CarePre: An Intelligent Clinical Decision Assistance System

ZHUOCHEN JIN and SHUYUAN CUI, Tongji University
SHUNAN GUO, East China Normal University
DAVID GOTZ, University of North Carolina at Chapel Hill
JIMENG SUN, Georgia Institute of Technology
NAN CAO, Tongji University

Clinical decision support systems are widely used to assist with medical decision making. However, clinical decision support systems typically require manually curated rules and other data that are difficult to maintain and keep up to date. Recent systems leverage advanced deep learning techniques and electronic health records to provide a more timely and precise result. Many of these techniques have been developed with a common focus on predicting upcoming medical events. However, although the prediction results from these approaches are promising, their value is limited by their lack of interpretability. To address this challenge, we introduce CarePre, an intelligent clinical decision assistance system. The system extends a state-of-the-art deep learning model to predict upcoming diagnosis events for a focal patient based on his or her historical medical records. The system includes an interactive framework together with intuitive visualizations designed to support diagnosis, treatment outcome analysis, and the interpretation of the analysis results. We demonstrate the effectiveness and usefulness of the CarePre system by reporting results from a quantities evaluation of the prediction algorithm, two case studies, and interviews with senior physicians and pulmonologists.

CCS Concepts: • **Computing methodologies** → **Reasoning about belief and knowledge**; • **Human-centered computing** → *Visual analytics*;

Additional Key Words and Phrases: Personal health records, neural networks, reasoning about belief and knowledge, visual analytics, user interface design

ACM Reference format:
Zhuochen Jin, Shuyuan Cui, Shunan Guo, David Gotz, Jimeng Sun, and Nan Cao. 2020. CarePre: An Intelligent Clinical Decision Assistance System. *ACM Trans. Comput. Healthcare* 1, 1, Article 6 (February 2020), 20 pages.
https://doi.org/10.1145/3344258

This research was supported in part by NSFC grants 61602306, Fundamental Research Funds for the Central Universities, and the National Grants for the Thousand Young Talents in China.
Authors' addresses: Z. Jin, S. Cui, and N. Cao (corresponding author), Intelligent Big Data Visualization Lab, Tongji University; emails: {zcjin.idvx, sycui.idvx, nan.cao}@gmail.com; S. Guo, East China Normal University; email: g.shunan@gmail.com; D. Gotz, University of North Carolina at Chappel Hill; email: gotz@unc.edu; J. Sun, Georgia Institute of Technology; email: jsun@cc.gatech.edu.

2637-8051/2020/02-ART6 $15.00
https://doi.org/10.1145/3344258

1 INTRODUCTION

Medical decision making is fraught with uncertainty. It involves not only deciding what disease a patient may have but also which treatments to choose from a set of possible alternatives [35]. Motivated by these challenges, clinical decision support systems (CDSS) have been increasingly used in recent years.

CDSS are computer-based systems that are integrated into the clinical workflow to help physicians determine which questions to ask, which tests to order, and which procedures to perform [39, 43]. However, typical CDSS require manually curated knowledge bases that are difficult to maintain and keep up to date, thus limiting their usage in real-world clinical scenarios [32].

The rapid development of machine learning techniques and the increasing availability of electronic health records (EHRs) has stimulated great interest in harnessing EHR data to help drive CDSS. It is widely believed that high-quality EHR data in the context of CDSS has potential to reduce errors and provide more precise results [5, 7, 8, 26]. To this end, many techniques have been developed to extract meaningful insights from EHR data with a common focus on the prediction of upcoming medical events (e.g., a diagnosis or treatment) [13, 20]. In particular, a series of deep learning–based prediction models [11, 12, 50] have successfully demonstrated that high accuracy predictions are possible. However, the utility of these methods is greatly limited by their lack of interpretability. The ideal intelligent medical event prediction system must provide results that are both accurate and interpretable through a user-friendly interface.

However, achieving both accuracy and interpretability is challenging, as they are often achieved via contradictory design decisions. The highest accuracy prediction is often obtained when using more complex prediction methods, whereas simpler models with lower accuracy are often more interpretable [6]. Attempts have been made to improve the interpretability of more complex prediction models [12, 50]. However, these approaches are still too complex for users with little or no technical training, such as medical doctors.

To address the preceding issues, we introduce CarePre, an intelligent clinical decision assistance system. CarePre predicts the risks of a patient being diagnosed in the future with certain diseases based on his or her historical EHRs. The system extends a state-of-the-art deep prediction model that is specifically designed for predicting medical events, and employs intuitive visualization techniques to help interpret the prediction results without reducing the complexity of the underlying model. In particular, CarePre supports interpretation by (1) framing the prediction results in the context of a group of similar patients and (2) analyzing the factors that influence the prediction results to help physicians make more informed clinical decisions. The contributions of this article include the following:

- *System design*: We introduce a comprehensive clinical decision assistance system for predicting a patient's risk of future diagnoses for certain diseases and estimating the outcome of different treatments based on the patient's EHRs. The system design is guided by results from a pilot study with two senior physicians.
- *Exploratory analysis*: We propose an interactive framework that supports detailed exploration for both (1) interpretation of prediction results in the context of historical and similar medical records, and (2) analysis of potential treatment outcomes.
- *Evaluation*: We evaluate the system via a quantitative evaluation of the algorithm; two case studies using real-world medical records of a group of cardiac and respiratory patients, respectively; and interviews with seven senior physicians. We describe the case studies and feedback collected from the interviews. These results provide evidence regarding the usefulness of the system.

2 RELATED WORKS

In this section, we provide an overview of previous research that is most relevant to our work, including (1) CDSS, (2) prediction models in medicine, and (3) visualization of EHRs.

2.1 Clinical Decision Support Systems

Existing CDSS can primarily be summarized into two major types: knowledge based and non-knowledge based [3]. Knowledge-based systems, which are most commonly used, typically organize knowledge about diseases and the associations of symptoms in the form of if-then rules. For example, Dayan et al. [15] introduced the traumatic brain injury (TBI) prediction rules in a CDSS to foresee risks of TBI. Laleci et al. [28] utilized a guideline-based CDSS to help manage the personal care plans of elders. Rodriguez-Borja et al. [42] introduced a "send & hold" system, utilizing clinical decision support rules to reduce the avoidable vitamin testing.

Non-knowledge-based systems are usually developed based on machine learning techniques that can automatically learn the associations between symptoms and diseases from EHR data [3]. It has been shown that EHR data not only helps improve the precision of analysis results [5, 18] but also greatly improves the robustness of a CDSS due to the availability of rich and diverse EHR data gathered during the daily clinical encounters [11, 49]. When compared to knowledge-based systems, these systems greatly reduce the human efforts required to manually build and update a large knowledge database [32]. However, these systems typically suffer from a lack of interpretablity of the analysis results [32] and a lack of user-friendly interfaces to facilitate efficient results inspection [4, 8]. CarePre leverages the advantages of machine learning techniques and EHRs while also providing a comprehensive visualization-based design to support result inspection and interpretation.

2.2 Prediction Models in Medicine

Prediction models have played an increasingly important role in the medical domain, for both diagnosis and prognosis [45]. Recent research has often focused on leveraging deep learning techniques to make predictions more accurate and precise [52]. These techniques have been used to support public health analysis [9, 54, 55], medical research [14, 47, 53], and clinical practice [1, 23, 29]. Some deep learning techniques have been developed to assess risk for specific conditions, such as the diagnosis of heart disease [1, 40, 51], cancer [10, 14, 53], and mental health [2, 22, 41].

Most relevant to our work are the studies that also focused on predicting upcoming medical events (e.g., a future diagnosis or treatment) based on EHRs. Examples in this area include Jagannatha and Yu [23], who used EHR data to train a bidirectional recurrent neural network (RNN) for medication and disease prediction. Choi et al. [11] developed Doctor AI, a generic RNN model that use historical EHR to predict the clinical events and the time to the next visit. Following this work, Choi et al. [12] further introduced Retain, a state-of-the-art, high-accuracy prediction model that was specifically designed to predict "signal" events (i.e, heart failure) based on EHR data. Our system extends this model to predict multiple events, as motivated by our design requirements.

Interpreting results from prediction models is a recognized challenge, and it is especially difficult for models that leverage deep learning. Existing interpretation techniques can largely be categorized into two categories: (1) global model analysis, which employs visualization techniques to represent the internal structure of a deep learning model [30, 31, 46], and (2) instance-based analysis, which monitors changes to results in response to changes in model input [25, 27, 36]. CarePre adopts the instance-based analysis approach via a set of interactive visualization designs that allow users to adjust/delete/add medical events within a patients' historical medical records and explore their impact on the prediction result.

