Machine Learning for Multimodal Electronic Health Records-based Research: Challenges and Perspectives

Ziyi Liu¹,², Jiaqi Zhang¹, Yongshuai Hou¹, Xinran Zhang¹,³, Ge Li, Yang Xiang¹*

¹Department of Network Intelligence, Peng Cheng Laboratory, Shenzhen, China
²School of Computer Science, Carnegie Mellon University, Pittsburgh, USA
³Mathematical and Computational Sciences, University of Toronto Mississauga, Mississauga, Canada
⁴School of Electronic and Computer Engineering, Peking University, Shenzhen, China

*Corresponding author

Abstract

Background

Electronic Health Records (EHRs) contain rich information of patients’ health history, which usually include both structured and unstructured data. There have been many studies focusing on distilling valuable information from structured data, such as disease codes, laboratory test results, and treatments. However, relying on structured data only might be insufficient in reflecting patients’ comprehensive information and such data may occasionally contain erroneous records.

Objective

With the recent advances of machine learning (ML) and deep learning (DL) techniques, an increasing number of studies seek to obtain more accurate results by incorporating unstructured free-text data as well. This paper reviews studies that use multimodal data, i.e. a combination of
structured and unstructured data, from EHRs as input for conventional ML or DL models to address the targeted tasks.

Materials and Methods

We searched in the Institute of Electrical and Electronics Engineers (IEEE) Digital Library, PubMed, and Association for Computing Machinery (ACM) Digital Library for articles related to ML-based multimodal EHR studies.

Results and Discussion

With the final 94 included studies, we focus on how data from different modalities were combined and interacted using conventional ML and DL techniques, and how these algorithms were applied in EHR-related tasks. Further, we investigate the advantages and limitations of these fusion methods and indicate future directions for ML-based multimodal EHR research.

Introduction

Electronic Health Records (EHRs) offer an efficient way to maintain patient information and are becoming more and more widely used by healthcare providers around the world. [1] Both structured and unstructured free-text data could be stored in EHRs in an organized way. Structured data can include numerical data such as lab test results, coded data such as procedure codes and diagnosis codes, categorical data such as medication lists and vital signs, and demographic information. Unstructured data such as clinical notes and discharge summaries can record more expressive and detailed descriptions of the patient’s pain level, differences from the last visit, unique characteristics of their case, and anything that the physician thinks might be helpful for diagnosis and treatment, and supplement the nuances not seen in structured data.
For example, Fox et al. validated a sample of hip fracture medical records and showed that coded records only have 66.7% sensitivity and 78.9% specificity for representing complications compared to postoperative notes. [3] By means of EHRs, physicians are able to keep track of patients’ health status and visit history to facilitate diagnosis. Meanwhile, informaticians could develop algorithms to extract patterns from the vast amount of medical data of EHRs and provide insightful suggestions to physicians’ decision-making. [4,5]

Machine learning (ML) have been used in a broad range of predictive tasks in healthcare, e.g. drug-drug interaction recognition, [6] disease progress prediction [7] and tumor detection. [8] In the past decades, traditional ML methods for EHR mining focused primarily on using structured data, which can usually be easily processed and analyzed using computer programs, and have demonstrated the effectiveness of corresponding predictive variables. [4,9] However, such structured data may suffer from missing- or erroneous-value problems, and thus cannot always guarantee consistent and accurate prediction results when used alone, as pointed out by [10,11]. Unstructured free-text data are harder to understand by programs, but are less prone to such errors, and it has been shown that textual features extracted by Natural Language Processing (NLP) pipelines can lead to positive results and serve as compensation to structured features. [12] With the development of more advanced ML and deep learning (DL) techniques, studies started to merge the two modalities, which are complementary in nature, to fully exploit the predictive potential of EHRs. Modality fusion strategies play a significant role in these studies.

There have been many reviews summarizing studies with EHR data as shown in Table 1. However, most reviews either do not discuss the fusion strategies of modalities [13–15] or focus
on conventional ML models only. \cite{12,16,17} \cite{18} discusses fusion strategies in DL, but is of structured data and imaging data. In addition, some reviews are limited to one type of disease \cite{12} or one type of task. \cite{12,15,16} In our paper, we focus exclusively on studies that use conventional ML and DL techniques with multimodal EHR data, and place our emphasis on the strategies for combining different modalities. This definition of multimodality within this paper refers to structured data and unstructured free-texts in EHRs, which is converging with heterogeneity, but different from the traditional definition that often also includes audio, image, and videos as data sources. We don’t place restrictions on certain types of diseases or tasks.

In the following sections, we review the cohort that supports multimodal analysis, tasks that involve multimodal analysis, and ML models and fusion strategies used in these studies. We also introduce a new taxonomy, interaction strategies, that is targeted to aligning the simple-form structured data to different processing granularities of unstructured data beyond fusion strategies. We thus provide a comprehensive analysis of recent progress in multimodal EHR analysis. In addition, we summarize the limitations of these methodologies and present potential future directions in this field. We expect that through this review, researchers can have a thorough view of the advancement of ML techniques in combing multimodal EHRs and a better understanding of how ML and DL models could be designed to align data from different modalities.

