Sun, GRAM: graph-based attention model for healthcare representation learning, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 787–795.

[16] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, Jimeng Sun, Doctor ai: Predicting clinical events via recurrent neural networks, in: Machine Learning for Healthcare Conference, PMLR, 2016, pp. 301–318.

[17] Jose F Rodrigues-Jr, Marco A Gutierrez, Gabriel Spadon, Bruno Brandoli, Sihem Amer-Yahia, LIG-doctor: Efficient patient trajectory prediction using bidirectional minimal gated-recurrent networks, Inform. Sci. 545 (2021) 813–827.

[18] Edward Choi, Cao Xiao, Walter F. Stewart, Jimeng Sun, Mime: Multilevel medical embedding of electronic health records for predictive healthcare, 2018, arXiv preprint arXiv:1810.09593.

[19] Chenxi Sun, Hongna Dui, Hongyan Li, Interpretable time-aware and co-occurrence-aware network for medical prediction, BMC Med. Inform. Decis. Making 21 (1) (2021) 1–12.

[20] Leopold Franz, Yash Raj Shrestha, Bibek Paudel, A deep learning pipeline for patient diagnosis prediction using electronic health records, 2020, arXiv preprint arXiv:2006.16926.

[21] Xianlong Zeng, Yunyi Feng, Soheil Moosavinasab, Deborah Lin, Simon Lin, Chang Liu, Multilevel self-attention model and its use on medical risk prediction, in: PACIFIC SYMPOSIUM on BIOCOMPUTING 2020, World Scientific, 2019, pp. 115–126.

[22] Zhenchao Sun, Hongzhi Yin, Hongxu Chen, Tong Chen, Lizhen Cui, Fan Yang, Disease prediction via graph neural networks, IEEE J. Biomed. Health Inf. 25 (3) (2020) 818–826.

[23] R. Chitravathi, G. Kanimozhi MCA, Disease prediction using snn over big data, Int. J. Innov. Technol. Explor. Eng. (IJITEE) (2019).

[24] Gokul S. Krishnan, Sowmya Kamath S, Ontology-driven text feature modeling for disease prediction using unstructured radiological notes, Comput. Sist. 23 (3) (2019) 915–922.

[25] Lin Qiu, Sruthi Gorantla, Vaibhav Rajan, Bernard C.Y. Tan, Multi-disease predictive analytics: A clinical knowledge-aware approach, ACM Trans. Manag. Inform. Syst. (TMIS) 12 (3) (2021) 1–34.

[26] Jamil Zaghir, Jose F Rodrigues-Jr, Lorraine Goeuriot, Sihem Amer-Yahia, Real-world patient trajectory prediction from clinical notes using artificial neural networks and UMLS-based extraction of concepts, J. Healthc. Inform. Res. (2021) 1–23.

[27] Junyu Luo, Muchao Ye, Cao Xiao, Fenglong Ma, Hitanet: Hierarchical time-aware attention networks for risk prediction on electronic health records, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 647–656.

[28] Robert Chen, Walter F Stewart, Jimeng Sun, Kenney Ng, Xiaowei Yan, 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 (10) (2019) e005114.

[29] Fenglong Ma, Jing Gao, Qiuling Suo, Quanzeng You, Jing Zhou, Aidong Zhang, Risk prediction on electronic health records with prior medical knowledge, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1910–1919.

[30] Laila Rasmy, Yang Xiang, Ziqian Xie, Cui Tao, Degui Zhi, Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction, NPJ Digit. Med. 4 (1) (2021) 1–13.

[31] Yang Xiang, Jun Xu, Yuqi Si, Zhiheng Li, Laila Rasmy, Yujia Zhou, Firat Tiryaki, Fang Li, Yaoyun Zhang, Yonghui Wu, et al., Time-sensitive clinical concept embeddings learned from large electronic health records, BMC Med. Inform. Decis. Making 19 (2) (2019) 139–148.

[32] P. Priyanga, Veena V. Pattankar, S. Sridevi, A hybrid recurrent neural network-logistic chaos-based whale optimization framework for heart disease prediction with electronic health records, Comput. Intell. 37 (1) (2021) 315–343.

[33] Farman Ali, Shaker El-Sappagh, SM Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, Kyung-Sup Kwak, A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion, Inf. Fusion 63 (2020) 208–222.

[34] Zhichang Zhang, Yanlong Qiu, Xiaoli Yang, Minyu Zhang, Enhanced character-level deep convolutional neural networks for cardiovascular disease prediction, BMC Med. Inform. Decis. Making 20 (3) (2020) 1–10.