2.3 Visual Analysis of EHRs

Many visual analysis systems have been developed for representing and analyzing EHRs. Most of these summarize a large set of EHR data into a flow-based representation that reveals the frequent patterns of medical event sequences [33, 38] and the outcomes yielded by different care plans [37, 48]. However, these techniques are typically challenged by event sequences that have various length and contain large numbers of event types. These real-world properties of medical data can often lead to cluttered and less meaningful visualizations when sequences vary dramatically. To overcome this limitation, Gotz and Stavropoulos [19] introduced DecisionFlow,

in which sequences with different length and large numbers of even types are visualized based on several key events. This hides the complexity introduced by other non-key event types. Guo et al. [21] introduced ET^2, in which the sequences are aligned based on dynamic time wrapping and segmented into stages shown with more details to help illustrate the progression of a disease in context of a care plan. Du et al. [16, 17] introduced visual analysis systems to predict upcoming events or recommend the next procedure by summarizing a set of similar event sequences without using any prediction model, thus producing results with limited accuracy. Kwon et al. [27] tried to interpret the prediction results of upcoming events based on medical sequence data. However, the system is more designed for artificial intelligence scientists rather than physicians. CarePre considers the requirements of physicians and leverages many of the advances contributed by these visualization techniques, and supports multiple visualization-based views to help physicians explore and interpret prediction results for better clinical decisions.

3 PILOT STUDY

Our pilot study followed a multi-session design and involved two senior physicians ($\boldsymbol{P_{1-2}}$) with more than 15 years of clinical experience in two hospitals in a major city in China. In particular, both physicians were associate directors of internal medical department and have sufficient diagnose experience. The goal of the pilot study was to determine detailed requirements to guide the subsequent system design.

Session 1: Initial requirements. We conducted a 1-hour interview with each of the two physicians, during which we discussed the challenges they encountered when making diagnosis. Both physicians believe it is important to refer to a patient's medical history when making a diagnosis. For example, $\boldsymbol{P_1}$ mentioned, "Historical medical records can help doctors make a more comprehensive judgement on the potential illness of a patient, which is especially important for analyzing complications in chronic diseases." However, their diagnosis and treatment decisions are often made based on the patient's *current* symptoms and lab test results. As mentioned by $\boldsymbol{P_2}$, "We usually need to diagnose large number of patients every day. Time is limited for us to fully review their medical history." They expressed a desire for a tool that could automatically provide relevant diagnosis information based on a patient's medical history to avoid mistakes (*diagnosis-supporting requirements*). Moreover, they wished the tool to be able to estimate potential outcomes when the doctors were to make treatment plans (*prognosis-supporting requirements*). "When the patient's condition is complex and we have multiple treatment plans in mind, predicting the outcome for each treatment plan is especially helpful in making decisions," $E1$ explained, "It is also a more intuitive way for our patients to understand our decisions."

Session 2: Prototyping and refinement. Based on results of Session 1, an interactive design prototype was developed using Figma[1] by a professional designer (a co-author of this article). The prototype was demonstrated to the two physicians to further refine the initial requirements to the following detailed requirements:

- *R1* *Support for predicting the risks of potential diseases*: According to the physician interviewees, predicting the risks of potential diseases based on historical record is useful in verifying their initial diagnoses. For example, $\boldsymbol{P_1}$ stated, "I'll confirm the diagnoses with high risk and make further investigation to assess those with low risk." $\boldsymbol{P_2}$ commented, "I'll be more confident in making decisions if the prediction is aligned with my judgement." Thus, the system should be able to automatically assess a patient's historical medical record to predict the risks of a set of potential diseases identified by the physicians.
- *R2* *Support for exploring predictions under the context of historical sequences*: Both physicians mentioned that it is important to review predictions with historical records. $\boldsymbol{P_1}$ explained that "Historical events can provide evidence for us to determine whether the prediction is trustworthy," and $\boldsymbol{P_2}$ felt that "Reviewing history sequence can help us better understand the predictions." Hence, the system should be able to illustrate the

[1] https://www.figma.com/.

prediction results within the context of the patient's historical medial record to facilitate data exploration and result interpretation.

R3 *Support for easy comparison between the focal patient and similar patients*: The physicians expressed their need for leveraging the past experience in diagnosing and treating patients with similar clinical pathways. As mentioned by P_2, "Knowing how patients with similar symptoms [were] diagnosed and treated before will provide us with more guidance in making [a] diagnosis and treatment plans." Therefore, the system should be able to identify patients with similar historical medical records and summarize their clinical pathways for easy comparison.

R4 *Support for identifying contributing factors in predictions*: One common demand proposed by the physicians is to make the prediction result easier to interpret. As P_2 explained, "We sometimes do not trust the predictions generated by machines, especially when [they disagree] with our judgement. It would make a difference if the model [could] tell us the reason [for] the prediction." To this end, the system should be able to identify and communicate the key factors that have a large impact on the prediction result.

R5 *Support for exploring the outcomes of possible treatment plans*: When discussing the requirements for supporting the prognosis, P_1 expressed a desire for simulating possible treatment plans: "We usually have multiple treatment plans in mind. It would be great if the tool [could] help us predict and compare the outcomes after [making] those treatment plans." P_2 felt the same way and added, "It can also help us understand the effect of the drugs on the disease through the exploration." Accordingly, the system should help physicians explore changes to possible treatment plans and illustrate the impact of those changes on the predicted outcome.

The entire prototyping stage took place over 2 months, during which regular meetings with domain experts were held. The prototype was iteratively refined to incorporate clinicians' comments and new requirements. This process resulted in a series of eight different design versions, culminating in the final design described in the next section.

4 CAREPRE SYSTEM

Following the aforementioned requirements, we designed the CarePre system. This section provides an overview of the system design and its key algorithms.

4.1 System Overview

CarePre is an intelligent system designed to assist physicians or other health professionals when making decisions related to diagnosis and prognosis. The key functionalities of CarePre are (1) prediction of a patient's risk of being diagnosed with certain disease and (2) estimation of the most influential treatments, as determined based on a patient's historical EHRs. In particular, the system converts raw EHR data for a large number of patients into sequences of medical events. Based on those sequences, the system predicts the future occurrence probabilities of several given diagnosis events for a focal patient.

Figure 1 illustrates the CarePre user interface. It consists of 10 views, many of which utilize data visualization techniques to facilitate an intuitive data representation and interpretation. These views can be categorized into three classes based on their functionality: (1) diagnosis-supporting views (Figure 1(a)–(c)), (2) similar patients' retrieval and comparison views (Figure 1(d)–(g)), and (3) treatment outcome analysis views (Figure 1(b), (h), and (i)).

These views support the system's interaction pipeline as shown in Figure 2. The pipeline includes three main steps. First, the physician makes an initial diagnosis using his or her own knowledge and experiences, based on a focal patient's current symptoms and lab tests. A series of potential diagnoses are automatically identified from this stage or manually entered into the system (Figure 1(b-1)), and CarePre is able to estimate the patient's risks in terms of being diagnosed in the future with the diseases given his or her historical medical records (Figure 1(b-2)). Second, the doctor can explore the details of the historical medical records (Figure 1(b-3)), and issue a query

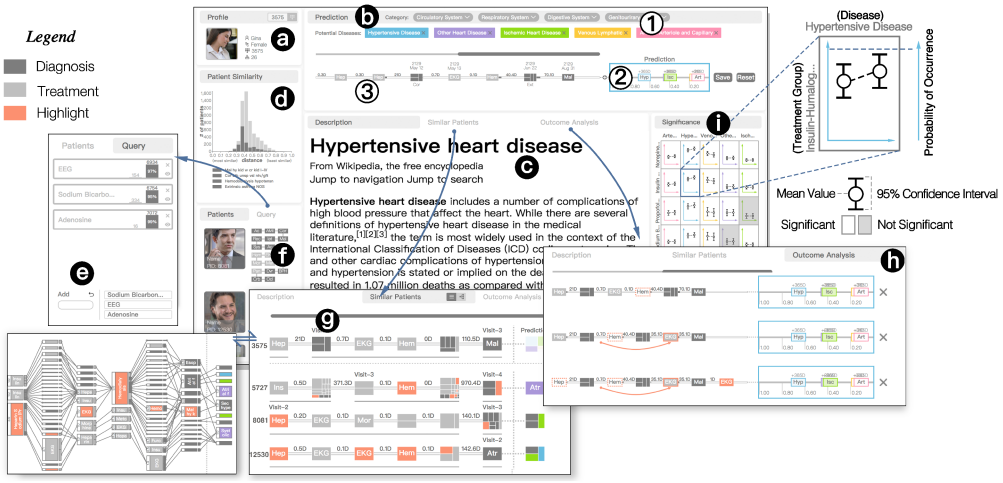

Fig. 1. The CarePre system contains nine interactively coordinated views, including a profile view showing the personal information of a patient (a); a prediction view illustrating the prediction results, as well as the historical medical records of the patient (b); a description view providing the detailed description of a disease selected from the prediction view (c); a patient similarity view measuring the similarity between the focal patient and the archived patients (d); a query view supporting a key-event-based query capability to select specific patients (e); a patient list showing similar patients retrieved from part (d) or part (e) (f); a similar patients view comparing the prediction results to the outcomes of similar patients (g); an outcome analysis view allowing the examination of the outcomes of different treatment plans (h); and a significance view showing the influence of treatments on the risks of diseases (i).