Table 1. Existing reviews with similar studied aspects.

| Modalities of Included Studies | ML/DL | Main Focus and Whether Modality Fusion Discussed | Task |
|--------------------------------|-------|-----------------------------------------------|------|
| AUTHORS         | TYPE               | DL TECHNIQUES AND THEIR APPLICATION TO DIFFERENT EHR-BASED CLINICAL TASKS. PRESENTS MODALITY FUSION AS FUTURE DIRECTION. | ALL CLINICAL TASKS REGARDLESS OF APPLICATION DOMAIN |
|-----------------|--------------------|-----------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| Shickel, 2018   | Primarily structured + a few with both structured and unstructured | - DL techniques and their application to different EHR-based clinical tasks. - Presents modality fusion as future direction. | All clinical tasks regardless of application domain |
| [13]            |                    |                                                                                                                |                                                     |
| Xiao, 2018      | Primarily structured + a few with both structured and unstructured | - Identifying analytical tasks and introduces commonly used models for each task. - Presents modality fusion as future direction. | All clinical tasks regardless of application domain |
| [14]            |                    |                                                                                                                |                                                     |
| Huang, 2020     | Imaging + structured data | - Fusion strategies of the modalities and implementation guidelines of multimodal models. | All clinical tasks regardless of application domain |
| [18]            |                    |                                                                                                                |                                                     |
| Si, 2021 [15]   | Multimodal + single-modal | - Resources, methods, applications and potential of EHR representation learning. - Presents modality fusion as future direction. | All representation learning studies regardless of application domain |
|                 |                    |                                                                                                                |                                                     |
| Authors         | Description                                                                 | Methodology                                | Results                                                                                           | Field                        |
|-----------------|------------------------------------------------------------------------------|--------------------------------------------|---------------------------------------------------------------------------------------------------|------------------------------|
| Sheikhalishahi  | Clinical notes.                                                              | ML + three DL articles                     | Application of NLP to clinical notes in 10 different chronic diseases groups. Does not discuss modality fusion. | Chronic diseases             |
| Zeng, 2019 [16]| Primarily structured + a few with both structured and unstructured.          | Rule-based, ML, DL                         | Application and state-of-the-art of NLP methods for computational phenotyping. Presents modality fusion as future direction. | Six tasks of computational phenotyping regardless of application domain |
| Ford, 2016 [12]| Primarily structured and unstructured + a few with unstructured only.        | Rule-based, ML                             | Information extraction methods from unstructured data and improvement over structured data only. Does not focus on modality fusion. | Case-detection for named clinical conditions |

**Search Methods**

The focus of this review is on the application of conventional ML and DL techniques to multimodal EHR-based tasks, in which multimodal is defined as including both structured and unstructured free-text data. To obtain the set of related articles, we first searched in Institute of
Electrical and Electronics Engineers (IEEE) Digital Library, PubMed, and Association for Computing Machinery (ACM) Digital Library with query terms in the following format: (“multimodal” AND (“machine learning” OR “deep learning”) AND “EHR” NOT “imaging”). We used various forms of each keyword, and included different ML and DL algorithms. The search resulted in 598 studies. Our inclusion criteria are that the research study must be using both structured and unstructured free-text EHR data in the proposed model, and that the model must use conventional ML or DL techniques. Based on these criteria, we screened 222 articles according to the title and abstract at the first step, and after full-text review, 82 articles were left. We then snowballed on other relevant review articles and included studies. Our final pool has 94 studies.

During full text review, we analyzed the studies from three perspectives: data, task, and model. For data, we focused on the dataset used by the studies, including size, language, availability and data types used (categorical, numerical, free-text, etc.). For task, we looked at where and how the proposed model is applied, and recorded elements such as task type, disease domain, and outcome. Finally, for the model, we focused on how each modality is represented, categorized the fusion strategy of the modalities, and recorded implementation details of the model.

Cohorts
The population size in each study varies greatly from hundreds to millions (of patients). For example, [19] used a dataset with only 300 patients to train a complex disease identification model, while [20] used a dataset of 6 million patients for Atopic Dermatitis identification. The dataset size used in most studies is over 1,000. Cohorts from large datasets are mostly generated
automatically according to certain criteria (i.e. phenotyping) related to the research aims [21], while the cohorts from most small datasets were often manually chosen and annotated, which could have higher accuracy and be more targeted towards their tasks. [22] Moreover, most large datasets are in English with patients from the US and UK, and datasets in other languages such as Chinese, [23] Dutch, [24] Swedish, [25,26] and Japanese [26] tend to be smaller.

The multimodal EHR datasets generally contain clinical text (e.g. clinical notes), codes (e.g. ICD codes), categorical data (e.g. medication list), and numerical data (e.g. laboratory measurements). Most datasets also include multi-visit information, although some proposed models did not specifically model such sequential information. [20,22,27] Most studies used private datasets, while others used publicly available datasets [28–30] or shared their datasets. [21,31] The most popular public dataset is the Medical Information Mart for Intensive Care III (MIMIC-III), which comprises de-identified health-related data associated with 53,423 distinct hospital admissions for adult patients. [32] Other public datasets such as Vanderbilt University Medical Center EHR, [33] Mount Sinai Data Warehouse [34] and I2B2 [35] were also used more than once in the included studies and can be used to conduct further multimodal EHR research.

**Tasks**

Traditional ML-based tasks related to EHRs can be roughly divided into *clinical information regularization* and *clinical decision-making*. Usually, unstructured information, such as case reports and nursing notes, is used as the main data source in the first type of task. The information is extracted through NLP processes and transformed into a structured format that is easier to store and analyze. [36] In comparison, for the second task, structured information such as ICD codes and patient demographics data is more often used. [37,38] EHR analysts have also
been exploring the fusion of the two types of information. For example, Payrovnaziri and Barrett extended their prior work by adding unstructured text features into previously used structured data to predict one-year mortality risk of patients with AMI. [39] Experiments demonstrate that the performance of the algorithmic model that combines both structured and unstructured information is superior to the results obtained when only one of them is used. [40]

According to our review, multimodal data that combines structured and unstructured information is currently mainly used in the clinical decision-making task, such as the prediction of different types of diseases and patient risk assessment. [41–44] Another direction, which is more related to the clinical information regularization task is patient representation learning, where researchers attempt to model multimodal EHR data in a shared semantic space for general clinical tasks. [28] The performance of these learned representations is verified on baseline tasks such as the prediction of hospital length of stay and the rate of readmission, and shows that representation learned from multimodal data has a more comprehensive generalization ability in solving real-world problems. [7] Other tasks with multimodal EHR data, such as patient trail matching and improving phenotyping algorithms, have also presented promising results. [45] These studies may provide further directions on applicable scenarios of multimodal data for EHR researchers, and the improved performance on a wide range of tasks has demonstrated the properness and explorable space of this direction.

