MediCoSpace: Visual Decision-Support for Doctor-Patient Consultations using Medical Concept Spaces from EHRs

Citation for published version (APA):
van der Linden, S., Sevastjanova, R., Funk, M., & El-Assady, M. (2023). MediCoSpace: Visual Decision-Support for Doctor-Patient Consultations using Medical Concept Spaces from EHRs. *ACM Transactions on Management Information Systems*, *14*(2), Article 15. https://doi.org/10.1145/3564275

Document license:
CC BY

DOI:
10.1145/3564275

Document status and date:
Published: 01/06/2023

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 04. Oct. 2023

MediCoSpace: Visual Decision-Support for Doctor-Patient Consultations using Medical Concept Spaces from EHRs

SANNE VAN DER LINDEN, Eindhoven University of Technology
RITA SEVASTJANOVA, University of Konstanz
MATHIAS FUNK, Eindhoven University of Technology
MENNATALLAH EL-ASSADY, ETH AI Center

Healthcare systems are under pressure from an aging population, rising costs, and increasingly complex conditions and treatments. Although data are determined to play a bigger role in how doctors diagnose and prescribe treatments, they struggle due to a lack of time and an abundance of structured and unstructured information. To address this challenge, we introduce *MediCoSpace*, a visual decision-support tool for more efficient doctor-patient consultations. The tool links patient reports to past and present diagnoses, diseases, drugs, and treatments, both for the current patient and other patients in comparable situations. *MediCoSpace* uses textual medical data, deep-learning supported text analysis and concept spaces to facilitate a visual discovery process. The tool is evaluated by five medical doctors. The results show that *MediCoSpace* facilitates a promising, yet complex way to discover unlikely relations and thus suggests a path toward the development of interactive visual tools to provide physicians with more holistic diagnoses and personalized, dynamic treatments for patients.

CCS Concepts: • **Human-centered computing → Visualization toolkits**; **User interface management systems;**

Additional Key Words and Phrases: Visual analytics, natural language processing, interaction design, electronic health records

ACM Reference format:
Sanne van der Linden, Rita Sevastjanova, Mathias Funk, and Mennatallah El-Assady. 2023. *MediCoSpace*: Visual Decision-Support for Doctor-Patient Consultations using Medical Concept Spaces from EHRs. *ACM Trans. Manage. Inf. Syst.* 14, 2, Article 15 (January 2023), 20 pages.
https://doi.org/10.1145/3564275

1 INTRODUCTION

With an aging population and rising healthcare costs, the pressure on healthcare systems is increasing [19]. This pressure is especially felt by doctors and nurses. While not a universal cure, information systems promise to make a significant impact to help doctors with documentation, information retrieval, and decision support. One example is the introduction of the electronic health record (EHR) system, where EHRs contain medical narratives. These are textual notes about the

---

Authors' addresses: S. van der Linden and M. Funk, Eindhoven University of Technology, De Zaale, 5612 AZ Eindhoven, the Netherlands; emails: {s.v.d.linden, m.funk}@tue.nl; R. Sevastjanova, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany; email: rita.sevastjanova@uni-konstanz.de; M. El-Assady, ETH AI Center, Binzmühlestrasse 13, 8092 Zürich, Switzerland; email: melassady@ethz.ch.

2158-656X/2023/01-ART15
https://doi.org/10.1145/3564275

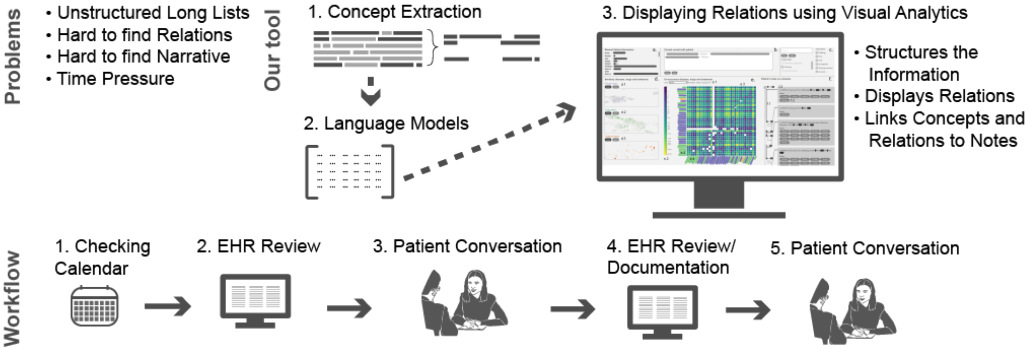


Fig. 1. Doctor's simplified workflow, current problems for diagnosing non-straightforward conditions, and how our tool can help.

patient's condition and progress, to which doctors mainly contribute and on which diagnoses heavily rely [33]. However, analyzing the information contained in these notes is hard because of the complexity of medical issues, the different formats of the EHR systems, and the usability of hospital systems displaying these notes. These systems display textual notes as lengthy lists of narrative text that require extensive scrolling, which leads to information overload and consequently narrative fragmentation [58]. These lengthy lists are not surprising because, on average, individual notes contain 642 words [51] and patients have hundreds of them. Chronically ill patients often have the most notes, e.g., a patient with chronic kidney disease in the U.S. has on average 338 notes [49]. Physicians in the U.S. spend on average 5 [47] to 9 [59] minutes to review the patient information in the EHR per patient encounter (in total 1.5 hours per day [59]). However, this slightly differs per specialty, e.g., endocrinologists spend the most time reviewing EHRs (33% more) and cardiologists the least (44% less) [59].

Especially, diagnosing patients with non-trivial conditions is a tedious task, which could take up to 4.8 years [16] and generate many notes. It is not surprising that physicians misdiagnose approximately 5% [18] to 15% [45] of their patients, ranging from 5% in radiology to 12% in emergency medicine [45]. These errors can have serious consequences regarding the patient's chances of health and treatment success, medical costs (testing for diagnostic purposes accounts for approximately 10% of the healthcare costs in the U.S. [45]), and the doctor's time (0.1% of hospital visits and 0.4% of hospital admission in the U.S. are results of diagnostic error-associated adverse events [45]).