[35] Tong Wang, Ping Xuan, Zonglin Liu, Tiangang Zhang, Assistant diagnosis with Chinese electronic medical records based on CNN and BiLSTM with phrase-level and word-level attentions, BMC Bioinformatics 21 (2020) 1–16.

[36] Tingyi Wanyan, Martin Kang, Marcus A Badgeley, Kipp W Johnson, Jessica K De Freitas, Fayzan F Chaudhry, Akhil Vaid, Shan Zhao, Riccardo Miotto, Girish N Nadkarni, et al., Heterogeneous graph embeddings of electronic health records improve critical care disease predictions, in: International Conference on Artificial Intelligence in Medicine, Springer, 2020, pp. 14–25.

[37] Aishwarya Mishra, Dayashankar Singh, Heart disease predictions using numerous classification techniques and dynamic LSTM model, in: 2019 International Conference on Communication and Electronics Systems, ICCES, IEEE, 2019, pp. 1136–1141.

[38] Haleh Ayatollahi, Leila Gholamhosseini, Masoud Salehi, Predicting coronary artery disease: a comparison between two data mining algorithms, BMC Public Health 19 (1) (2019) 1–9.

[39] N. Balakumar, B. Prabadevi, Modified monarch butterfly based feature selection for multi medical data classification using deep neural network, Int. J. Recent Technol. Eng. (IJRTE) 29 (2019) 31.

[40] Seunghyun Park, You Jin Kim, Jeong Whun Kim, Jin Joo Park, Borim Ryu, Jung Woo Ha, Interpretable prediction of vascular diseases from electronic health records via deep attention networks, in: 18th IEEE International Conference on Bioinformatics and Bioengineering, BIBE 2018, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 110–117.

[41] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, Jing Gao, Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1903–1911.