**Machine Learning Methods**

In total, there are 63 (67.0%) papers using primarily conventional ML and 30 (31.9%) papers using primarily DL, with one paper not explicitly introducing methods or models used but
mentioning they used ML. As discussed above, most tasks with multimodal EHR deal with classification (clinical decision-making) or patient representation problems. For classification tasks such as disease prediction and risk assessment, relevant studies used mainly conventional ML models, such as Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVMs), and Naive Bayes (NB). For instance, Chen et al. took the UMLS concepts extracted from clinical notes and billing codes as input and incorporated active learning to SVM-based phenotyping algorithms. Slightly different from traditional approaches, Zhao and Weng designed a weighted Bayesian Network Inference (BNI) model where they combined structured EHR data with a prior probability calculated from free-text PubMed abstracts of the expert selected variables for pancreatic cancer prediction.

Despite the popularity of conventional ML for classification tasks, DL models were greatly explored over the recent years, and are being applied to both classification tasks and representation learning tasks. CNN has been shown to be good at extracting locality-invariant features, which can be useful for identifying key concepts in clinical notes. On the other hand, RNN, and especially its variant Long Short-Term Memory Networks (LSTM), is well known for discovering sequential dependencies, which is often used to model clinical notes (text as sequence) and temporal information in EHR. More recently, Transformer-based pre-trained models are also adapted to the medical domain. Huang et al. developed ClinicalBERT, which pre-trained the BERT model on clinical notes in MIMIC-III. This model has also been used by many multimodal EHR studies to build embeddings for free-text data. Among the included studies, Autoencoders and word embeddings are the most commonly used techniques to compute the representation. Other approaches that applied
recent DL techniques also exist. Lee et al. built a harmonized representation space for patients, medical concepts, and medical events by modeling the time-variant patients nodes with LSTM, and time-invariant concept and event nodes with Graph Convolutional Networks (GCN), which fused multimodal inputs. [60] After obtaining these representations, they can be used as input to both conventional ML and DL models for downstream tasks.

To compare, most conventional ML methods take each feature, either from structured or unstructured EHRs, independently owing to their shallow structures. For example, it is hard for RF and SVM to capture the sequential dependencies conveyed in unstructured data, and thus some deeper semantic information might be missed during modeling. On the other hand, DL models have a stronger capacity to model different levels of dependencies between feature dimensions so that complex interactions between modalities might be built. Specifically, DL models are superior in modeling free-text, e.g. using hierarchical and sequential structures, which is more conformable with the natural structures of free-text.

Fusion Strategies

The remarkable characteristic of multimodal EHR research is that useful information can be conveyed by both structured and unstructured data, and thus features extracted from the two modalities should be fused effectively. As a reference, there exists a generally accepted taxonomy by previous studies which categorizes multimodal data fusion in ML into three types: early fusion, joint fusion, and late fusion, as shown in Figure 1. [18] We firstly tried to align each study to this taxonomy, shown in Table 2.
Figure 1. Early/Joint/Late Fusion by Huang et al., 2020 [18]

Table 2. Different fusion types for multimodal EHR data that can be aligned to.

| Fusion Type | Definition | Commonly Used By | Examples |
|-------------|------------|------------------|----------|
| Early Fusion | Features are extracted from modalities by statistical methods, existing NLP tools, word embedding models, or other DL models, and combined prior to passing to the model | ML, DL | statistical methods [26,48,61] Existing NLP tools [21,46,62] Word embedding models [39,63,64] Other DL models [22,59,65] |
Joint Fusion
Combines the modalities at intermediary layers of the neural network, and losses can be propagated back to the feature extraction phase to dynamically update the feature weights

Late Fusion
Separate models are used for each modality, and the final decision leverages the decision of each individual model by some ensemble strategies

Interaction Strategies
Unlike imaging and speech data which usually need to be preprocessed separately and transformed into feature vectors before fusion, free-text data is more flexible in the sense that its fusion with structured data can start at a much earlier stage and be in a more direct way because of the similarity in modality. Each word in free-text is semantically meaningful by itself, whereas extracting such segments from an image or a speech signal sequence often requires additional handling. Different levels of text processing could lead to features capturing different granularities of semantics. For example, raw free-text data could be used directly to compute word embeddings that can reflect relationships between words, and concepts and topics could be extracted from free-text to represent its meaning at the sentence/document level.
Therefore, as free-text data has higher granularities, the three fusion strategies alone might be insufficient to fully represent the various levels of interactions between multi modalities. For example, using raw words, word embeddings or named entities as input to a neural network model can all be aligned to early fusion, but they represent quite distinct combination scenarios where different complementary and interdependent information could be carried and different model structures could be applied. We design a novel and finer-grained taxonomy, interaction strategies, to better represent how multimodal data are fused and help us understand how different models deal with multimodal data according to their operational mechanism. We define five categories based on the degree to which free-text data is processed before combining with structured data (Figure 2). We also discuss pros and cons of each level of interaction in Table 3.
Figure 2. Different interaction levels for multidata fusion. We also demonstrate the comparison of different interaction levels in the last sub-figure. The legend is similar to that in Figure 1.