Visual analytic tools could aid in getting a more holistic overview of the patient, especially, for the doctor's decision-making process for diagnosing and devising treatment plans. For example, Sultanum et al. [56] redesign the structure of the EHR notes, by linking notes with similar medical concepts together, to find the patient narrative. However, there are limits to how much information can be assessed in relation: the relationships between diseases (including diagnoses and symptoms), drugs, and treatments. Likewise, we observe lacking tool support for linking potentially discovered relationships back to textual notes and proposing interesting parts of the patient history compared to similar patients.

Our contributions are threefold: (1) Insights into decision-support based on EHRs and problem characteristics of analyzing EHRs for diagnosing patients based on interviews with doctors. (2) A novel visual analytics decision-support tool, *MediCoSpace*, for augmenting doctor-patient consultations to give doctors in hospitals and general practitioners a data-driven overview of possible relations between diseases, drugs, and treatments, both historically and present, and related to similar patients. (3) Results from user expert evaluations that show how *MediCoSpace* could broaden the doctors' solution space, offer new areas of interest, reduce personal biases and stimulate communication between different medical specialties. In the following sections, we introduce related work leading to the problem characterization and requirements. We then explain the data processing pipelines and describe the visualization features of the tool, which are evaluated by medical experts. Finally, we discuss the findings and conclude with future research opportunities.

2 RELATED WORK

This section focuses on previous work on the processing and analysis of EHRs.

2.1 Text Processing and Semantic Concept Extraction from EHRs

Since EHR notes consist of free text, text analysis methods, including approaches that incorporate knowledge graphs or apply language models, can be used to extract the essential concept information. Linking data to an ontology is a common practice. Li et al. [39] present a design framework for named behavioral ontology learning from text. The framework describes linguistic and statistical approaches to address tasks such as variable and synonymous relationship extraction. Knowledge graphs [5, 31, 52] provide insights into (hierarchical) relations and the structure of medical concepts in relation to medical ontology knowledge. For instance, Li et al. [40] present a visual analytics approach by linking medical event sequences to a subgraph in a medical knowledge graph using a domain-knowledge–guided recurrent neural network (DG-RNN) model. Such approaches are effective, yet limited to the information stored in the particular knowledge graph in use.

Deep-learning-based language models (e.g., BERT [11]) have reached high performance in diverse natural language processing tasks. These models are pre-trained on large corpora, learning language structures in an unsupervised manner. Furthermore, domain or task-specific fine-tuning, i.e., adapting the pre-trained weights according to language characteristics of a specific domain (also known as *additional pre-training*) or downstream task [12], is also commonly used in the medical domain. There are several medical domain-specific adaptations of BERT, such as BioBERT [38] and PubMedBert [23] (both additionally pre-trained models on large-scale biomedical corpora), and ClinicalBERT [29] and clinical-kb-bert [25] (both additionally pre-trained on the MIMIC-III [35] dataset to capture patient-record related information and clinical-kb-bert is also pre-trained on UMLS [7] ontology knowledge). Neural language models can be used for different analysis purposes and downstream tasks. First, we can fine-tune them for the named entity recognition task. For instance, Sun et al. [57] fine-tune BioBERT on a machine reading comprehension task that allows it to predict named entity (chemicals, diseases, and proteins) occurrences. Second, we can use them to generate contextualized embedding representations (e.g., on word, sentence, or even document level). To understand named-entity similarity, we can thus use a medical domain-adapted language model to compute their embedding representations and apply a similarity function to determine their similarity. Since this is a very general approach and is not restricted to specific named-entity categories, we apply it in our work. Also, Loureiro et al. [41] use a language model for a medical entity linking task with the MedMentions [44] dataset, whereby the embedding similarity is one step (in addition to entity classification) in their processing pipeline to link entities to an ontology.

2.2 Physician-centric Visual Analysis of EHRs

Doctors use information systems to access and extend EHRs. Currently, Epic [14] is one of the most common commercial EHR systems, which, according to doctors, still suffers from problems, see the top of Figure 1. In general, in the research community, interactive EHR visualizations are most often visualized using bar/line/pie charts, glyphs, and timelines [60]. For example, LifeLines was one of the first tools to visualize textual notes from EHRs as events on a timeline [17]. Researchers have used this timeline structure to visualize EHRs of individual patients abundantly [6, 10, 24, 26, 28, 42, 56, 58] to, for example, display cause and effect [48] or disease progression [50]/risk [40] prediction. Also, Sultanum et al. [56] researched the importance of visualizing text for assisting doctors. Moreover, van der Linden et al. [58] visualized EHRs in a multiscale way to find the fragmented narratives based on the different tasks of the doctor.

Furthermore, stepping away from individual patients, many researchers have visualized patient cohorts as flow visualizations [21, 22, 30, 34, 36, 64, 65] for disease progression, which have limitations in identifying relations. Therefore, Jin et al. [34] visualized causal relations between medical events and two groups. Furthermore, many researchers have used text [20] or basic plots (e.g., line plots) [15] to display summary statistics and heatmaps [30] for visualizing research around the diagnosis process using medical concepts. For example, Hur et al. [30] focused on diagnosis predictions, for which they used different heatmaps (one for the entire cohort, one for the patient, or one to show the difference between them) to show the weights of the medical concepts used in their model.

While these tools make important steps, the diagnosis and creation of non-trivial treatment plans are more difficult than more trivial ones. To our knowledge, no medical decision-support system addresses relationships between diseases, drugs, and treatments combined with advanced search support within the patient's history and across similar patients, and links this back to patient reports to discover and leverage possibly overlooked relations.

3 PROBLEM CHARACTERIZATION

In this section, we describe the first steps of our **user-centered design** (**UCD**) process [63].