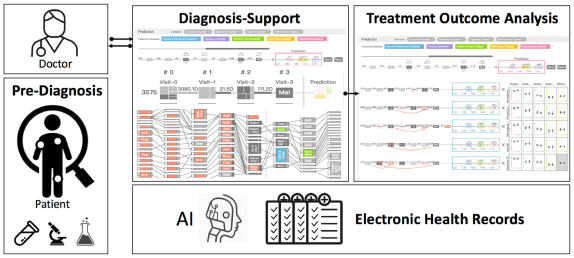

Fig. 2. The interaction pipeline of the CarePre system consists of three steps, including a pre-diagnosis step in which physicians initially diagnosis a focal patient according to his or her symptoms and lab tests (1); a diagnosis support step in which the system automatically estimates the risk of each of the potential diseases determined in the previous step, and in which physicians can verify the results by comparison to the medical records of a set of similar patients (2); and the treatment outcome analysis step in which physicians can compare and evaluate the expected outcomes of different treatment plans (3).

to fetch a set of similar patients to help contextualize and interpret the prediction results (Figure 1(d), (f), and (g)). Third, the doctor can examine alternative treatment plans by examining and comparing the expected outcomes of each as estimated by the system (Figure 1(h) and (i)).

4.2 Usage Scenario

To help understand how a doctor could use CarePre to support diagnosis and prognosis, let us consider the following usage scenario. Imagine a doctor, John, who is diagnosing a patient and trying to make a treatment

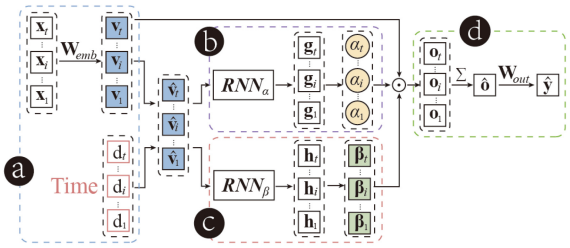

Fig. 3. The structure of Retain. Taking a sequence $x_1, \ldots, x_t$ as input, the model can predict the distribution of possible diagnosis in three steps: the embedding step (a), the attention steps (b, c), and the final prediction step (d).

plan. He has several suspected diagnoses in mind, and wants to use CarePre to confirm his thought and help him make treatment decisions.

After loading the patient's records into the system, John first takes a look at the prediction view (Figure 1(b-1)) to check whether the predicted diagnoses are aligned with his initial judgment. To better understand the prediction result, he clicks on each of the predicted diagnoses to investigate the impact of each historical event on the prediction. To find evidence from a larger population, John retrieves a group of similar patients from the patient similarity view (Figure 1(d)) to observe their clinical pathways. He then switches to the similar patients' view to observe the aggregated clinical paths of the retrieved patients. John clicks and highlights the paths with suspected diagnoses as outcome to see whether events along the path are in accordance with the influential events in the history of the focal patient. With the aid of both evidence in the historical records of similar patients and patterns captured by the prediction model, John is able to confirm the diagnoses of the analyzed patient, so he continues to make treatment plans.

After reviewing the significance view (Figure 1(h)), John is able to identify treatments that have a large influence on the risk of the predicted diseases. Combining his own domain knowledge, he has several optional treatment plans. John tries out different treatment plans by appending the treatments to the end of the patient's history and inspects the change of risk in predicted diseases. He saves the result of different treatment plans in the outcome analysis view (Figure 1(h)), and after a comprehensive comparison, he finally decides on a optimal treatment plan.

4.3 Diagnosis Support

CarePre system assists a typical diagnosis procedure by predicting the next medical event given an event sequence representing a patient's medical record (R1). More specifically, the system predicts the next diagnostic event (i.e., the potential diseases a patient may have) based on the patient's previous diagnoses and treatments. The prediction results and the patient's historical medical data are illustrated in an interactive visualization to facilitate data and result exploration (R2).

4.3.1 Prediction Model. To predict the next diagnosis, we developed a deep learning model to predict the likelihood of occurrence for a set of potential diseases selected by physicians according to a patient's historical medical record. The model extends the design used in Retain [12] to predict multiple medical events at the same time. Our model was trained using a subset of the MIMIC dataset [24], which contains the EHRs of 46,521 patients. Prior to training, the data were cleaned by removing rarely occurring and irrelevant event types.

Figure 3 illustrates the structure of the model. The model predicts subsequent medical events based on an input event sequence $[x_1, \ldots, x_t]$, where x_i is a multi-hot vector that captures the occurrences of events at each time point. Given this input, an embedding layer is used to project each of the input events into a latent feature vector v_i (Figure 3(a)). After that, v_i is further concatenated with d_i, the duration between the i-th event in the sequence and the prediction time, which is denoted as $\hat{v}_i = [v_i, d_i]$. This combined vector is the input for two RNNs, as shown in Figure 3(b) and (c).

The first network, RNN_α (Figure 3(b)), takes the information of all events at each time point into consideration to ensure a high-accuracy prediction result. The outputs of the model (i.e., $(\alpha_1, \ldots, \alpha_t)$) are weights that indicate the accumulated influence on the prediction results at each time point.

The second network, RNN_β (Figure 3(c)), estimates the influence of each individual event at each time point on the prediction results. These estimates facilitate the interpretation of the prediction results. The output, $(\beta_1, \ldots, \beta_t)$, are vectors at different time points, with each field in a vector representing the influence of an individual event on the prediction results. A positive/negative field value corresponds to an event that is associated with an increase/decrease in the occurrence probability of the predicted event, respectively.

The results from the overall model are calculated using a softmax layer, which predicts the occurrence probability of each event as follows:

$$\hat{\mathbf{y}}_t = softmax(\mathbf{W}_{out}\hat{\mathbf{o}}_t + \mathbf{e}_{out}),$$

where $\mathbf{W}_{out}$ and $\mathbf{e}_{out}$ are the parameters to be learned in the softmax-layer; $\hat{\mathbf{o}}_t$ is the context vector at time point t, which we define as a combination of the previous outputs:

$$\hat{\mathbf{o}}_t = \sum_{i=1}^{t} \alpha_i \boldsymbol{\beta}_i \odot \mathbf{v}_i$$

where $\odot$ is the element-wise multiplication operator. The original Retain model is designed with the purpose of producing bivariance survival predictions. To extend the use case of the original model and support the real-world application requirements of predicting the risk of multiple diseases, we adapted the loss function with a specially designed cross-entropy loss as follows:

$$L = -\frac{1}{N}\sum_{k=1}^{N}\frac{1}{T^{(k)}}\sum_{t=1}^{T^{(k)}}(\mathbf{b}_w\mathbf{y}_t^T log(\hat{\mathbf{y}}_t) + (1-\mathbf{y}_t)^T log(1-\hat{\mathbf{y}}_t)),$$

where N is the number of samples, $T^{(k)}$ is the length of the sequence in each sample, y_t is the ground truth, and $\hat{y}_t$ represents the prediction results. b_w is a vector that is included within the loss function to address the presence of highly skewed training data. Each field in b_w is calculated as $1/log(n)$, where n indicates the number of occurrences of an event within the training samples. b_w helps overcome skewed distributions within the training samples by reducing the marginal importance of additional event occurrences for high-frequency events. Finally, we estimate the influence of a historical event s occurring at timestamp t to the prediction results based on α_t and $\boldsymbol{\beta}_t$ as follows:

$$\textbf{Influence}(s,t) = \alpha_t \mathbf{W}_{out}(\boldsymbol{\beta}_t \odot \mathbf{W}_{emb}[:,s]),$$

where W_{emb} is the weight matrix of the input layer that transforms the input sequence into feature vectors, and W_{out} is the weight matrix of the output layer (i.e., the softmax layer) that transforms the latent vector into probabilities.

4.3.2 Visualization. We represent a patient's electronic medical record as a sequence of medical events, which are displayed using rectangular nodes arranged horizontally in order of event occurrence, as shown in Figure 4(a). To avoid overlaps (during periods of time with multiple medical events) and large gaps (during periods of time where medical events are infrequent), the event nodes are spaced with equal distance between them. The actual event times are marked above the event nodes using text labels.