Table 3. Pros and cons of the five interaction strategies.

| Interaction Strategy | Pros | Cons |
|----------------------|------|------|
| Data-level           | Could lead to the discovery of mutual and representative features which are present in more than one modality. | Data from the two modalities are considered as coming from only one modality, e.g. consider a code as a word. Resulting representation is more likely to capture shared rather than complementary information from the modalities. |
| Concept-level        | Can be easily implemented and is computationally efficient. Extracted textual features provide a clean and concise representation of the unstructured data, and thus increases the model’s interpretability and is | Could lead to information loss, such as contextual and temporal information in free-text data. Concatenation is a relatively naive way to combine modalities that does not exploit full relationships between the modalities. |
| Layer-level | Description | Benefits | Disadvantages |
|-------------|-------------|----------|---------------|
| Embedding-level | Word embeddings and neural networks can greatly preserve information in free-text and well-formed patterns and features in structured data could be discovered than using frequency counts. Dependency on external tools is reduced. | The boundary between modalities is blurred with concatenated single-modal representations, which might cause the model to rely on dominating modality instead of learning the cross-modality interaction. Decreased model interpretability. | |
| Intermediate Layer-level | Retains all advantages of embedding-level interaction and provides greater interaction flexibility. Easier to learn complementary information from all modalities and dynamically update feature weights accordingly by incorporating the combination step within the model. | Low quality features from one modality could be harmful and add additional noise to the model. | It is debatable whether one or two layers could fully capture the complex relationships between the modalities depending on at which layer the interaction happens. |
| Decision-level | Performance of each modality can be evaluated separately, which can be useful when the number of modalities | Models in this interaction strategy are not exposed to multimodal data at all, and thus may not be able to | |
is large, or there are multichannel representations, and decisions need to be made regarding which ones to keep. effectively use the rich information contained in the multimodal data.

Data-level Interaction

In data-level interaction, free-text data remains unprocessed and is combined with structured data in its raw form. Methods using this strategy usually take as input both structured and unstructured raw data and learn a shared embedding space by word embedding models. Bai et al. proposed JointSkip-gram, which is based on the Word2Vec learning schema. [71] It takes raw clinical notes and medical codes as input, and for each code, learns to predict all other codes and words in the same visit, while for each word, learns to predict its neighboring words and all codes. The resulting representation is able to capture not only similarities within the same modality, but also cross different modalities through shared weights.

Concept-level Interaction

Concept-level interaction refers to extracting features firstly by statistics or established packages from each modality, and combining them using simple ways such as concatenation. As the free-text features can be either raw data or clinical concepts, e.g. named entities, we call this interaction concept-level. This is a typical type of early fusion, and appears most often in conventional ML studies, before the advance of DL. Feature values of structured data are often represented as the frequencies of events, while for free-text data, they are extracted using Term Frequency-Inverse Document Frequency (TF-IDF), bag of words, topic modeling methods such as Latent Dirichlet Allocation (LDA), [72] or established NLP pipelines such as clinical Text
Analysis and Knowledge Extraction System (cTAKES), MetaMap, KnowledgeMap, and Health Information Text Extraction (HITEx) is a typical example of concept-level interaction with machine learning, where features such as disease concepts, lab results, and medications were extracted from both codified structured data and clinical notes using HITEx. Then the frequencies of events were passed to penalized logistic regression to predict rheumatoid arthritis.

**Embedding-level Interaction**

Embedding-level interaction shows another way of integrating free-text, in which representation of free-text data is obtained by passing the raw text through a pre-trained word embedding or neural network model. Structured data can be represented similarly as in concept-level interaction or also be passed through a neural network model as embedding-like representations. The embedding vectors from each modality are then concatenated or summed up as the input for the final output model. This type of interaction is used by many DL studies according to our review due to the ability of embeddings to capture more comprehensive information from the raw data. Beeksma et al. embedded unstructured free-text data using Word2Vec, and concatenated with structured codes, laboratory, and medication data for life expectancy prediction using LSTM. Darabi et al. developed two separate modules for code and text representation. Code module is composed of a skip-gram model followed by a Transformer encoder and the text model uses BioBERT followed by a Bi-GRU. Finally, patient representation is a concatenation of the code and text representations.

**Intermediate Layer-level Interaction**
In intermediate layer-level interaction, the interaction of modalities happens between the embedding and the output layer. Intermediate layer-level interaction is used only by DL studies, but less popular than embedding-level interaction. Liu et al. proposed CNN and Bi-LSTM based models for chronic disease prediction, which first represented free-text data using word embedding pre-trained on medical data, and passed to the base model. [31] Structured data were concatenated to the layer before the last fully connected layer. Xu et al. adopted a novel approach with Memory Networks, where they used clinical notes encoded by hierarchical LSTM as the query, and fed structured clinical sequences into the memory. [42] Then the information similar to the query (unstructured data) is extracted from the memory (structured data) and combined with the query to form a representation for acute kidney injury predictions.

**Decision-level Interaction**

Decision-level interaction is the same as late fusion, in which separate models are built for each modality, and modalities only interact with each other after the output layer. Shin et al. used TF-IDF and topic modeling to represent two types of clinical notes, and built separate logistic regression models for each feature set. [69] The final division is the average of all model outputs. This level of interaction is not commonly used in multimodal EHR studies, especially those with DL.

In addition to these semantic granularity-based interactions, there also are interaction scenarios that are specific to EHRs and from other perspectives. For example, the multimodal data can be fused in either the visit level, i.e. the interaction happens within each visit, or the patient level,
i.e. the interaction happens after all the visits are encoded. In these scenarios, either of the above-mentioned interactions on different levels could be leveraged.

**Limitations**

The acquisition of multimodal EHR data is one of the most significant limitations. Since different EHR systems have different standards, not all systems store data as having both “structured” and “unstructured” parts, especially in different countries, where data might be all structured, all semi-structured, or all unstructured. This makes multimodal EHR datasets more scarce than single-modal data. And as mentioned in the cohort section, most current multimodal EHR data are in English, it could introduce distribution biases on population that might lead to algorithmic biases. Finally, unlike structured data, unstructured data is more sensitive to the missing data problem. It is possible to impute a missing blood pressure value, but difficult to impute a missing clinical note.