3.1 Physician's Workflow

By interviewing a cardiologist (D1), a general practitioner (D2), a medical student (D3), and two medical doctors in internal medicine (D4) and cardiology (D5) about their workflows and comparing them to the processes from Balogh et al. [4] and Adler–Milstein et al. [1], we identified the following general workflow for diagnosing and making treatment plans for non-trivial conditions for doctors of all specialties and experience levels, see Figure 1. Accordingly, the interviewed doctors were from different specialties and experience levels. First, the doctor looks up patient appointment details. The patient already went through *the experiencing health problems* and *engaged with the healthcare system* stages from Balogh et al. [4]. Second, the doctor reviews the EHR for the medical history and the current disease(s) as preparation (related to *the information integration and interpretation stage* [1, 4]). Third, the doctor speaks with the patient and might conduct physical tests (related to *the information gathering stage* [1, 4]). Fourth, the doctor reviews the EHR in more detail to find previous and present diagnoses, issues and physiology, and how the patient appears to progress. Based on this, next steps (related to *the formulation of next steps* [1]) could be conversations with colleagues and possibly diagnostic testing (related to *the information gathering stage* [1, 4]). Also, the doctor matches the symptoms to the most probable diseases to form a working diagnosis (related *to working/leading diagnosis stage* [1, 4]), after which they research the best treatment option online and communicate this to the patient. This is often an iterative process possibly with multiple cycles based on certain outcomes. These final steps also correspond to the final stages of Balogh et al.'s [4] and Adler–Milstein et al.'s [1] processes. We noticed that it differs per specialty how much information the doctors require from the EHR. For example, internal medicine requires a deep dive into the EHR because patients often have vague symptoms, while a cardiologist often needs less information because medical imaging often indicates the main problems directly.

Also, doctors indicated that it is hard to find the correct disease based on ambiguous symptoms. The occurrence frequency of a disease also needs to be taken into account, as well as the patient's lifestyle context; and sometimes it is hard to mentally let go of an initial diagnosis. Visual analytics can assist the doctor (in stages two and four of our workflow) in finding relations between symptoms, diseases, treatments, and drugs to get a more holistic overview of the patient to guide the doctor in information gathering, finding working diagnoses, and possible treatments.

3.2 Tool Requirements

In designing *MediCoSpace*, we focus on the fourth workflow step, the in-depth EHR review. From this, we derive the following requirements based on a thematic analysis [8] of the interviews:

R1: Ability to see relations between diseases, drugs, and treatment of past and present for the current patient.
R2: Ability to see when certain concepts or co-occurrences are mentioned.
R3: Ability to find similar diseases, drugs, or treatments based on a current disease, drug, or treatment.
R4: Ability to compare the patient's relations to similar patients to see similarities and differences.
R5: Ability to link the relations back to the original textual notes of this patient.
R6: Ability to save interesting findings.

4 DATA PROCESSING AND FEATURE EXTRACTION PIPELINE

This section describes our data sources and processing pipelines, see Figure 2, to show relations between diseases, drugs, and treatments concepts (i.e., medical entities). In the remainder of this article, we use the following mini case: the task involves generating a diagnosis for a patient with non-straightforward conditions (mainly cardiovascular diagnoses and a vague current diagnosis; weakness). We want to see if *MediCoSpace* can help the doctor with diagnosing this more clearly. We picked a patient with mainly cardiovascular diagnoses because of the time restrictions of the evaluation (cardiologists require less of an in-depth analysis).

4.1 Data Sources

We used the Unified Medical Language System (UMLS) version 2019AB [7] as ontology knowledge of medical terminology. We also used the MIMIC-III [35] dataset as patient record input. Both are the most extensive, freely available sources suited for our research. The patient from the mini case had 112 notes with, on average, 300 words and four hospital admissions of, on average, 7 days. This patient had (sub-)diagnoses from 20 (sub-)specialties, e.g., cardiology. We compared this patient to a population of 19 random patients with similar main cardiovascular diagnoses, see Section 5. They had, on average, 151 notes containing 460 words, two hospital admissions of, on average, 14 days, and (sub-)diagnoses from 19 (sub-)specialties.

To measure named-entity similarity, we use an adapted version of BERT called *clinical-kb-bert* [25], where the MIMIC-III as well as knowledge base information from UMLS is added into the model pre-training. We chose this model since the authors show that *clinical-kb-bert* outperforms the corresponding model with no knowledge base information and other state-of-the-art models. We adapted their model with a pooling layer. By initializing this adapted model with their

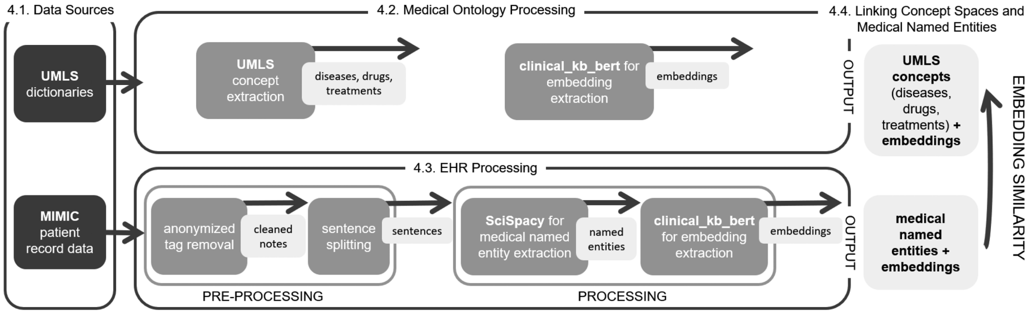


Fig. 2. Data processing pipelines to extract medical concepts from the text. Using two different data sources, we run two parallel pipelines; (1) processing data from medical ontologies to generate the medical concept spaces; and (2) processing the EHRs to extract patient medical entities. Lastly, we link the outcomes of both pipelines to power the visual workspace. The light gray boxes under the arrows describe the in- and outputs of each step.

pre-trained weights, we further fine-tuned the model on MedNLI dataset [54] and cosine similarity loss. The MedNLI dataset contains 11,232 training, 1,395 development, and 1,422 test instances. All other training hyper-parameters from Hao et al. [25] were maintained. We applied several metrics to evaluate the model's performance on the sentence-pair similarity task. We only used sentences from the MedNLI dataset and not from the notes of the mini case patient for the evaluation. We used the training and test split as proposed by MedNLI [54] for the evaluation. Compared to other models we tested, such as *biobert-nli*, *bio-clinicalBert*, *PubMedBert*, the fine-tuned *clinical-kb-bert* reached the highest performance, see Table 1.