Successive event nodes are depicted with a duration bar connecting the nodes, and each bar is labeled with the time span between events. When multiple events occur at the same time (as is common in medical data), a treemap-based representation is used to compactly represent the multi-event information within a single rectangular node. All events are color coded by event type, with dark gray representing treatments and light gray representing diagnoses. Hovering the mouse over event nodes highlights the corresponding node into orange

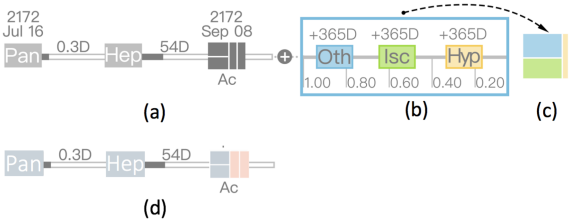

Fig. 4. The visual design of the prediction view includes the medical event sequence leading up to the time point of prediction (a), the prediction box showing the most likely diagnoses in order of predicted probabilities (b), the treemap summarizing the probabilities of diagnoses (c), and the color of the rectangle indicating the influence of the event after clicking on a diagnosis event (d).

and triggers the display of a tooltip showing additional details of the event. Scrolling and zooming operations are allow for further exploration of the patient's medical history.

The prediction results are visualized within a box located to the right of the event sequence visualization (Figure 4(b)). The prediction box contains a series of rectangular nodes, one for each of the most likely predicted diagnosis events for the patient. Each rectangular nodes is color coded by diagnosis type, where the set of possible diagnoses are pre-chosen by a physician using the dropdown list shown in Figure 1(b-1). This choice is determined by the physician based on his or her pre-diagnosis of the patient's condition.

The order (from left to right) of the event nodes inside the prediction box are determined by the predicted occurrence probability of the events. Therefore, the left-most event box within the prediction box corresponds to the diagnosis that is predicted to be most likely for the patient. The predicted likelihood of each diagnosis event decreases as the boxes move toward the right of the prediction box. The color saturation for each box indicates the prevalence of the predicted diagnosis within the medical records for a population of similar patients.

Users can click on a diagnosis event to view more details about the predicted diagnosis. Available information includes a general description of the diagnosis, symptoms, causes, diagnosis methods, treatments, and typical prognosis. These details are displayed within the description view for physicians to review. The color of rectangular nodes will change to blue or red to indicate the influences of the historical events after clicking on a diagnosis event. Red color indicates that the event can lead to the disease, whereas blue color implies that the event can lower the risk of the disease. The color saturation presents the degree of influence.

4.4 Similar Patient Retrieval and Comparison

As identified in the pilot study, a key requirement for clinicians is the ability to compare the focal patient and prediction results to other patients with similar medical records (R3). CarePre allows users to retrieve similar patients in two ways: (1) via brushing a patient similarity histogram (Figure 1(d)) and (2) via explicit queries using key medical events (Figure 1(e)). Similar patients retrieved via either interface are displayed in a patient list (Figure 1(f)) that depicts a detailed event sequence for each similar patient (Figure 1(g)) to allow detailed comparisons.

4.4.1 Patient Similarity and Sequence Alignment. To support the preceding functions, CarePre adopts a distance measure to quantify the similarity between events sequences that is robust to differences in sequence length and event timings. To this end, CarePre uses the event-to-vector and sequence alignment techniques introduced in ET^2 [21]. Specifically, a vector representation of each event in a set of sequences is first calculated based on a neural network model. Sequences are then aligned temporally using a dynamic time working algorithm (DTW) [34], and distances are calculated using the event vectors. The algorithm measures similarity between sequences by estimating the similarity between each pair of events respectively in these sequences based on the Euclidean distance of the corresponding event vectors.

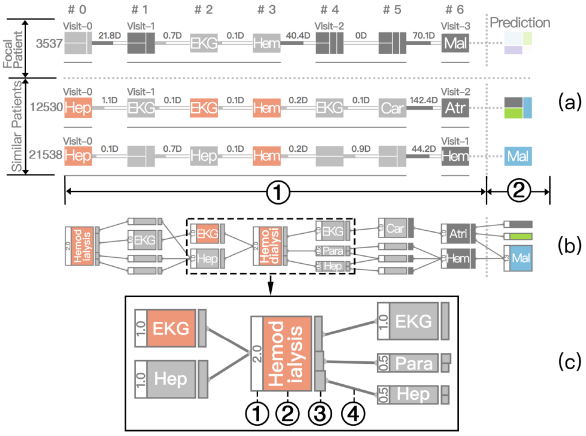

Fig. 5. Medical event sequences for similar patients are visualized as either individual sequences (a) or an aggregated flow diagram (b). Part (c) shows a more detailed illustration of the aggregate view.

4.4.2 Visualization. The patient similarity view displays event sequence data for both the focal patient and the patients most similar to him or her. By default, the system retrieves a group of sequences with normalized distance to the focal sequence under 0.1 for analysis but also allows users to determine similar patients for comparison via brushing on the similarity histogram or querying by key medical events. The event sequences for similar patients are aligned to the focal patient and visualized in parallel, as shown in Figure 5(a). We divide each of the similar sequences into two parts: (1) a history section, which best matches with the focal patient's historical medical records up to the current point in time (Figure 5(a-1)), and (2) an outcome section, which depicts the outcomes observed for the similar patients in comparison to the predicted outcome results for the focal patient (Figure 5(a-2)). This view adopts a visual design that is similar to the prediction view.

To support more effective one-to-many comparison between the focal patient and the set of selected similar patients, we aggregate the medical event sequences for the similar patients into a flow-based visualization that illustrates the overall evolution of diseases and treatments within the group over time. In each time stage, we group patients together if they experience identical events. Each medical event is visualized using a compound rectangle, with the height proportional to the population of patients with the corresponding event at the corresponding time stage. Patients with multiple events co-occurring within the time stage are split by the number of co-current events (as shown in Figure 5) to make sure the total number of the population remains consistent in all time stages. The number of patients (after weighting by event-co-occurrence) flowing through each event is displayed as text in the leading rectangle (Figure 5(b-1)). The middle rectangle of each node (Figure 5(b-2)) displays the event name. Finally, several connection glyphs on the right edge of the node (Figure 5(b-3)) depict connections (via the linking lines) to subsequent nodes that occur in the next time stage (Figure 5(b-4)). The height of each connection glyph indicates the number of patients whose medical record includes the corresponding event transition. The width of the connection glyphs corresponds to the average duration of the transition.

4.5 Treatment Outcome Analysis

The CarePre system provides a set of interactive analysis capabilities to identify key factors that effect the prediction results (R4) and make more informed treatment decisions by simulating possible treatment plans (R5). This is accomplished through interactions that edit the focal patient's event sequence within the prediction view (Figure 1(b)) and visual comparison of the edited sequences in the outcome analysis view (Figure 1(h)).

The outcome analysis capability is summarized in Figure 6. Users can edit the focal patient's original event sequence using four interactive operations: (1) adding a new event, (2) removing an existing event, (3) adjusting

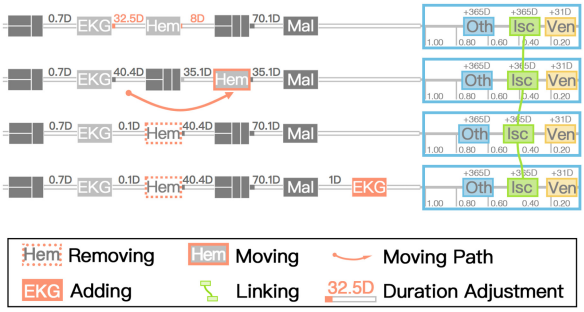

Fig. 6. We enable four interactions for outcome analysis: Removing, Moving, Duration Adjustment, and Adding. The adjusted event sequences are highlighted on the left side with annotations as shown in the figure. The corresponding predicted outcomes are shown on the right side of the view.

the order of events, and (4) changing the duration between events. Updated prediction results are calculated in real time in response to any edit operation performed, and users have the option to save an edited event sequence (and the resulting prediction) as a new entry within the outcome analysis view. This allows clinicians to compare edited event sequences to explore how changes in a patient's medical record (i.e., a new treatment or the absence of a co-morbidity) impacts the prediction results. To support this activity, the view highlights each of the user's event sequence edits in orange (Figure 6) and uses coordinated highlighting to link predictions of the same medical event across edited sequences (e.g., the same predicted diagnosis appearing for two different edited versions of the focal patient's medical record). Users can also zoom in/out on the prediction box to retrieve more detailed views. These interactions help communicate changes in risk between sequences, especially when the same events (but with different probability) are predicted for the alternative edited sequences.

A common use case for these features is when a physician investigates the potential outcomes of different treatments. The physician can create multiple edited sequences by adding the potential treatment events to the end of the focal patient's original medical record. Viewing the predicted results under the assumption of alternative treatments can help the clinician understand the impact of each treatment option. Alternatively, a physician could create alternative versions of a patient's medical record by removing individual events. This would facilitate model interpretation by allowing a clinician to see the impact of a given feature on the prediction result.