Another limitation is on the modality interaction techniques and models used. Although the advance of DL propels new techniques such as embedding and intermediate layer level interaction, the “interaction” strategy in these studies is still simple and possibly insufficient to capture complex interactions between different modalities. In addition, base models in these studies are often restricted to traditional ML and simple DL models. This could be due to the fact that processing and extracting useful information from free-text data by traditional informaticians, statisticians or physicians remains a challenge. There have been studies using more complex models such as Graph Convolutional Networks to wrap up data from multimodalities,[60] but the number remains limited, and the exploration is still incipient.
Future Directions

Include more data modalities

Although we focus on structured and unstructured text as our multimodal data, other modalities such as images and videos could also be integrated using the aforementioned fusion or interaction techniques. [18,78] Integrating more data modalities could be helpful if each data modality contains incomplete but complementary information, which is frequently seen in EHRs. For example, Patient history, imaging diagnostic report and associated lab test values could naturally be used together with medical imaging for making clinical decisions. [78] However, whether to include data from more sources is also restricted by the data availability and sharing policies in many cases. Furthermore, in some scenarios, prior knowledge, including terminologies, ontologies and knowledge bases, have also been proven useful when performing ML-aided clinical prediction. [52,68] It can serve as another data modality and the relevant techniques deserve a deeper investigation.

Explore more ways of interaction between modalities

There are still many limitations in the fusion methods research of multimodal data. This issue is also a major direction that needs attention in future research. Most of the current research uses methods of direct concatenation to process multimodal data, either from data or embedding level. This method is easy to implement, but the representation gap of different types of data is ignored during this process. For example, embeddings trained from unstructured data are quite different from those trained from structured data in the semantic level, e.g. the semantic granularity is different. [79] Future research on the fusion strategy of multimodal data should find more ways
to align the representation level of different types of data. Current works have tried different solutions, but a consensus on a widely accepted strategy for multimodal interaction, especially in the fusion method of structured and unstructured data, has not been retained yet. The interaction strategies of multimodal methods using different types of data, such as interaction of image data [80], structured data and text data, may be worth exploration and adoption among each other.

*How and how much pretrained models can help?*

Pretrained models (PTM) have been proved helpful in many natural language processing [81] as well as multimodal classification tasks (e.g. integrating image and text). [82] Using PTMs, comprehensive contextual information for both different modalities and their interactions could be learned in a self-supervised way [83] or with minimum supervisions [84]. For EHR mining, however, the explorations on multimodal data that contain both structured and unstructured text are limited. [71] is one representative study in this paradigm but it only applies shallow embedding techniques and fails to encoding deeper contextual information. [85] is a pioneer study that pretrains structured and unstructured data together using BERT, but it only uses the diagnosis information as the structured modality. Adopting more techniques from the general multimodal pretraining (e.g. VideoBERT [86]) or treating it as a machine translation problem [87] might be ways for further exploration.

*Enable more flexible data acquisition and sharing*

As mentioned above, the availability of multimodal data is one of the limitations that hinder the development of corresponding research. There are many factors (e.g. gender, ethnicity, political issues, weather, region, environmental humidity) that could influence the research directions or
even clinical decisions, it might not be sufficient to rely on the few public datasets to make conclusive clinical claims to the general population. Therefore, it is necessary to encourage more institutions or hospitals to share flexible data to include a more general and broader range of population to facilitate clinical research. In ML, federated learning (FL) [88] has the capability to collect patients’ data with the safety of patients' privacy protection and data security from multicenter. It might be leveraged to collect multimodal EHR features from multicenters to train a large-scale model in a discrete distribution without collecting patients’ EHR directly.

Conclusion

This review summarizes current advances in multimodal EHR studies using both structured and unstructured free-text EHR data with conventional ML and DL techniques. We proposed interaction strategies, a new taxonomy targeted to the combination of structured and free-text EHR data. Our finding suggests that there is a growing interest in multimodal EHR, but most studies combine the modalities with relatively simple strategies, which despite being shown to be effective, might not fully exploit the rich information embedded in these modalities. We acknowledge that there are still limitations with this review such as the limitation in the coverage of queried papers and included algorithms. As this is a fast-growing field and new models are constantly being developed, there might exist studies that fall outside of our definition of strategies or use a combination of these strategies. Nonetheless, we believe that the development of this field will give rise to more comprehensive EHR analysis and will be of great support to the clinical decision-making process.

Funding
There is no funding resource to be disclosed.

**Author contributions**

YX and ZYL conceived the study. ZYL designed search method, analyzed model aspect of included papers, and drafted most parts of the manuscript. JQZ participated in study design, analyzed task aspect of included papers, and drafted Task section. YSH analyzed cohort aspect of included papers, and drafted Cohort section. XRZ contributed to future direction, YX supervised the research, provided feedback to the proposed taxonomy, contributed to future direction, and critically revised the manuscript. GL participated in manuscript review. All authors provided feedback and approved the final version of the manuscript.

**Conflict of interest statement**

The authors have no competing interests to declare.

**References**

1. Jha AK, DesRoches CM, Campbell EG, et al. Use of Electronic Health Records in U.S. Hospitals. *http://dx.doi.org/101056/NEJMSa0900592* 2009;360:1628–38. doi:10.1056/NEJMSA0900592

2. ST R, JC D, H X, et al. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association : JAMIA* 2011;18:181–6. doi:10.1136/JAMIA.2010.007237

3. KM F, M R, WG H, et al. Accuracy of medical records in hip fracture. *Journal of the American Geriatrics Society* 1998;46:745–50. doi:10.1111/J.1532-5415.1998.TB03810.X
4 Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics* 2012;13:395–405. doi:10.1038/nrg3208

5 Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International Journal of Medical Informatics* 2008;77:291–304. doi:10.1016/J.IJMEDIINF.2007.09.001

6 Segura-Bedmar I, Martínez P, de Pablo-Sánchez C. Using a shallow linguistic kernel for drug–drug interaction extraction. *Journal of Biomedical Informatics* 2011;44:789–804. doi:10.1016/J.JBI.2011.04.005

7 Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *npj Digital Medicine* 2018;1:1–10. doi:10.1038/s41746-018-0029-1

8 Wang Z, Yu G, Kang Y, et al. Breast tumor detection in digital mammography based on extreme learning machine. *Neurocomputing* 2014;128:175–84. doi:10.1016/J.NEUCOM.2013.05.053