4.2 Medical Ontology Processing for Concept Space Generation

As described before, we used a medical ontology (UMLS) to build a *backbone* for our analysis. In particular, we extracted concepts related to diseases, drugs, and treatments UMLS vocabularies ICD [62], ATC [61], and CCS [2], respectively. To map these concepts to medical entities found in the patient record data, we used our fine-tuned language model to compute an embedding for each concept to be able to apply similarity functions. The result of this processing pipeline is a dictionary of three concept spaces, whereby each concept space consists of a set of concepts (i.e., medical entities) that are represented by their names (i.e., strings) and corresponding embedding vectors.

4.3 EHR Processing for Patient Medical Entity Extraction

The second processing pipeline is related to the processing of EHR data. First, we applied pre-processing methods to clean the data (removed the anonymized, tagged names) and separate the original text into meaningful sentences using the spaCy [27] library. After the data pre-processing, we used the ScispaCy [3] model to extract a set of medical named entity candidates from sentences. It uses a task-specific model for medically named entity extraction introduced by Lample et al. [37]. We chose this model for the purpose of generalization, i.e., we selected a flexible, fast model with a performance close to state-of-the-art that does not need entity information to identify entities [3]. From ScispaCy we chose the model that has the best performance on a named entity *mention* detection task, independent of the specific entity categories [3]. This allows us to apply the same data processing pipeline to other data than the MIMIC-III dataset. This occurs if we want to extend *MediCoSpace* in the future to, for example, include medical paper abstracts as an extra knowledge

Table 1. We Fine-tuned Multiple State-of-the-art Models on the MedNLI Dataset for the Natural Language Inference Task

	cosine pearson	cosine spearman
biobert-nli	0.79	0.78
bio-clinicalBert	0.77	0.75
PubMedBert	0.76	0.77
clinical-kb-bert	**0.81**	**0.79**

These models learned to predict whether a "hypothesis" is true, false, or undetermined given a "premise." We can understand how well these models perform by evaluating the sentence pair similarity. The sentence similarity task is typically evaluated through different correlation metrics. The evaluation metrics show that *clinical-kb-bert* outperforms other models on this task.

source for doctors to read about new research for certain non-trivial treatments. To be able to link the extracted entity candidates back to the original UMLS concepts, we used our fine-tuned language model to represent each candidate as an embedding vector. Hence, the outcome of this pipeline is a dictionary with medical entity candidates that are represented by their names (i.e., strings) and embedding vectors, as well as the original text and its associated metadata (e.g., time when the record was created).

4.4 Linking Concept Spaces and Patient Medical Entities

Since physicians use synonyms, abbreviations, or misspelled names to refer to concepts in the EHR data, we mapped the medical entity candidates to concepts in the UMLS ontology. We computed the cosine similarity

$$\cos(\mathbf{A},\mathbf{B}) = \frac{\mathbf{A}\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} \mathbf{A}_i\mathbf{B}_i}{\sqrt{\sum_{i=1}^{n}(\mathbf{A}_i)^2}\sqrt{\sum_{i=1}^{n}(\mathbf{B}_i)^2}} \tag{1}$$

between the medically named entity candidate vectors and the UMLS concept vectors. We then linked entity candidates to UMLS concepts whose similarity exceeds 0.9 (we set this threshold after evaluating several manual test iterations). The patient had 1,321 concepts: 970 diseases, 321 drugs, and 30 treatment concepts. The entire population yielded 5,410 concepts (3,841 diseases, 1,462 drugs, and 107 treatment concepts). The reason for computing embedding similarity for entity linking instead of using other entity linking methods (e.g., linking entities to a knowledge base) was the generalizability of the approach. This allows us to apply this method to other data with different named entity categories, without requiring us to adapt the knowledge base. Moreover, it took approximately one hour to compute all concepts for the mini case patient on a conventional laptop.

5 VISUAL ANALYSIS WORKSPACE

We designed *MediCoSpace* using a UCD process [63] and relevant NIST standards for good design principles [63]. We will only describe the most important principles and UCD process steps for our tool. Overall, our tool consists of a limited number of major components (conceptual model and information density principles [63]). First, the general patient information provides basic patient information and a quick summary (clinical decision support—patient information summary principle [63]) (Figure 3(a)). Second, the doctor can record the current note about the patient (Figure 3(b)). Third, the doctor can select interesting concept relations in the co-occurrence heatmap (Figure 3(e)) (R1/4) and find related concepts (R3) in the similarity plots (Figure 3(d)). The note text (Figure 3(f))