[42] Riccardo Miotto, Li Li, Brian A. Kidd, Joel T. Dudley, Deep patient: an unsupervised representation to predict the future of patients from the electronic health records, Sci. Rep. 6 (1) (2016) 1–10.
[43] Ruiwei Feng, Yan Cao, Xuechen Liu, Tingting Chen, Jintai Chen, Danny Z Chen, Honghao Gao, Jian Wu, ChroNet: A multi-task learning based approach for prediction of multiple chronic diseases, Multimedia Tools Appl. (2021) 1–15.
[44] Qiuling Suo, Fenglong Ma, Ye Yuan, Mengdi Huai, Weida Zhong, Jing Gao, Aidong Zhang, Deep patient similarity learning for personalized healthcare, IEEE Trans. Nanobiosci. 17 (3) (2018) 219–227.
[45] Nenad Tomašev, Xavier Glorot, Jack W Rae, Michal Zielinski, Harry Askham, Andre Saraiva, Anne Mottram, Clemens Meyer, Suman Ravuri, Ivan Protsyuk, et al., A clinically applicable approach to continuous prediction of future acute kidney injury, Nature 572 (7767) (2019) 116–119.
[46] Vagisha Gupta, Shelly Sachdeva, Subhash Bhalla, A novel deep similarity learning approach to electronic health records data, IEEE Access 8 (2020) 209278–209295.
[47] Yafeng Ren, Hao Fei, Xiaohui Liang, Donghong Ji, Ming Cheng, A hybrid neural network model for predicting kidney disease in hypertension patients based on electronic health records, BMC Med. Inf. Decis. Making 19 (2) (2019) 131–138.
[48] Dongfang Ma, Ximin Li, Shenghong Mou, Zhiyuan Cheng, Xiaoqian Yan, Ying Lu, Ruijian Yan, Shiyue Cao, Prediction of chronic kidney disease risk using multimodal data, in: 2021 the 5th International Conference on Compute and Data Analysis, 2021, pp. 20–25.
[49] Weicheng Zhu, Narges Razavian, Variationally regularized graph-based representation learning for electronic health records, in: Proceedings of the Conference on Health, Inference, and Learning, 2021, pp. 1–13.
[50] Branimir Ljubic, Shoumik Roychoudhury, Xi Hang Cao, Martin Pavlovski, Stefan Obradovic, Richard Nair, Lucas Glass, Zoran Obradovic, Influence of medical domain knowledge on deep learning for alzheimer's disease prediction, Comput. Methods Programs Biomed. 197 (2020) 105765.
[51] Xi Sheryl Zhang, Fengyi Tang, Hiroko H Dodge, Jiayu Zhou, Fei Wang, Metapred: Meta-learning for clinical risk prediction with limited patient electronic health records, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2487–2495.
[52] Muskan Kapoor, Mehak Kapoor, Rohit Shukla, Tiratha Raj Singh, Early diagnosis of alzheimer's disease using machine learning based methods, in: 2021 Thirteenth International Conference on Contemporary Computing (IC3-2021), 2021, pp. 70–76.
[53] Edward Choi, Zhen Xu, Yujia Li, Michael Dusenberry, Gerardo Flores, Emily Xue, Andrew Dai, Learning the graphical structure of electronic health records with graph convolutional transformer, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 606–613.
[54] Vinith M Suriyakumar, Nicolas Papernot, Anna Goldenberg, Marzyeh Ghassemi, Chasing your long tails: Differentially private prediction in health care settings, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021, pp. 723–734.
[55] Awais Ashfaq, Anita Sant'Anna, Markus Lingman, Sławomir Nowaczyk, Readmission prediction using deep learning on electronic health records, J. Biomed. Inform. 97 (2019) 103256.
[56] Sergio Ramírez-Gallego, Bartosz Krawczyk, Salvador García, Michał Woźniak, Francisco Herrera, A survey on data preprocessing for data stream mining: Current status and future directions, Neurocomputing 239 (2017) 39–57.
[57] Dipak Kalra, Thomas Beale, Sam Heard, The openehr foundation, Stud. Health Technol. Inform. 115 (2005) 153–173.
[58] Joshua C Mandel, David A Kreda, Kenneth D Mandl, Isaac S Kohane, Rachel B Ramoni, SMART on FHIR: a standards-based, interoperable apps platform for electronic health records, J. Am. Med. Inform. Assoc. 23 (5) (2016) 899–908.
[59] Tim Benson, Grahame Grieve, Principles of Health Interoperability: SNOMED CT, HL7 and FHIR, Springer, 2016.
[60] Chen Lin, Yuan Zhang, Julie Ivy, Muge Capan, Ryan Arnold, Jeanne M Huddleston, Min Chi, Early diagnosis and prediction of sepsis shock by combining static and dynamic information using convolutional-LSTM, in: 2018 IEEE International Conference on Healthcare Informatics, ICHI, IEEE, 2018, pp. 219–228.
[61] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative adversarial nets, Adv. Neural Inf. Process. Syst. 27 (2014).
[62] Mehak Gupta, Thao-Ly T Phan, H Timothy Bunnell, Rahmatollah Beheshti, Concurrent imputation and prediction on EHR data using bi-directional GANs: Bi-GANs for EHR imputation and prediction, in: Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2021, pp. 1–9.
[63] Narges Razavian, David Sontag, Temporal convolutional neural networks for diagnosis from lab tests, 2015, arXiv preprint arXiv:1511.07938.
[64] Yiwen Meng, William Farran Speier, Michael K Ong, Corey Arnold, Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression, IEEE J. Biomed. Health Inf. (2021).
[65] Yiwen Meng, William Speier, Michael Ong, Corey W. Arnold, HCET: Hierarchical clinical embedding with topic modeling on electronic health records for predicting future depression, IEEE J. Biomed. Health Inf. 25 (4) (2020) 1265–1272.
[66] Xiao Luo, Priyanka Gandhi, Zuoyi Zhang, Wei Shao, Zhi Han, Vasu Chandrasekaran, Vladimir Turzhitsky, Vishal Bali, Anna R Roberts, Megan Metzger, et al., Applying interpretable deep learning models to identify chronic cough patients using EHR data, Comput. Methods Programs Biomed. 210 (2021) 106395.
[67] David M. Blei, Andrew Y. Ng, Michael I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[68] Daniel D. Lee, H. Sebastian Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[69] Rebecka Weegar, Karin Sundström, Using machine learning for predicting cervical cancer from Swedish electronic health records by mining hierarchical representations, PLoS One 15 (8) (2020) e0237911.
[70] Yikuan Li, Shishir Rao, Jose Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, Gholamreza Salimi-Khorshidi, BEHRT: transformer for electronic health records, Sci. Rep. 10 (1) (2020) 1–12.
[71] Maria Hügle, Gabriel Kalweit, Thomas Hügle, Joschka Boedecker, A dynamic deep neural network for multimodal clinical data analysis, in: Explainable AI in Healthcare and Medicine, Springer, 2021, pp. 79–92.
[72] Jong Hak Moon, Hyungyung Lee, Woncheol Shin, Young-Hak Kim, Edward Choi, Multi-modal understanding and generation for medical images and text via vision-language pre-training, IEEE J. Biomed. Health Inf. 26 (12) (2022) 6070–6080.
[73] Sungjin Park, Seongsu Bae, Jiho Kim, Tackeun Kim, Edward Choi, Graph-text multi-modal pre-training for medical representation learning, in: Conference on Health, Inference, and Learning, PMLR, 2022, pp. 261–281.
[74] Haohui Lu, Shahadat Uddin, Farshid Hajati, Matloob Khushi, Mohammad Ali Moni, Predictive risk modelling in mental health issues using machine learning on graphs, in: Australasian Computer Science Week 2022, Australasian Computer Science Week 2022, ACSW 2022, 2022, pp. 168–175.
[75] Jane Lange, Latent continuous time markov chains for partially-observed multistate disease processes (Ph.D. thesis), University of Washington, 2014.
[76] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, 2014, arXiv preprint arXiv:1406.1078.
[77] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[78] Yu Zhu, Hao Li, Yikang Liao, Beidou Wang, Ziyu Guan, Haifeng Liu, Deng Cai, What to do next: Modeling user behaviors by time-lstm, in: IJCAI, Vol. 17, 2017, pp. 3602–3608.
[79] Jasmine Collins, Jascha Sohl-Dickstein, David Sussillo, Capacity and trainability in recurrent neural networks, 2016, arXiv preprint arXiv:1611.09913.
[80] Alex Graves, Greg Wayne, Ivo Danihelka, Neural turing machines, 2014, arXiv preprint arXiv:1410.5401.
[81] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, Timothy Lillicrap, Meta-learning with memory-augmented neural networks, in: International Conference on Machine Learning, PMLR, 2016, pp. 1842–1850.
[82] Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al., Hybrid computing using a neural network with dynamic external memory, Nature 538 (7626) (2016) 471–476.
[83] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, Neural machine translation by jointly learning to align and translate, 2014, arXiv preprint arXiv:1409.0473.
[84] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, 2017, pp. 5998–6008.
[85] Junyuan Shang, Tengfei Ma, Cao Xiao, Jimeng Sun, Pre-training of graph augmented transformers for medication recommendation, 2019, arXiv preprint arXiv:1906.00346.
[86] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, Graham Neubig, Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing, ACM Comput. Surv. 55 (9) (2023) 1–35.
[87] Peter D. Turney, Distributional semantics beyond words: Supervised learning of analogy and paraphrase, Trans. Assoc. Comput. Linguist. 1 (2013) 353–366.
[88] Qi Liu, Matt J. Kusner, Phil Blunsom, A survey on contextual embeddings, 2020, arXiv preprint arXiv:2003.07278.
[89] Alessio Miaschi, Felice Dell'Orletta, Contextual and non-contextual word embeddings: an in-depth linguistic investigation, in: Proceedings of the 5th Workshop on Representation Learning for NLP, 2020, pp. 110–119.
[90] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, Jeff Dean, Distributed representations of words and phrases and their compositionality, in: Advances in Neural Information Processing Systems, 2013, pp. 3111–3119.
[91] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, Léon Bottou, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (12) (2010).
[92] Quoc Le, Tomas Mikolov, Distributed representations of sentences and documents, in: International Conference on Machine Learning, PMLR, 2014, pp. 1188–1196.