9 Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *Journal of the American Medical Informatics Association : JAMIA* 2014;21:221. doi:10.1136/AMIAJNL-2013-001935

10 WR H, MG W, PJ E, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical care* 2013;51. doi:10.1097/MLR.0B013E31829B1DBD
11 E B-D, AD W, Y Y, et al. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Medical care* 2005;**43**:480–5. doi:10.1097/01.MLR.0000160417.39497.A9

12 Ford E, Carroll JA, Smith HE, et al. Extracting information from the text of electronic medical records to improve case detection: A systematic review. *Journal of the American Medical Informatics Association* 2016;**23**:1007–15. doi:10.1093/JAMIA/OCV180

13 B S, PJ T, A B, et al. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. *IEEE journal of biomedical and health informatics* 2018;**22**:1589–604. doi:10.1109/JBHI.2017.2767063

14 Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. *Journal of the American Medical Informatics Association* 2018;**25**:1419–28. doi:10.1093/JAMIA/OCY068

15 Si Y, Du J, Li Z, et al. Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. *Journal of Biomedical Informatics* 2021;**115**. doi:10.1016/J.JBI.2020.103671

16 Zeng Z, Deng Y, Li X, et al. Natural Language Processing for EHR-Based Computational Phenotyping. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2019;**16**:139–53. doi:10.1109/TCBB.2018.2849968

17 Sheikhalishahi S, Miotto R, Dudley JT, et al. Natural language processing of clinical notes on chronic diseases: Systematic review. *JMIR Medical Informatics* 2019;**7**. doi:10.2196/12239
18 Huang S-C, Pareek A, Seyyedi S, et al. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *npj Digital Medicine* 2020;3:1–9. doi:10.1038/s41746-020-00341-z

19 Murray SG, Avati A, Schmajuk G, et al. Automated and flexible identification of complex disease: Building a model for systemic lupus erythematosus using noisy labeling. *Journal of the American Medical Informatics Association* 2019;26:61–5. doi:10.1093/JAMIA/OCY154

20 Ananthakrishnan AN, Cai T, Savova G, et al. Improving case definition of Crohn’s disease and ulcerative colitis in electronic medical records using natural language processing: A novel informatics approach. *Inflammatory Bowel Diseases* 2013;19:1411–20. doi:10.1097/MIB.0B013E31828133FD

21 Teixeira PL, Wei W-Q, Cronin RM, et al. Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. *Journal of the American Medical Informatics Association: JAMIA* 2017;24:162. doi:10.1093/JAMIA/OCW071

22 Mugisha C, Paik I. Pneumonia Outcome Prediction Using Structured And Unstructured Data From EHR. In: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2020. 2640–6. doi:10.1109/BIBM49941.2020.9312987

23 Jiang H, Li Y, Zeng X, et al. Exploring fever of unknown origin intelligent diagnosis based on clinical data: Model development and validation. *JMIR Medical Informatics* 2020;8. doi:10.2196/24375

24 Afzal Z, Engelkes M, Verhamme KMC, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. *Pharmacoepidemiology and Drug Safety* 2013;22:826–33. doi:10.1002/PDS.3438
25 Henriksson A, Zhao J, Dalianis H, et al. Ensembles of randomized trees using diverse distributed representations of clinical events. *BMC Medical Informatics and Decision Making* 2016;16. doi:10.1186/S12911-016-0309-0

26 Makino M, Yoshimoto R, Ono M, et al. Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Scientific Reports* 2019;9:1 2019;9:1–9. doi:10.1038/s41598-019-48263-5

27 Tou H, Yao L, Wei Z. Automatic infection detection based on electronic medical records. In: 2017 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2017. 1684–7. doi:10.1109/BIBM.2017.8217913

28 Zhang D, Yin C, Zeng J, et al. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Medical Informatics and Decision Making* 2020;20. doi:10.1186/S12911-020-01297-6

29 Hu SY, Santus E, Forsyth AW, et al. Can machine learning improve patient selection for cardiac resynchronization therapy? *PLoS ONE* 2019;14. doi:10.1371/JOURNAL.PONE.0222397

30 Landi I, Glicksberg BS, Lee HC, et al. Deep representation learning of electronic health records to unlock patient stratification at scale. *npj Digital Medicine* 2020;3. doi:10.1038/S41746-020-0301-Z

31 Liu J, Zhang Z, Razavian N. Deep EHR: Chronic Disease Prediction Using Medical Notes. 2018;:440–64.http://proceedings.mlr.press/v85/liu18b.html (accessed 13 Jul 2021).

32 AE J, TJ P, L S, et al. MIMIC-III, a freely accessible critical care database. *Scientific data* 2016;3. doi:10.1038/SDATA.2016.35
33 DM R, JM P, MA B, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clinical pharmacology and therapeutics 2008;84:362–9. doi:10.1038/CLPT.2008.89

34 Miotto R, Li L, Kidd BA, et al. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Scientific Reports 2016;6. doi:10.1038/SREP26094

35 Uzuner Ö, Stubbs A. Practical Applications for Natural Language Processing in Clinical Research: The 2014 i2b2/UTHealth Shared Tasks. Journal of biomedical informatics 2015;58:S1. doi:10.1016/J.JBI.2015.10.007

36 Spasic I, Nenadic G. Clinical text data in machine learning: Systematic review. JMIR Medical Informatics 2020;8. doi:10.2196/17984

37 Gultepe E, Green JP, Nguyen H, et al. From vital signs to clinical outcomes for patients with sepsis: A machine learning basis for a clinical decision support system. Journal of the American Medical Informatics Association 2014;21:315–25.