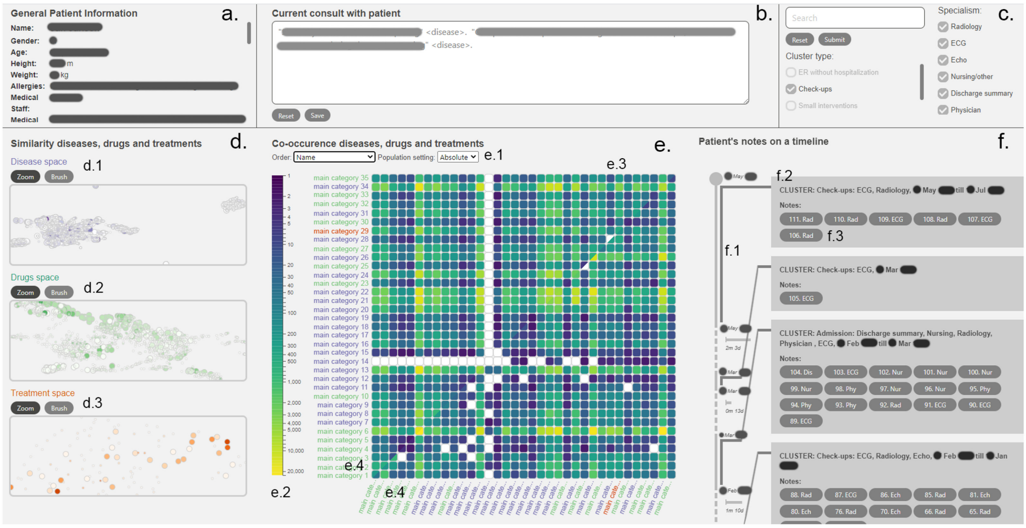


Fig. 3. *MediCoSpace*'s interface with general patient information at the top (a), the current note the doctor is working on (b), and the search and filter options (c). The bottom part consists of similarity plots (d.1–3), a co-occurrence overview heatmap (e) with frequency filter (e.2) and reorder options (e.1), and a timeline (f1) with clusters of notes (f.2–3). Textual information from the data sources is blanked or replaced by "main category [number]", e.g., e.4.

is highlighted (R2/5) based on the selected relation. Finally, the user can save interesting findings back into the current note (R6). These components (placed in this order) help see multiple facets of the disease, drugs, and treatment relations in the patient's history in a single view (clinical decision support—patient history principle [63]), while taking the doctor's workflow (workflow principle [63]) into account. The data from the mini case is displayed in the visualizations.

5.1 Similarity Plots

The similarity plots show similarities between drugs, treatments, and diseases (R3). Since doctors are not used to complex visualizations that are not used in day-to-day clinical practice, we chose simple scatter plot visualizations. Each scatter plot displays either the diseases (Figure 3(d).1), drugs (Figure 3(d).2), or treatments (Figure 3(d).3). Each concept–disease, drug, or treatment–is represented by a circle. We calculated the circles' coordinates using the UMAP [43] dimensionality reduction algorithm on concept embedding vectors, see Section 4. The smaller circles (Figure 4(f)) are concepts that only occur in the population, and the bigger circles (Figure 4(c)) are concepts that occur in the patient and possibly the population. The closer two circles are, the more similar they are. The user can hover over a circle to see the concept name. The darker the color is, the higher the concept occurrence is in the text. The user can zoom in and select concepts (indicated by a pink border) and concepts that co-occurred with these (get a black border) in the notes by brushing, see Figure 4(b), (c). Moreover, *MediCoSpace* links the brushed concepts to highlights in the heatmap (clinical decision support—contextual patient details and visual design—highlight principles [63]), see Section 5.2 and Figure 4(d)–(e). The user can only brush one scatter plot at a time to avoid confusion. The user can clear the brush by clicking a button, see Figure 4(a). We based the colors for the circles and highlighting on ColorBrewer [9] (visual design—color principle [63]).