Finally, to support further analysis on the contribution of key events to the predicted outcomes, CarePre computes the degree to which each potential treatment is associated with each of the prediction targets within the similar patient population. These associations are displayed in the significance view (Figure 1(i)) as a matrix where each row is a treatment group and each column is a predicted disease. The rows are clustered to group-related treatments using the event vector technique presented earlier in this section.

Each cell in the matrix includes a diagram that shows the change of a disease's mean occurrence probability (shown in the y-axis) and the 95% confidence interval within the subset of similar patients with the treatment (left plot) versus those without the treatment (right plot). Cells with statistically significant differences are highlighted with a white background.

5 EVALUATION

This section presents the results from three forms of evaluation: quantitative experiments to measure the performance of the prediction model; two case studies with a cardiologist and a pulmonologist, respectively; and interviews with seven physicians in both the United States and China.

Table 1. Comparison of Prediction Performance

	Retain	Retain Extended
Neg Log Likelihood	0.2834 ± 0.0036	**0.2768** ± 0.0036
AUC	0.8294 ± 0.0022	**0.8307** ± 0.0026
Precision	0.8126 ± 0.0053	0.8126 ± 0.0054
Recall@2	0.6859 ± 0.0081	**0.6943** ± 0.0082
Recall@4	0.8954 ± 0.0027	**0.8973** ± 0.0032

5.1 Evaluation of the Prediction Model

To verify that the expansion of prediction scope in our model (with the extension of multi-event prediction) has no negative effect on the precision of prediction, we compared the performance of our model (with extensions for multi-event prediction) to the original Retain single-event prediction model [12] from which our model was derived.

More specifically, medical records for patients with cardiovascular disease and at least one hospitalization were chosen from the MIMIC dataset, and their medical records were transformed into one or more event sequences based on a 6-month time window. Each sequence ended at a hospital admission event and started 6 months prior. As a result, 7,537 patients were selected and 64,269 sequence samples were generated. These samples were divided into the training and testing sets using a 7:3 ratio. We further cleaned the sequences by preserving only diagnosis and treatment events. Both the original and extended Retain models were trained using the training samples to predict the risk of five highly prevalent heart and cardiovascular diseases. The disease risks were simultaneously predicted using a single extended Retain model. Meanwhile, five independent models (one for each disease) were trained for the original Retain model design. As shown in Table 1, the negative log likelihood of the extended model is lower than that of the original model. The accuracy of both the extended model and the original model are around 0.83, whereas the accuracy of the extended model is higher. The precision of two models is around 0.81, and the extended model outperformed the original Retain with regard to both the top-2 and top-4 recall. Overall, the performance evaluation result shows that our extended model performed similarly to (slightly better than) the original, whereas our extended model is able to support prediction for multiple events and has a wider application.

5.2 Case Study I: Cardiovascular Disease

We conducted a case study with a senior inpatient cardiovascular doctor with more than 20 years of clinical experience in China. During the study, we first introduced the CarePre system and the doctor was invited to use the system for himself. After getting familiar with the system's functions, the doctor was asked to perform a series of tasks, including interpreting prediction results, making a treatment decision, and estimating the future outcomes for different treatment plans. The study lasted for about 2 hours, and the doctor was encouraged to ask questions or make comments at any time.

Figure 7 shows the results of our study. After reviewing the patient's historical medical records (Figure 7(a)), the doctor said, "This patient is being treated with some typical medicines such as metoprolol and furosemide." He also noted that the patient suffered from diabetes after noticing regular insulin injections within the patient's medical records. According to the prediction results, the patient had a high risk of heart failure in the future. The doctor mentioned, "It is possible as the diabetes may lead to coronary disease and finally develop into heart failure." The doctor then turned to the similar patient view (the aggregated form), which displayed the disease progression of the 10 most similar patients automatically retrieved by the system. He was impressed by the capability of this view in summarizing complex sequence progressions. After a brief inspection, the doctor selected the group of 6 patients with heart failure for further inspection, and the corresponding disease progression paths were automatically highlighted by the system. He believed that this view was "very informative," and that the

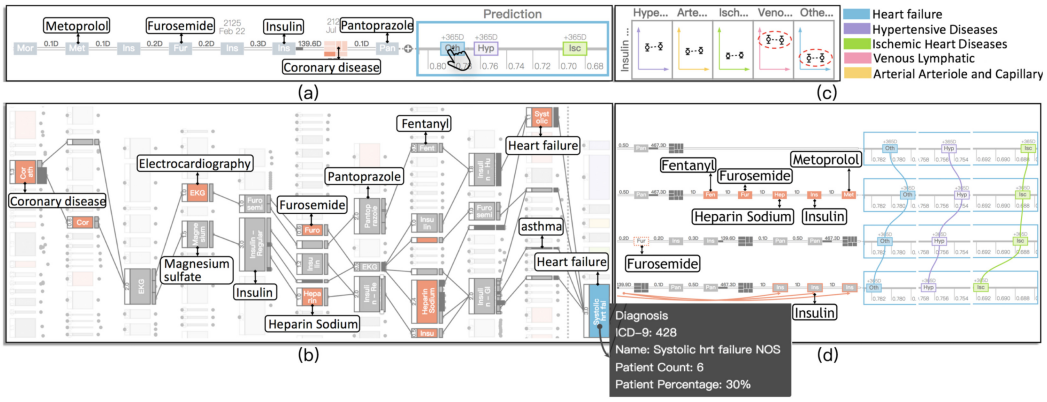

Fig. 7. A case study based on a subset of MIMIC data with cardiovascular disease. The results shown in this figure were identified by our expert user.

click-to-highlight function was able to "clearly show the progression of an outcome in context of treatments." He felt that this view gave him more confidence in the prediction results as "it provides specific evidence [to support the prediction results]." He also mentioned that this view would be particularly useful for medical researchers as "it illustrates many examples following different treatment plans" (see annotations in Figure 7(b)).

The doctor was also interested in the system's outcome analysis function. Specifically, he first made a care plan for the focal patient by adding multiple treatments (e.g., fentanyl, furosemide, insulin, and metoprolol) to the end of the patient's existing medical record, as shown in Figure 7(d). In response, the risk of heart failure and hypertensive disease both decreased. Next, the doctor removed all of the events for furosemide (a common medicine used for heart failure patients to treat fluid build-up) from the sequence, resulting in a significant increase in the risk of heart failure. This also revealed in the view seen in Figure 7(c), for example, that injections of insulin had a significant effect on reducing the risk of heart failure. Finally, the doctor moved all insulin to the end of the sequence to mimic a scenario in which the patient delayed diabetes treatment. This resulted in a further increase in heart failure risk. The correctness of the various predictions were verified by the doctor.

5.3 Case Study II: Respiratory Diseases

We conducted a another case study with a pulmonologist who has more than 8 years of clinical experience in China. We first introduced the CarePre system and encouraged the doctor to use the system. The tasks included diagnosis prediction, similar patient comparison, and outcome analysis. The study lasted for about 1.5 hours.

Figure 8 shows the results of our study. After reviewing the patient's historical medical records (Figure 8(a)), the doctor said, "This patient is being diagnosed with pneumonia and treated with heparin sodium and furosemide." He noted that this patient suffered from both pulmonary diseases and coronary diseases. "This patient must be an [elderly person]. Pulmonary diseases can easily lead to pulmonary heart diseases for the elderly. [However], heart disease can also result in pulmonary diseases such as pulmonary edema. Furosemide is typically used to treat pulmonary edema." As shown in the prediction results, the patient has a low risk of pulmonary diseases in the future, which indicates that the applied treatment plan was effective for the focal patient. The doctor mentioned, "It seems that the treatment plan was effective. I am also wondering which medicine played the most important role in the plan." He further clicked on the green rectangle (obstructive chronic bronchitis, which is the most common complication of pulmonary heart disease) to analyze the effects of the historical events, as shown in Figure 8(a). According to the results, the event heparin sodium was a key factor that reduces the probability of obstructive chronic bronchitis. The doctor said, "As far as I know, heparin sodium doesn't have a direct effect on the obstructive chronic bronchitis. I need more information to confirm these results."