38 Zhao J, Henriksson A, Asker L, et al. Predictive modeling of structured electronic health records for adverse drug event detection. BMC Medical Informatics and Decision Making 2015;15. doi:10.1186/1472-6947-15-S4-S1

39 Payrovnaziri SN, Barrett LA, Bis D, et al. Enhancing prediction models for one-year mortality in patients with acute myocardial infarction and post myocardial infarction syndrome. Studies in Health Technology and Informatics 2019;264:273–7. doi:10.3233/SHTI190226
40 AP N, J Y, L R, et al. Assessing occurrence of hypoglycemia and its severity from electronic health records of patients with type 2 diabetes mellitus. *Diabetes research and clinical practice* 2016;121:192–203. doi:10.1016/J.DIABRES.2016.09.012

41 Meng Y, Speier W, Ong M, et al. HCET: Hierarchical clinical embedding with topic modeling on electronic health records for predicting future depression. *IEEE Journal of Biomedical and Health Informatics* 2021;25:1265–72. doi:10.1109/JBHI.2020.3004072

42 Xu Z, Chou J, Zhang XS, et al. Identifying sub-phenotypes of acute kidney injury using structured and unstructured electronic health record data with memory networks. *Journal of Biomedical Informatics* 2020;102. doi:10.1016/J.JBI.2019.103361

43 Amrollahi F, Shashikumar SP, Razmi F, et al. Contextual Embeddings from Clinical Notes Improves Prediction of Sepsis. *AMIA Annual Symposium Proceedings* 2020;2020:197./pmc/articles/PMC8075484/ (accessed 13 Jul 2021).

44 Zeng Z, Yao L, Roy A, et al. Identifying Breast Cancer Distant Recurrences from Electronic Health Records Using Machine Learning. *Journal of Healthcare Informatics Research* 2019;3:283–99. doi:10.1007/S41666-019-00046-3

45 Zhang X, Xiao C, Glass LM, et al. DeepEnroll: Patient-Trial Matching with Deep Embedding and Entailment Prediction. In: *Proceedings of The Web Conference 2020*. New York, NY, USA: : Association for Computing Machinery 2020. 1029–37. doi:10.1145/3366423.3380181

46 Xu H, Fu Z, Shah A, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. *AMIA . Annual Symposium proceedings / AMIA Symposium AMIA Symposium* 2011;2011:1564–72.
47 Gustafson E, Pacheco J, Wehbe F, et al. A Machine Learning Algorithm for Identifying Atopic Dermatitis in Adults from Electronic Health Records. *Proceedings - 2017 IEEE International Conference on Healthcare Informatics, ICHI 2017* 2017:83–90. doi:10.1109/ICHI.2017.31

48 Scheurwegs E, Luyckx K, Luyten L, et al. Data integration of structured and unstructured sources for assigning clinical codes to patient stays. *Journal of the American Medical Informatics Association* 2016;23:e11–9. doi:10.1093/JAMIA/OCV115

49 Lin C, Karlson EW, Canhao H, et al. Automatic Prediction of Rheumatoid Arthritis Disease Activity from the Electronic Medical Records. *PLoS ONE* 2013;8.

50 Fodeh SJ, Li T, Jarad H, et al. Classification of Patients with Coronary Microvascular Dysfunction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2020;17:704–11. doi:10.1109/TCBB.2019.2914442

51 Chen Y, Carroll RJ, Hinz ERMP, et al. Applying active learning to high-throughput phenotyping algorithms for electronic health records data. *Journal of the American Medical Informatics Association* 2013;20.

52 Zhao D, Weng C. Combining PubMed Knowledge and EHR Data to Develop a Weighted Bayesian Network for Pancreatic Cancer Prediction. *Journal of biomedical informatics* 2011;44:859. doi:10.1016/J.JBI.2011.05.004

53 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature 2015* 521:7553 2015;521:436–44. doi:10.1038/nature14539

54 Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space. *1st International Conference on Learning Representations, ICLR 2013* -
Workshop Track Proceedings Published Online First: 16 January 2013.https://arxiv.org/abs/1301.3781v3 (accessed 13 Aug 2021).

55 Devlin J, Chang M-W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference 2018;1:4171–86.* https://arxiv.org/abs/1810.04805v2 (accessed 13 Aug 2021).

56 Yin W, Kann K, Yu M, et al. Comparative Study of CNN and RNN for Natural Language Processing. Published Online First: 7 February 2017.https://arxiv.org/abs/1702.01923v1 (accessed 13 Jul 2021).

57 Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Computation* 1997;9:1735–80. doi:10.1162/NECO.1997.9.8.1735

58 Huang K, Altosaar J, Ranganath R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. Published Online First: 10 April 2019.https://arxiv.org/abs/1904.05342v3 (accessed 13 Jul 2021).

59 Darabi S, Kachuee M, Fazeli S, et al. TAPER: Time-aware patient EHR representation. *IEEE Journal of Biomedical and Health Informatics* 2020;24:3268–75. doi:10.1109/JBHI.2020.2984931

60 Lee D, Jiang X, Yu H. Harmonized representation learning on dynamic EHR graphs. *Journal of Biomedical Informatics* 2020;106. doi:10.1016/J.JBI.2020.103426

61 Wang L, Sha L, Lakin JR, et al. Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients With Dementia for Earlier
Palliative Care Interventions. *JAMA Network Open* 2019;2:e196972–e196972.
doi:10.1001/JAMANetworkOpen.2019.6972

62 Liao KP, Cai T, Gainer V, et al. Electronic medical records for discovery research in rheumatoid arthritis. *Arthritis Care and Research* 2010;62:1120–7.