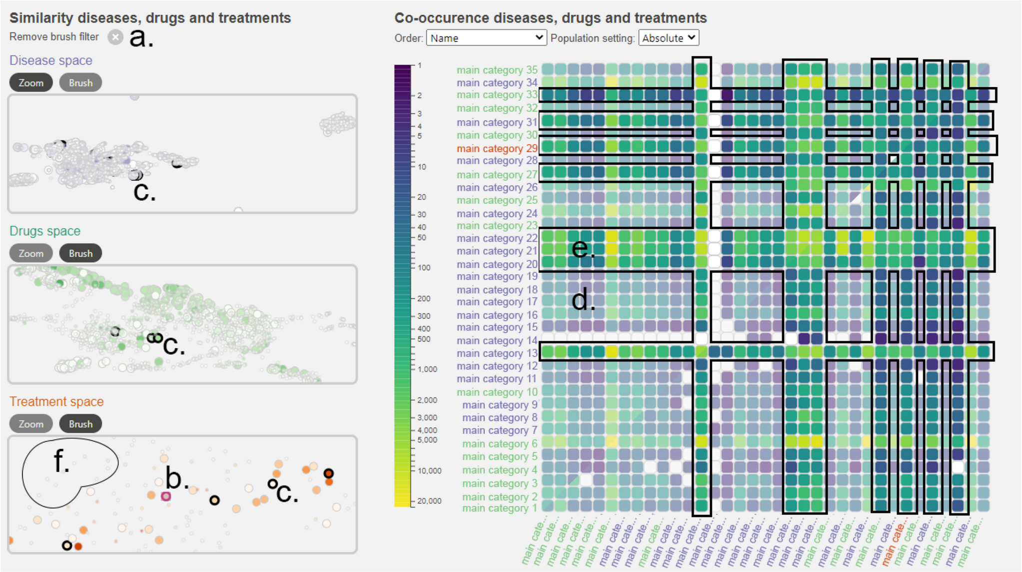


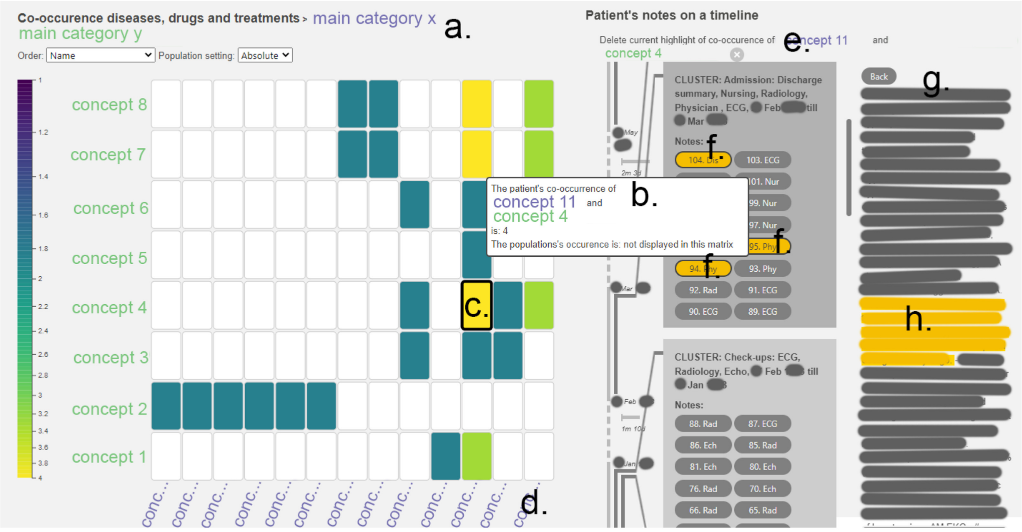
Fig. 4. Connected brushing of the similarity plots. The circles in the brushed area are marked pink (b), and all the circles that co-occurred in the notes with this circle are marked black (c) shows a few examples. Corresponding cells in the heatmap are kept (e), while unrelated cells are lowered in opacity (d). The user can remove the brush by clicking a button (a). Smaller circles only occur in the population (the ellipses surrounding (f) contain a few examples).

5.2 Co-occurrence Heatmap

This central component displays the relations between drugs, diseases, and treatments concepts (requirement R1) occurring in the patient's notes by using the co-occurrence of these concepts in the same notes. For example, if we have concepts c_a, c_b, and c_c, we want to display the co-occurrence of each concept pair in a heatmap matrix. This results in a 3 × 3 matrix with a shortened version of the name of the concept c_a, c_b, and c_c as labels on both axes. The label color depends on the category of the concept (e.g., disease) and is the same as the similarity plot colors. Every concept pair occurs twice in the heatmap (except for the diagonal), once above and below the diagonal in mirrored positions.

If all concept pairs were displayed, the heatmap became too overwhelming for the doctors. Therefore, we used the parent concepts (e.g., a heart disease can have cardiovascular system as parent) from the UMLS dataset to display a heatmap with main categories (overview heatmap) at first (visual design—view simplification, functional grouping, and data visualization principles [63]). Its value is a summation of the co-occurrence of child concept pairs, where one child belongs to one of the two parent concepts and the other child to the other. The cells above the diagonal contain the co-occurrence frequency in the population of similar patients, and the cells below the diagonal display the co-occurrence frequency of the patient, see Figure 6(a). This allows us to compare the concept relations of the patient to the population (requirement R4). Cells on the diagonal are split in half, one for the population and one for the patient.

We used the Viridis color map [13] because it represents the data well and is colorblind-friendly (visual design—color principle [63]), see Figure 3(e).2. We used a logarithmic color scale to better display frequency difference under 100 because these are more important to the doctors than small differences between very high frequencies.

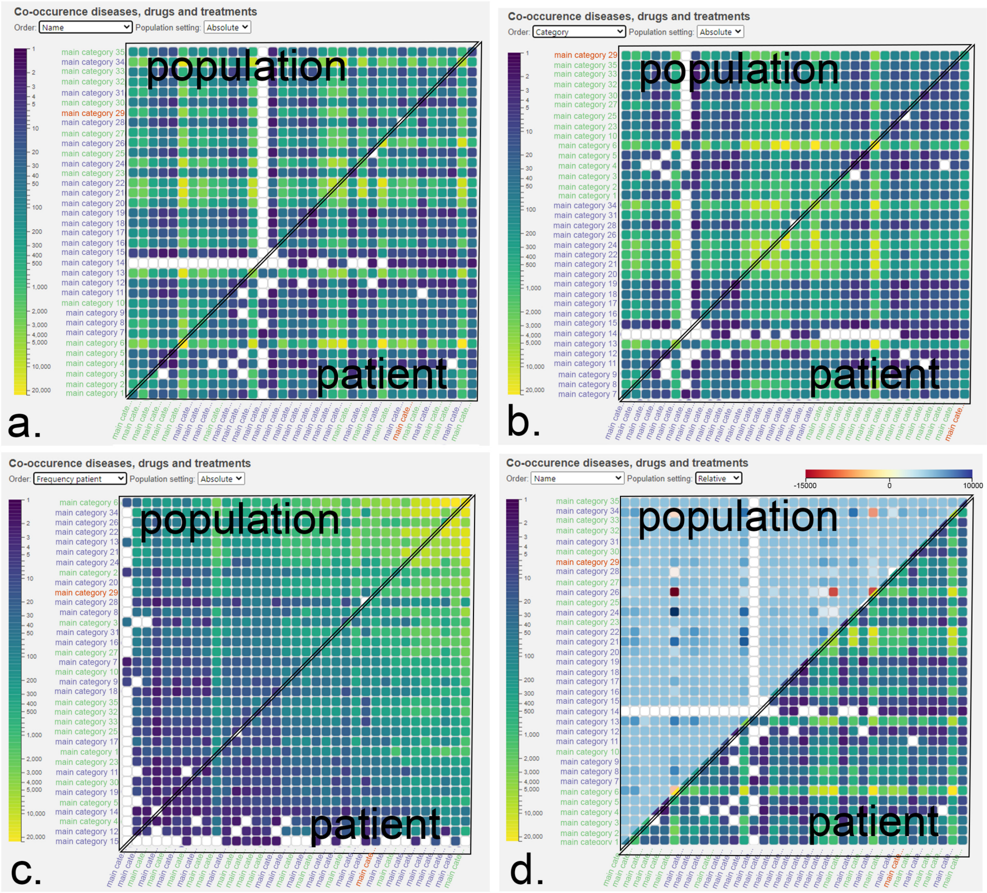


The patient's co-occurrence of concept 11 and concept 4 is: 4
The populations's occurence is: not displayed in this matrix

Fig. 5. When a cell is clicked in the overview heatmap, the child concepts of one parent are displayed as rows and the concepts of the other as columns (d). The parent concepts are displayed as breadcrumbs (a). Hovered cells highlight (c) and trigger a pop-up with the concept name and frequency (b). When a cell is clicked in this detailed heatmap, the co-occurrence of the selected concepts are highlighted in the clusters of notes (f) and in the opened notes (h). The entire note will be displayed (g), and the highlighting can be removed upon clicking a button (e). The actual concept names are replaced by "concept [number]".