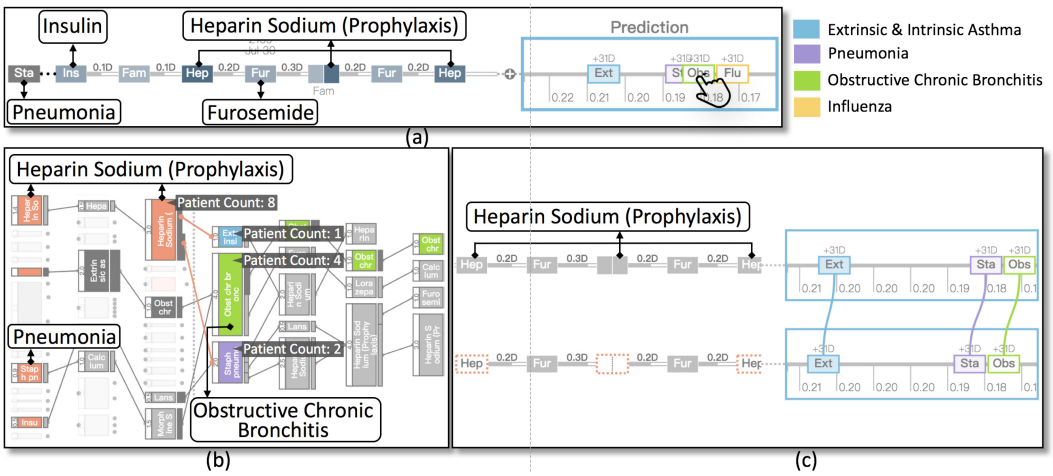

Fig. 8. A case study based on a subset of MIMIC data with respiratory diseases. The results shown in this figure were identified by our expert user.

The doctor then explored the similar patient view (Figure 8(b)), which shows the disease progression of similar patients. The doctor further filtered a group of patients who had the same events in the first stage as the focal patients. The corresponding disease progression paths of the focal patients were then highlighted by the system. The doctor found that eight patients were treated with heparin sodium. Among these eight patients, one of them had asthma and two of them had pneumonia, whereas the rest of them did not have any pulmonary diseases in the outcome. However, the patients who did not take heparin sodium had obstructive chronic bronchitis. The doctor said, "The progression of the similar patients confirmed that heparin sodium can lower the risk of obstructive chronic bronchitis. The result actually reminded me of a new research work in medical science. This research work evaluated the therapeutic use of heparin in patients with COPD (chronic obstructive pulmonary disease)" [44]. He commented that the progression of the similar patients provided more information to confirm the prediction results via the summarization of the raw data. He also mentioned, "It can inspire me to make better decisions. We can also learn from the progression of similar patients."

The doctor also performed the outcome analysis task using the system. Specifically, he removed all of the events for heparin sodium from the sequence, resulting in a significantly increased risk of pulmonary diseases (Figure 8(c)). This result again confirmed the importance of heparin sodium for patients who suffered from pulmonary heart diseases. The doctor said, "This function is really novel and useful. Knowing the effect of the treatment in advance can be very helpful in clinical scenarios. Providing the outcome analysis result can also make the patients more confident in making treatment plans."

5.4 Domain Expert Interview

In addition to the case studies, we conducted in-depth interviews with three senior physicians (E_{1-3}) separately and a group discussion with four pulmonologists (E_{4-7}). To ensure diversity, interviewees involved were from different backgrounds. In particular, E_1 and E_2 were two Chinese cardiologists, both with more than 15 years of clinical experience; E_3 was a senior physician in the United States; and E_4 through E_7 were four Chinese pulmonologists, with 6 to 8 years of clinical experience. To avoid bias, none of these doctors participated in our pilot study. Each interview session started with a half-hour interactive demonstration of our system. The doctors were then invited to use the system on their own. We provided the doctors with the corresponding case study tasks in their area of expertise as a reference, but they were encouraged to freely experiment with the system. After the experts finished exploring the system's functionality, we conducted a semi-structured interview guided

by several questions, including the overall usefulness, ease of use, general pros and cons of the prototype system, visualization designs, and insights obtained from using the system. Each of the four interview sessions lasted for approximately 1.5 hours, and the interview process was recorded. We summarize the collected feedback into the following four themes, including feedback on three key functionalities of CarePre, followed by the user-friendliness of our visualization and interaction design.

5.4.1 Diagnosis Support. All experts appreciated the design of CarePre in supporting diagnosis and believed that the capability of our system in estimating the risk of potential diseases is useful. E_1 said, "I need to take care of over 50 patients a day . . . sometimes I am just too tired to avoid making mistakes . . . if the system is developed based on statistical analysis of similar medical records, I'd love to trust the results . . . and it can actually help us reduce the risk of making a mistake." E_2 mentioned that a potential use of CarePre would be helping inexperienced doctors or medical students make more accurate diagnoses. E_2 also mentioned, "This tool can [help] reduce a doctor's burden." Similarly, E_3 mentioned, "Doctors' time is valuable, [and] quickly estimating the risk of a patient [using the system] is a useful function." The pulmonologists ($E_{4–7}$) also agreed that the system can help doctors make diagnoses more efficiently, and they felt that the system "can be applied to assist consultation." Moreover, they also found revealing the impact of historical event on the prediction result useful. For example, as E_4 commented, "We sometimes hesitate to trust the machine learning models because they usually fail in providing reasons. It can help raise our confidence if the system can illustrate how the model [arrives] at the result."

5.4.2 Similar Patient Retrieval and Comparison. According to the experts' feedback, comparing the focal patient to similar patients "accords with the idea of evidence-based medicine." As E_5 explained, "This is similar to what we usually do in Propensity Score Matching (PSM) but is much more advanced." E_6 agreed and added, "We used PSM to study the effect of clinical decisions based on statistical analysis. However, it is generally hard to find [a] proper study group [due to the high complexity of observational data] . . . it is very powerful that your tool can automatically identify patients with similar progression and make comparison with rich contextual information." The usefulness of our similar patient comparison mechanisms was also recognized by two cardiologists, as E_1 said, "Comparing to the similar patients in detail not only gives me more confidence of the prediction results but also provides me with rich treatment examples." E_2 also commented, "Medical records of similar patients are an important reference for a doctor to make a proper diagnosis, but sometimes the doctor cannot fully review a patient's entire medical records [due to limited time or unavailable of the data] . . . the system provides a more efficient way for us to retrieve the similar patients [when compared to the system we are currently using]." E_3 felt that "the most valuable part of the system would be the impact different treatment approaches would have on similar patients."

5.4.3 Treatment Outcome Analysis. The treatment outcome analysis was considered a highlight of the system and was the most discussed during all interviews. E_1 felt that the idea of virtually making different care plans and comparing their potential outcomes was a "cool and valuable" feature to support making a prognosis. E_2 felt the same way and stated that "this system provided an interactive way for exploring some complicated situations and their influence on the patient." This was also mentioned by E_7, as he commented, "We used to study the effect of a single treatment with [a] RCT (randomized clinical trial), [whereas] this tool can simulate the effect of a combination of treatment, which is awesome." E_3 felt that knowing the impact of treatment on the risk of diseases is convenient for doctors to further investigate key factors that influences the outcome. E_4 was especially excited about this feature and mentioned that "this can be used to help junior doctors try out the dose of insulin in treating diabetics." E_5 agreed and added, "This system can help junior doctors learn from the experience of senior doctors in a more interactive way instead of plain text [in] the textbook."

5.4.4 User-Friendliness. All interviewees felt that the system was more complex than any tools they had used in their daily work. Comments such as "looks overwhelming" and "seems difficult to learn" were frequently

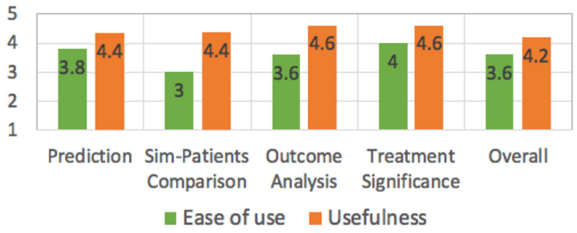

Fig. 9. The questionnaire results.

mentioned at the beginning of each interview session. However, after they became familiar with the system's functionalities and explored the system by themselves for a while, they felt that the design was "intuitive" and the usage of system "gradually becomes clear." For example, when the similar patient view was first introduced, the senior physicians felt that it was "complex and took time to learn," especially the aggregated representation. However, once they became familiar with the design, they felt that this view was "informative" and "clearly illustrated different care plans and the corresponding outcomes." E_6 also stated, "I was confused by so many views and had no idea where to look at, but then I realized that each view is especially designed for a specific task, and that [what analytical task to perform] is all I need to care about." E_7 agreed and added "especially that the [exploration] pipeline for each task is relatively fixed." E_5 felt that the interaction "helped a lot in reducing the complexity," as he explained while tweaking the treatment plan: "Tools we used typically employ a lot of context-menu [to interact with users], which is difficult to memorize, [whereas] this tool allows me to manipulate the visualization more directly."

5.5 Post-Interview Questionnaire

All experts (seven interviewees and two case study participants) were invited to complete a questionnaire after their interviews.