[93] Sumit Chopra, Raia Hadsell, Yann LeCun, Learning a similarity metric discriminatively, with application to face verification, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, IEEE, 2005, pp. 539–546.

[94] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[95] Kuzman Ganchev, Joao Graça, Jennifer Gillenwater, Ben Taskar, Posterior regularization for structured latent variable models, J. Mach. Learn. Res. 11 (2010) 2001–2049.

[96] Fengyi Tang, Cao Xiao, Fei Wang, Jiayu Zhou, Li-wei H Lehman, Retaining privileged information for multi-task learning, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 1369–1377.

[97] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[98] Jesse Vig, A multiscale visualization of attention in the transformer model, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Association for Computational Linguistics, Florence, Italy, 2019, pp. 37–42.

[99] Centers for Disease Control, Prevention, et al., Underlying cause of death, 1999–2018, in: CDC WONDER Online Database. Atlanta, GA: Centers for Disease Control and Prevention, 2018.

[100] Salim S Virani, Alvaro Alonso, Emelia J Benjamin, Marcio S Bittencourt, Clifton W Callaway, April P Carson, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Francesca N Delling, et al., Heart disease and stroke statistics—2020 update: a report from the American heart association, Circulation 141 (9) (2020) e139–e596.

[101] Chenguang Fang, Chen Wang, Time series data imputation: A survey on deep learning approaches, 2020, arXiv preprint arXiv:2011.11347.

[102] Najmeh Abiri, Björn Linse, Patrik Edén, Mattias Ohlsson, Establishing strong imputation performance of a denoising autoencoder in a wide range of missing data problems, Neurocomputing 365 (2019) 137–146.