MediCoSpace supports several linked interactions, starting with heatmap interactions. If the user clicks on a cell in the overview heatmap, the view changes to a detailed heatmap where the child concepts of one parent are displayed as rows and the concepts of the other as columns (clinical decision support—contextual patient details principle [63]), see Figure 5. The user can hover over a cell to highlight it and see the full name of the concept labels on the axes belonging to that cell and the co-occurrence frequency in the patient and population (error prevention—information suppression principle [63]]), see Figure 5(b). The visualization highlights (visual design—highlight principle [63]) the cell belonging to the same concept pair on the other side of the diagonal (always present in an unfiltered overview heatmap, but not always present in the detailed heatmap). *MediCoSpace* links this to the similarity plots, where the corresponding concepts are highlighted (two circles if a cell is hovered in the detailed heatmap and all the child concept circles of two-parent concepts if a cell is hovered in the overview heatmap). The main parent concepts are displayed as a breadcrumb trail above the heatmap (Figure 5(a)), and by clicking on the title of the heatmap, the user can go back to the overview heatmap. If the user clicks on a cell in this detailed heatmap (Figure 5(c)), the co-occurrence of the concepts this cell represents are highlighted (visual design—highlight principle [63]) in the note clusters and in the notes in yellow (requirements R2 and R5), see (Figure 5(f), (h)).

Second, there are several interactions to filter or reorder views: The heatmap and similarity plots can be filtered based on a range of co-occurrence frequencies using the slider legend on the right, see Figure 7(b), (e). Also, the user can switch the absolute values of co-occurrence frequencies of the population cells to the relative value (frequency concept pair (a, b) population−frequency concept pair (a, b) patient, related to the error prevention—pre-processed information principle [63]), see Figure 6(d). We used a diverging color map with colors not present in the visualization yet. Moreover, the user can reorganize the rows and columns based on alphabetical order of the labels



Co-occurence diseases, drugs and treatments
Order: Name Population setting: Absolute

Co-occurence diseases, drugs and treatments
Order: Category Population setting: Absolute

Co-occurence diseases, drugs and treatments
Order: Frequency patient Population setting: Absolute

Co-occurence diseases, drugs and treatments
Order: Name Population setting: Relative

Fig. 6. Different ordering options for the axis of the heatmaps: alphabetical order (a), concept category (b), or frequency of the patient or population (c). Instead of the absolute values, the relative values of the population compared to the patient can be displayed (d). Population data is displayed above the diagonal and patient data below.

(Figure 6(a)), the category (disease, drug, and treatment) of the labels (Figure 6(b)), and the average co-occurrence frequency of the patient or population (Figure 6(c)).

Third, there are interactions connected to other components (conceptual model—integration principle [63]). If the user brushes concepts in the similarity plots, the cells belonging to these concepts in the detailed heatmap or the parent concepts in the overview heatmap are kept (Figure 4(e), visual design—highlight changes principle [63]), and all other cell fades to a low opacity (Figure 4(d)). Furthermore, the user can save interesting relations by clicking on the corresponding axis label in the heatmap. These labels are added to the current note on the top. The idea is that clicking on these saved relations restores the respective heatmap filter on this information (requirement R6). The user can also select a time period on the timeline, and then only the concepts occurring in that time period are displayed (requirement R1), see Figure 7(c).

We explored design alternatives for the heatmap in previous iterations: the first iteration was a conceptual design [63], which visualized the relations using hierarchical edge bundling. After a cognitive walkthrough [63] and interviews with D1 and three domain experts (analyzed using

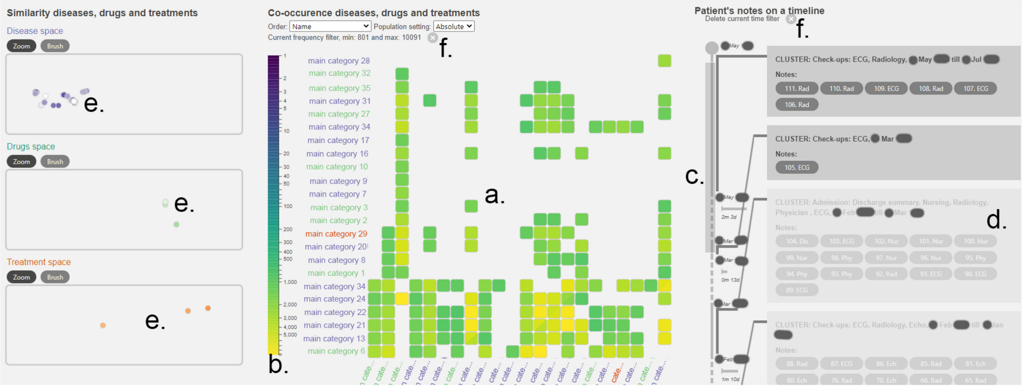


Fig. 7. The user filtered on co-occurrence frequency (b) and time (c). Only the cells (a), concepts (e), and notes (f) inside these ranges are kept. Clusters outside the time period are made transparent (d). Either filter can be removed by clicking a button (f).

a thematic analysis [8]), the results stated that this visualization might be too complex for doctors and chaotic due to the number of bundles. In the second iteration (a detailed design [63]), the current heatmap was split up into different heatmaps; one for all the patient data, one for selected population concepts, and one for specific time periods after each other. After a cognitive walkthrough [63] and interviews with D2 and four domain experts (analyzed using a thematic analysis [8]), the results stated that the second heatmap had low explorative qualities, the third heatmap became too big, and the overview was lost. Therefore, we merged all these features in the current interactive heatmap.