The questionnaire asked experts to rate the ease of use and usefulness of the key features of our system on a scale of 1 to 5, with 1 indicating difficult to use/useless and 5 indicating very easy to use/very useful. The results are summarized in Figure 9. The primary issue highlighted in these results was related to ease of use, which will be discussed in the next section.

6 DISCUSSION

The results from our case study and expert interviews were generally positive, with users confirming the usefulness of CarePre and expressing excitement regarding the treatment outcome analysis capabilities. However, they also identified several limitations, provided a number of constructive suggestions, and mentioned some interesting potential applications of the system.

6.1 Limitations and Solutions

The major limitations of the CarePre system include ease-of-use, data quality, and a lack input methods.

6.1.1 Ease of Use. Although all of the experts acknowledged the usefulness of the system, they also mentioned that learning how to use it took essential time away from a doctor. All of the experts were familiar with traditional statistical diagrams such as histograms and line charts; however, they needed some practice to read some of the more complex views introduced in CarePre. But they also believed that these new designs were more informative when compared to more familiar statistical charts. The experts also felt that the difficulty mainly comes from the lack of training. For example, E_2 said, "We (doctors) spend years in school to learn how to make [a] diagnosis based on those [traditional] statistical tools and diagrams . . . your tool is obviously more informative but we

just need more time to get familiar with it." In addition, operating multiple coordinated views also takes some effort. Both E_1 and E_3 mentioned that it would easier to use if the tool could directly generate and print out a report without as many interactions. E_3 also said, "It will be easier to use if you could somehow separate the views of three different functions apart into multiple pages and guide the operation in a step to step manner instead of packing them all together." The four young doctors ($E_{4–7}$) agreed that they can use the system easily after training. E_4 said, "It will take some time for us to learn how to use the system. However, the design is not difficult to understand. After the training, I am willing to use the system. I also hope to get some new findings using this system."

6.1.2 Data Quality. The Chinese physicians were concerned about the quality of the training data that directly influenced our analysis results. Both E_1 and E_2 mentioned that the quality of the EHRs collected in Chinese hospitals were much worse than those of the MIMIC dataset. They mentioned that the medical data in China was primarily free text, and that many hospitals in China were just starting to use EHR systems. That limits the longitudinal extent of data that could be used as input to the system. For this reason, they believed that CarePre might not be as useful in Chinese hospitals right away. E_2 reminded us that a prognosis estimate is usually based on the statistics of a very large collection of patients over a very long period of time. She pointed out that this feature, therefore, was useful only when the underlying data were rich enough to represent the rich variety of outcomes that patients face. E_3 also mentioned that the treatment outcome analysis should be based on a larger dataset collected within a longer time window (e.g., several years). E_4 added that the knowledge in the field of medicine updated very fast. As a result, the data used in the system need to be updated as well. To solve this problem, we need to have deeper cooperation with the doctors in the data processing step. With the help of doctors, we can uncover the most meaningful information in the dataset and update our system with more data from real clinical scenarios.

6.1.3 Lack of Input Methods. E_1 also felt that although CarePre was useful, the design was not sufficient, as it has limited ways for clinicians to enter new medical data. In particular, she said, "When compared to the existing system, your tool focuses more on the analysis but lacks a convenient method for me to enter medical records in the text form." E_5 mentioned that providing the information of historical medical records is not enough. He said, "Although historical medical records are important, doctors also want to input the vital signs of the patients into the system." In the future, another view with interactions should be designed in our system to help doctors input the information they are interested in for analysis efficiently. In this way, our system will be more comprehensive and useful.

6.1.4 Insufficient Context Information. All of the experts felt that the information shown in the description view was helpful, but they would like to have more. They suggested we collect more information, such as the latest diagnosis guidelines and information about new medicines. In the future, we will upload the up-to-date medical information and latest guidelines into the system to make the description view more valuable.

6.2 Implications

Our experts raised many implications of the CarePre system as well, which can be summarized into two broad categories.

6.2.1 From Knowledge Sharing to Experience Sharing. All of the experts believed that CarePre would be especially useful for junior physicians, medical students, or other inexperienced health professionals. They believed that since the prediction model in CarePre is trained based on the treatment records made by experienced physicians, it would capture those doctors' experiences. In comparison, most existing knowledge-based systems only share medical knowledge. In particular, E_1 and E_2 mentioned that in China there are many undeveloped rural areas with poor health systems where doctors are less experienced and less well trained. The CarePre system would help provide information to these doctors based on the experience of more senior clinicians. This maps

to typical doctor training techniques, where doctors first learn from medical textbooks before a long period of training under the supervision of senior doctors to help them gain knowledge through experience.

6.2.2 From Doctors to Other Users. Our experts also suggested many other potential application scenarios for the CarePre system. For example, E_1 believed that our system would be very useful for analysts in a medical insurance company. "It can help an insurance company estimate the risk of a patient in a more efficient way," said E_1. Both E_1 and E_2 mentioned that our tool could be very helpful for medical research, as it is "[built] based on statistical analysis and provides many advanced visual diagrams, illustrating the evidence of the analysis results." Both E_2 and E_3 felt that the CarePre system could also be directly used by a patient, as "it suggests the risk a patient may have" and "the patient may want to spend more time investigating the functionality of the system." These scenarios greatly expand the application scope of the CarePre system, although certain design changes may be required for different applications.

7 CONCLUSION

This article introduced an intelligent clinical decision assistance system, CarePre, which uses large-scale EHR data to help physicians make decisions during their clinical workflow. The system, designed based on requirements identified in a pilot study, provides clinical assistance through a state-of-the-art deep learning prediction model and an interactive visual interface for exploration and interpretation. The interaction pipeline of our system consists of three major steps: (1) diagnosis support, (2) retrieval and comparison of similar patients, and (3) treatment outcome analysis. We evaluated the system via case studies, expert interviews, and a quantitative evaluation of the predictive model. The results from these evaluations showed that the overall system provided valuable assistance to the clinical decision process. In the future, we plan to address the aforementioned issues and conduct a larger evaluation of the system in a local hospital to update the system's models based on local patients' conditions.

ACKNOWLEDGMENTS

We would like to thank all medical experts for participating our case studies and interviews, and we also thank all reviewers for their valuable advice.

REFERENCES

[1] U. Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. 2017. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences* 415 (2017), 190–198.

[2] Adrian Benton, Margaret Mitchell, and Dirk Hovy. 2017. Multi-task learning for mental health using social media text. arXiv:1712.03538.

[3] Eta S. Berner and Tonya J. La Lande. 2007. Overview of clinical decision support systems. In *Clinical Decision Support Systems*. Springer, New York, NY, 3–22.

[4] Eta S. Berner, George D. Webster, Alwyn A. Shugerman, James R. Jackson, James Algina, Alfred L. Baker, Eugene V. Ball, et al. 1994. Performance of four computer-based diagnostic systems. *New England Journal of Medicine* 330, 25 (1994), 1792–1796.

[5] David Blumenthal and Marilyn Tavenner. 2010. The "meaningful use" regulation for electronic health records. *New England Journal of Medicine* 363, 6 (2010), 501–504.

[6] Leo Breiman. 2001. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science* 16, 3 (2001), 199–231.

[7] Jerome H. Carter. 2001. *Electronic Medical Records: A Guide for Clinicians and Administrators*. ACP Press, Washington, DC.

[8] Jerome H. Carter. 2007. Clinical decision support systems. In *Design and Implementation Issues*, E. A. Berner (Ed.). Springer, New York, NY, 64–98.

[9] Ajay P. Chainani, Santosh S. Chikne, Nikunj D. Doshi, Asim Z. Karel, and Shanthi S. Therese. 2018. Disease inference from health-related questions via fuzzy expert system. In *Information and Communication Technology for Sustainable Development*. Springer, Singapore, 91–102.

[10] Kumardeep Chaudhary, Olivier B. Poirion, Liangqun Lu, and Lana X. Garmire. 2017. Deep learning based multi-omics integration robustly predicts survival in liver cancer. *Clinical Cancer Research* 24, 6 (2017), 1248–1259.