[103] Laila Rasmy, Firat Tiryaki, Yujia Zhou, Yang Xiang, Cui Tao, Hua Xu, Degui Zhi, Representation of EHR data for predictive modeling: a comparison between UMLS and other terminologies, J. Am. Med. Inform. Assoc. 27 (10) (2020) 1593–1599.

[104] Visar Berisha, Chelsea Krantsevich, P Richard Hahn, Shira Hahn, Gautam Dasarathy, Pavan Turaga, Julie Liss, Digital medicine and the curse of dimensionality, NPJ Digit. Med. 4 (1) (2021) 153.

[105] Andrew L Beam, Benjamin Kompa, Allen Schmaltz, Inbar Fried, Griffin Weber, Nathan Palmer, Xu Shi, Tianxi Cai, Isaac S Kohane, Clinical concept embeddings learned from massive sources of multimodal medical data, in: PACIFIC SYMPOSIUM on BIOCOMPUTING 2020, World Scientific, 2019, pp. 295–306.

[106] Roman Egger, Joanne Yu, A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts, Front. Soc. 7 (2022).

[107] Yuxin Chen, Jean-Baptiste Bordes, David Filliat, Comparison studies on active cross-situational object-word learning using non-negative matrix factorization and latent dirichlet allocation, IEEE Trans. Cogn. Dev. Syst. 10 (4) (2017) 1023–1034.

[108] Rigo E Ramos-Vargas, Israel Román-Godínez, Sulema Torres-Ramos, Comparing general and specialized word embeddings for biomedical named entity recognition, PeerJ Comput. Sci. 7 (2021) e384.

[109] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, Matthew McDermott, Publicly available clinical BERT embeddings, 2019, arXiv preprint arXiv:1904.03323.

[110] Yijia Zhang, Qingyu Chen, Zhihao Yang, Hongfei Lin, Zhiyong Lu, BioWordVec, improving biomedical word embeddings with subword information and mesh, Sci. Data 6 (1) (2019) 1–9.

[111] Wenzhong Guo, Jianwen Wang, Shiping Wang, Deep multimodal representation learning: A survey, IEEE Access 7 (2019) 63373–63394.

[112] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, S Yu Philip, A comprehensive survey on graph neural networks, IEEE Trans. Neural Netw. Learn. Syst. 32 (1) (2020) 4–24.

[113] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, Justin M Solomon, Dynamic graph cnn for learning on point clouds, Acm Trans. Graph. (Tog) 38 (5) (2019) 1–12.

[114] Diego Marcheggiani, Ivan Titov, Encoding sentences with graph convolutional networks for semantic role labeling, 2017, arXiv preprint arXiv:1703.04826.

[115] Jason Walonoski, Mark Kramer, Joseph Nichols, Andre Quina, Chris Moesel, Dylan Hall, Carlton Duffett, Kudakwashe Dube, Thomas Gallagher, Scott McLachlan, Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record, J. Am. Med. Inform. Assoc. 25 (3) (2018) 230–238.

[116] Lei Xu, Kalyan Veeramachaneni, Synthesizing tabular data using generative adversarial networks, 2018, arXiv preprint arXiv:1811.11264.

[117] Aniek F. Markus, Jan A. Kors, Peter R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies, J. Biomed. Inform. 113 (2021) 103655.

[118] Gregor Stiglic, Primoz Kocbek, Nino Fijacko, Marinka Zitnik, Katrien Verbert, Leona Cilar, Interpretability of machine learning-based prediction models in healthcare, Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 10 (5) (2020) e1379.

[119] Alireza Mehrtash, William M Wells, Clare M Tempany, Purang Abolmaesumi, Tina Kapur, Confidence calibration and predictive uncertainty estimation for deep medical image segmentation, IEEE Trans. Med. Imaging 39 (12) (2020) 3868–3878.

[120] Dennis Ulmer, Lotta Meijerink, Giovanni Cinà, Trust issues: Uncertainty estimation does not enable reliable ood detection on medical tabular data, in: Machine Learning for Health, PMLR, 2020, pp. 341–354.

[121] Alexander Amini, Wilko Schwarting, Ava Soleimany, Daniela Rus, Deep evidential regression, Adv. Neural Inf. Process. Syst. 33 (2020) 14927–14937.

[122] Awais Ashfaq, Markus Lingman, Murat Sensoy, Sławomir Nowaczyk, DEED: DEep Evidential Doctor (Ph.D. thesis), Halmstad university, 2022.

[123] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, Li Zhang, Deep learning with differential privacy, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308–318.

[124] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, Roger G Mark, MIMIC-III, a freely accessible critical care database, Sci. Data 3 (1) (2016) 1–9.