63 Henriksson A, Zhao J, Boström H, et al. Modeling electronic health records in ensembles of semantic spaces for adverse drug event detection. In: *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2015. 343–50.
doi:10.1109/BIBM.2015.7359705

64 Beeksma M, Verberne S, van den Bosch A, et al. Predicting life expectancy with a long short-term memory recurrent neural network using electronic medical records. *BMC Medical Informatics and Decision Making* 2019;19. doi:10.1186/S12911-019-0775-2

65 Liu R, Greenstein JL, Sarma S v, et al. Natural Language Processing of Clinical Notes for Improved Early Prediction of Septic Shock in the ICU. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2019. 6103–8. doi:10.1109/EMBC.2019.8857819

66 Bardak B, Tan M. Improving clinical outcome predictions using convolution over medical entities with multimodal learning. *Artificial Intelligence in Medicine* 2021;117.
doi:10.1016/J.ARTMED.2021.102112

67 Bagheri A, Groenhof TKJ, Veldhuis WB, et al. Multimodal Learning for Cardiovascular Risk Prediction Using EHR Data. In: *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*. New York, NY, USA: : Association for Computing Machinery 2020.
doi:10.1145/3388440.3414924
Xie X, Xiong Y, Yu PS, et al. EHR Coding with Multi-Scale Feature Attention and Structured Knowledge Graph Propagation. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. New York, NY, USA: Association for Computing Machinery 2019. 649–58. doi:10.1145/3357384.3357897

Shin B, Hogan J, Adams AB, et al. Multimodal Ensemble Approach to Incorporate Various Types of Clinical Notes for Predicting Readmission. In: *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. 2019. 1–4. doi:10.1109/BHI.2019.8834640

Xu K, Lam M, Pang J, et al. Multimodal Machine Learning for Automated ICD Coding. *Proceedings of Machine Learning Research*. 2019;**106**:197–215. http://proceedings.mlr.press/v106/xu19a.html (accessed 13 Jul 2021).

Bai T, Chanda AK, Egleston BL, et al. EHR phenotyping via jointly embedding medical concepts and words into a unified vector space. *BMC Medical Informatics and Decision Making* 2018;**18**. doi:10.1186/S12911-018-0672-0

Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *Journal of Machine Learning Research* 2003;**3**:993–1022. doi:10.1016/b978-0-12-411519-4.00006-9

Savova GK, Masanz JJ, Ogren P v, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association : JAMIA* 2010;**17**:507. doi:10.1136/JAMIA.2009.001560

AR A. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proceedings AMIA Symposium* 2001;:17–21. https://pubmed.ncbi.nlm.nih.gov/11825149/ (accessed 13 Aug 2021).
75 Denny JC, Irani PR, Wehbe FH, et al. The KnowledgeMap Project: Development of a Concept-Based Medical School Curriculum Database. AMIA Annual Symposium Proceedings 2003;2003:195/pmc/articles/PMC1480333/ (accessed 13 Aug 2021).

76 Zeng QT, Goryachev S, Weiss S, et al. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. BMC Medical Informatics and Decision Making 2006 6:1 2006;6:1–9. doi:10.1186/1472-6947-6-30

77 Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. Advances in Neural Information Processing Systems 2017;2017-December:5999–6009.https://arxiv.org/abs/1706.03762v5 (accessed 4 Sep 2021).

78 Venugopalan J, Tong L, Hassanzadeh HR, et al. Multimodal deep learning models for early detection of Alzheimer’s disease stage. Scientific Reports 2021 11:1 2021;11:1–13. doi:10.1038/s41598-020-74399-w

79 Rasmy L, Xiang Y, Xie Z, et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. npj Digital Medicine 2021 4:1 2021;4:1–13. doi:10.1038/s41746-021-00455-y

80 Zhang X, Chou J, Liang J, et al. Data-Driven Subtyping of Parkinson’s Disease Using Longitudinal Clinical Records: A Cohort Study. Scientific Reports 2019 9:1 2019;9:1–12. doi:10.1038/s41598-018-37545-z

81 Qiu X, Sun T, Xu Y, et al. Pre-trained models for natural language processing: A survey. Science China Technological Sciences 2020 63:10 2020;63:1872–97. doi:10.1007/S11431-020-1647-3
Li Y, Wang H, Luo Y. A Comparison of Pre-trained Vision-and-Language Models for Multimodal Representation Learning across Medical Images and Reports. *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020* 2020;:1999–2004. doi:10.1109/BIBM49941.2020.9313289

Akbari H, Yuan L, Qian R, et al. VATT: Transformers for Multimodal Self-Supervised Learning from Raw Video, Audio and Text. Published Online First: 22 April 2021.https://arxiv.org/abs/2104.11178v1 (accessed 14 Jul 2021).

Bouritsas G, Koutras P, Zlatintsi A, et al. Multimodal Visual Concept Learning With Weakly Supervised Techniques. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018;:4914–23.

Shang J, Ma T, Xiao C, et al. Pre-training of Graph Augmented Transformers for Medication Recommendation. *IJCAI International Joint Conference on Artificial Intelligence* 2019;2019-August:5953–9.https://arxiv.org/abs/1906.00346v2 (accessed 12 Sep 2021).

Sun C, Myers A, Vondrick C, et al. VideoBERT: A Joint Model for Video and Language Representation Learning. *Proceedings of the IEEE International Conference on Computer Vision* 2019;2019-October:7463–72.https://arxiv.org/abs/1904.01766v2 (accessed 14 Jul 2021).

Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* Published Online First: 1 September 2014.https://arxiv.org/abs/1409.0473v7 (accessed 4 Sep 2021).
Li T, Sahu AK, Talwalkar A, et al. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine* 2020;37:50–60. doi:10.1109/MSP.2020.2975749

Ponte P, Melko RG. Kernel methods for interpretable machine learning of order parameters. *Physical Review B* 2017;96. doi:10.1103/physrevb.96.205146

Y X, H J, Y Z, et al. Asthma Exacerbation Prediction and Risk Factor Analysis Based on a Time-Sensitive, Attentive Neural Network: Retrospective Cohort Study. *Journal of Medical Internet Research* 2020;22. doi:10.2196/16981

Murdoch WJ, Liu PJ, Yu B. Beyond Word Importance: Contextual Decomposition to Extract Interactions from LSTMs. 2018.

Liang J, Bai B, Cao Y, et al. Adversarial Infidelity Learning for Model Interpretation. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2020;20:286–96. doi:10.1145/3394486.3403071

F C, R R, GF G. Unintended Consequences of Machine Learning in Medicine. *JAMA* 2017;318:517–8. doi:10.1001/JAMA.2017.7797