[11] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. 2016. Doctor AI: Predicting clinical events via recurrent neural networks. In *Proceedings of the Machine Learning for Healthcare Conference (PMLR'16)*. 301–318.
[12] Edward Choi, Mohammad Taha Bahadori, Jimeng Sun, Joshua Kulas, Andy Schuetz, and Walter Stewart. 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Advances in Neural Information Processing Systems*. 3504–3512.
[13] Jesse C. Crosson, Christine Stroebel, John G. Scott, Brian Stello, and Benjamin F. Crabtree. 2005. Implementing an electronic medical record in a family medicine practice: Communication, decision making, and conflict. *Annals of Family Medicine* 3, 4 (2005), 307–311.
[14] Padideh Danaee, Reza Ghaeini, and David A. Hendrix. 2017. A deep learning approach for cancer detection and relevant gene identification. In *Proceedings of the Pacific Symposium on Biocomputing*. 219–229.
[15] P. S. Dayan, D. W. Ballard, E. Tham, J. M. Hoffman, M. Swietlik, S. J. Deakyne, E. A. Alessandrini, et al. 2017. Use of traumatic brain injury prediction rules with clinical decision support. *Pediatrics* 139, 4 (2017), e20162709.
[16] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2016. EventAction: Visual analytics for temporal event sequence recommendation. In *Visual Analytics Science and Technology*. IEEE, Los Alamitos, CA, 61–70.
[17] Fan Du, Catherine Plaisant, Neil Spring, and Ben Shneiderman. 2017. Finding similar people to guide life choices: Challenge, design, and evaluation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 5498–5544.
[18] Srinivas Emani, David Y. Ting, Michael Healey, Stuart R. Lipsitz, Andrew S. Karson, and David W. Bates. 2017. Physician beliefs about the meaningful use of the electronic health record: A follow-up study. *Applied Clinical Informatics* 8, 4 (2017), 1044–1053.
[19] David Gotz and Harry Stavropoulos. 2014. DecisionFlow: Visual analytics for high-dimensional temporal event sequence data. *Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1783–1792.
[20] Richard W. Grant, Jonathan S. Wald, Jeffrey L. Schnipper, Tejal K. Gandhi, Eric G. Poon, E. John Orav, Deborah H. Williams, Lynn A. Volk, and Blackford Middleton. 2008. Practice-linked online personal health records for type 2 diabetes mellitus: A randomized controlled trial. *Archives of Internal Medicine* 168, 16 (2008), 1776–1782.
[21] Shunan Guo, Zhuochen Jin, David Gotz, Fan Du, Hongyuan Zha, and Nan Cao. 2018. Visual progression analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 417–426.
[22] Nils Yannick Hammerla, James Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plötz. 2015. PD disease state assessment in naturalistic environments using deep learning. In *Proceedings of the Conference on Artificial Intelligence*. 1742–1748.
[23] Abhyuday N. Jagannatha and Hong Yu. 2016. Bidirectional RNN for medical event detection in electronic health records. In *Proceedings of the Conference of the Association for Computational Linguistics*. 473.
[24] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, H. Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3 (2016), 160035.
[25] Josua Krause, Adam Perer, and Kenney Ng. 2016. Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 5686–5697.
[26] Gilad J. Kuperman, Dean F. Sittig, M. Michael Shabot, and J. Teich. 1999. Clinical decision support for hospital and critical care. *Journal of Healthcare Information Management* 13 (1999), 81–96.
[27] Bum Chul Kwon, Min-Je Choi, Joanne Taery Kim, Edward Choi, Young Bin Kim, Soonwook Kwon, Jimeng Sun, and Jaegul Choo. 2018. RetainVis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 299–309.
[28] G. B. Erturkmen Laleci, Mustafa Yuksel, Bunyamin Sarigul, Mikael Lilja, Rong Chen, and Theodoros N. Arvanitis. 2018. Personalised care plan management utilizing guideline-driven clinical decision support systems. *Studies in Health Technology and Informatics* 247 (2018), 750–754.
[29] Vernon Lawhern, Amelia Solon, Nicholas Waytowich, Stephen M. Gordon, Chou Hung, and Brent J. Lance. 2018. EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering* 15, 5 (2018), 056013.
[30] Mengchen Liu, Jiaxin Shi, Kelei Cao, Jun Zhu, and Shixia Liu. 2018. Analyzing the training processes of deep generative models. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 77–87.
[31] Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu, and Shixia Liu. 2017. Towards better analysis of deep convolutional neural networks. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 91–100.
[32] George M. Marakas. 2003. *Decision Support Systems in the 21st Century*. Vol. 134. Prentice Hall, Upper Saddle River, NJ.
[33] Megan Monroe, Rongjian Lan, Hanseung Lee, Catherine Plaisant, and Ben Shneiderman. 2013. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2227–2236.
[34] Meinard Müller. 2007. Dynamic time warping. In *Information Retrieval for Music and Motion*. Springer, 69–84.
[35] Mark A. Musen, Blackford Middleton, and Robert A. Greenes. 2014. Clinical decision-support systems. In *Biomedical Informatics*. Springer, 643–674.
[36] Anh Nguyen, Jason Yosinski, and Jeff Clune. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*. IEEE, Los Alamitos, CA, 427–436.

[37] Adam Perer and David Gotz. 2013. Data-driven exploration of care plans for patients. In *Proceedings of CHI Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, 439–444.
[38] Adam Perer, Fei Wang, and Jianying Hu. 2015. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics* 56 (2015), 369–378.
[39] Abdur Rais and Ana Viana. 2011. Operations research in healthcare: A survey. *International Transactions in Operational Research* 18, 1 (2011), 1–31.
[40] Pranav Rajpurkar, Awni Y. Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y. Ng. 2017. Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv:1707.01836.
[41] Daniele Ravi, Charence Wong, Benny Lo, and Guang-Zhong Yang. 2016. Deep learning for human activity recognition: A resource efficient implementation on low-power devices. In *Wearable and Implantable Body Sensor Networks*. IEEE, Los Alamitos, CA, 71–76.
[42] Enrique Rodriguez-Borja, Africa Corchon-Peyrallo, Esther Barba-Serrano, Celia Villalba Martínez, and Arturo Carratala Calvo. 2018. "Send & hold" clinical decision support rules improvement to reduce unnecessary testing of vitamins A, E, K, B1, B2, B3, B6 and C. *Clinical Chemistry and Laboratory Medicine* 56, 7 (2018), 1063–1070.
[43] Amir Salehipour and Mohammad Mehdi Sepehri. 2012. Exact and heuristic solutions to minimize total waiting time in the blood products distribution problem. *Advances in Operations Research* 2012, 25.
[44] Janis K. Shute, Ermanno Puxeddu, and Luigino Calzetta. 2018. Therapeutic use of heparin and derivatives beyond anticoagulation in patients with bronchial asthma or COPD. *Current Opinion in Pharmacology* 40 (2018), 39–45.
[45] Ewout W. Steyerberg. 2008. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer.
[46] Hendrik Strobelt, Sebastian Gehrmann, Hanspeter Pfister, and Alexander M. Rush. 2018. LSTMVis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2018), 667–676.
[47] Thomas Unterthiner, Andreas Mayr, Günter Klambauer, and Sepp Hochreiter. 2015. Toxicity prediction using deep learning. arXiv:1503.01445.
[48] Krist Wongsuphasawat and David Gotz. 2011. Outflow: Visualizing patient flow by symptoms and outcome. In *Proceedings of the IEEE VisWeek Workshop on Visual Analytics in Healthcare*. 25–28.
[49] Po-Yen Wu, Chih-Wen Cheng, Chanchala D. Kaddi, Janani Venugopalan, Ryan Hoffman, and May D. Wang. 2017. –Omic and electronic health record big data analytics for precision medicine. *IEEE Transactions on Biomedical Engineering* 64, 2 (2017), 263–273.
[50] Shuai Xiao, Junchi Yan, Mehrdad Farajtabar, Le Song, Xiaokang Yang, and Hongyuan Zha. 2017. Joint modeling of event sequence and time series with attentional twin recurrent neural networks. arXiv:1703.08524.
[51] Yan Yan, Xinbing Qin, Yige Wu, Nannan Zhang, Jianping Fan, and Lei Wang. 2015. A restricted Boltzmann machine based two-lead electrocardiography classification. In *Wearable and Implantable Body Sensor Networks*. IEEE, Los Alamitos, CA, 1–9.
[52] Zhen-Jie Yao, Jie Bi, and Yi-Xin Chen. 2018. Applying deep learning to individual and community health monitoring data: A survey. *International Journal of Automation and Computing* 15 (2018), 643–655.
[53] S. Yousefi, F. Amrollahi, M. Amgad, C. Dong, J. E. Lewis, C. Song, D. A. Gutman, et al. 2017. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Scientific Reports* 7, 1 (2017), 11707.
[54] Liang Zhao, Jiangzhuo Chen, Feng Chen, Wei Wang, Chang-Tien Lu, and Naren Ramakrishnan. 2015. SimNest: Social media nested epidemic simulation via online semi-supervised deep learning. In *Proceedings of the 2015 IEEE International Conference on Data Mining (ICDM'15)*. IEEE, Los Alamitos, CA, 639–648.
[55] Bin Zou, Vasileios Lampos, Russell Gorton, and Ingemar J. Cox. 2016. On infectious intestinal disease surveillance using social media content. In *Proceedings of the International Digital Health Conference*, Vol. 6. ACM, New York, NY, 157–161.

Received November 2018; revised April 2019; accepted June 2019