5.3 Timeline with Clusters of Notes

The search/filter options (Figure 3(c)) and the timeline with clusters of notes (Figure 3(f)) were taken from previous work [58] because they provide a clear overview of the patient notes over time with integrated search. The textual notes are clustered around admission or polyclinical check-ups with similar diagnoses. All the notes are displayed as dark-gray buttons (error prevention—information suppression principle [63]) in the cluster, and the user can open them by clicking [58] (clinical decision support—contextual patient details principle [63]). In previous work, the cluster boxes contained word summaries, which were out of the scope of this work and thus removed.

6 EVALUATION: EXPERT USER STUDY

This section describes medical experts' opinions to see if the complex process of designing *MediCoSpace* was worthwhile.

6.1 Methodology

We conducted a formative usability test [63] and used the checklists from Sperrle et al. [55] to evaluate our tool with participants D3-D5, as D1 and D2 were preoccupied due to COVID-19. *MediCoSpace* was displayed locally, and the evaluation lasted one hour. First, we gave an introduction and the participants signed an informed consent form. We recorded the evaluation sessions and transcribed them anonymously, after which we deleted the recordings. We asked the doctors to think aloud to understand their thought process [63]. Second, we asked the doctor to imagine being in a doctor's office of the future with data-driven visualizations being commonplace. Third, we conducted a semi-structured interview about data-driven techniques in healthcare and their assessment of the tool in general. Fourth, we showed a demo of *MediCoSpace*, and fifth, we sketched the mini case, see Section 4. The user evaluation was kept open-ended to give a satisfying representation of insight capabilities [46] and not restrict participants [53]. Sixth, the participant filled in a system usability scale [56, 63] and scored the different visual components present in *MediCoSpace*. Finally, we concluded each session with a semi-structured interview about the usefulness of the tool, (non-)preferred components, and perceived usability.

The analysis consisted of a qualitative (interviews and think aloud protocols) and a quantitative part (the system usability scales, the component scores, and insights measurements [46, 53]). The qualitative parts were analyzed using a thematic analysis [8], where a theme needed the support of all doctors. The insight measurements consisted of the number of insights (distinct observations of the doctor), a hypothesis score (scale: 1 = reading text, 5 = well structured and substantiated hypotheses) to determine how much the insights helped with the forming of hypotheses, time to reach these insights, and if insights were expected or not [46, 53].

6.2 Results

This section discusses the results of the evaluation based on the themes of the thematic analysis, NIST standards, and quantitative results.

6.2.1 Clinical Reasoning and Decision Support. The first theme was clinical reasoning and decision support. The quality of the hypotheses and the doctor's reasoning (extracted from the think-aloud protocols and recorded interactions) informed us about the performance of our tool in relation to clinical reasoning and decision support. Based on the different insights over time the hypotheses increased in depth (the hypothesis score of the current insight increased based on information from previous insights). D3 and D4 reached a score of 3.0 and D4 a score of 3.5 at the end of the evaluation. The doctors did not reach the highest score because they still needed to get used to the presented information to form a complete mental model to navigate quickly between the information in the different components (conceptual model principle [63]). However, the average time to find insights was low, 8 seconds. Also, the doctors already found five unexpected insights (see Figure 8) during the short evaluation, from which four helped to give information about relevant non-cardiovascular diseases to form diagnoses.

The doctors used the heatmap, the timeline with notes, filters and general patient information to form the hypotheses. *MediCoSpace* offered the flexibility to complete the workflow of the mini case task in the way the individual doctors preferred (related to the workflow principle [63]). For example, after reading the general patient information, D4 started with the filters and then the heatmap, while D3 and D5 started with the timeline with notes. Also, four of the unexpected insights were found using the heatmap. Furthermore, the similarity plots were barely used during the evaluation and the participants were divided about the usefulness (score of 5.7, see Figure 9).

6.2.2 Information Density, Integration, and Organization. The second theme was: information density, integration, and organization. First, the information density could be improved. Currently, participants scored our tool 6.3, 6, and 5.3 for being an addition, improving the work efficiency (related to workflow principle [63]), and providing a quick overview respectively, see Figure 10. This was influenced by the amount of information displayed in the similarity view (see Section 5.1) and the co-occurrence heatmap (see Section 5.2), and the familiarity of the visualizations (related to the visual design—view simplification and data visualization principles [63]). This was reflected in the scores (between 3.3 and 7.7) for the questions about the usage, see Figure 10.

According to the participants, they wanted more pre-processed information to reduce the cells in the heatmap (error prevention—pre-processed information principle [63]) to filter out unimportant information to get an even more simplified view (visual design—view simplification principle [63])

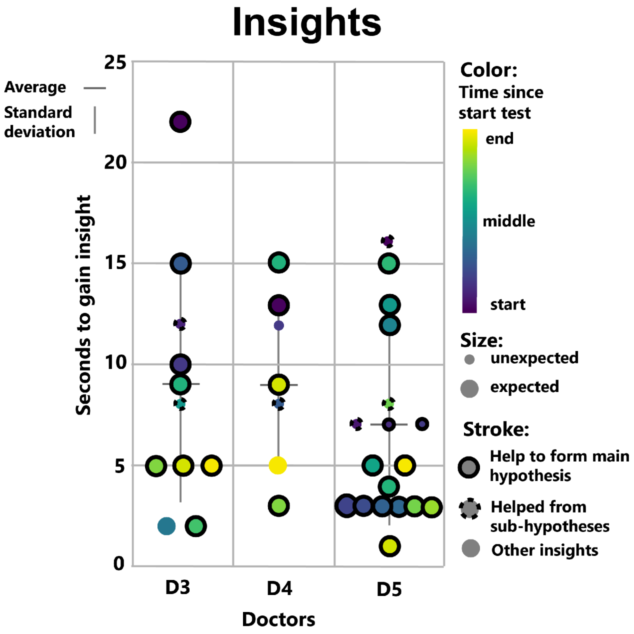


Fig. 8. Insight measurements, where every circle is an insight. The seconds to gain an insight started when the doctor was done with the previously gaining insight. No insight was incorrect.

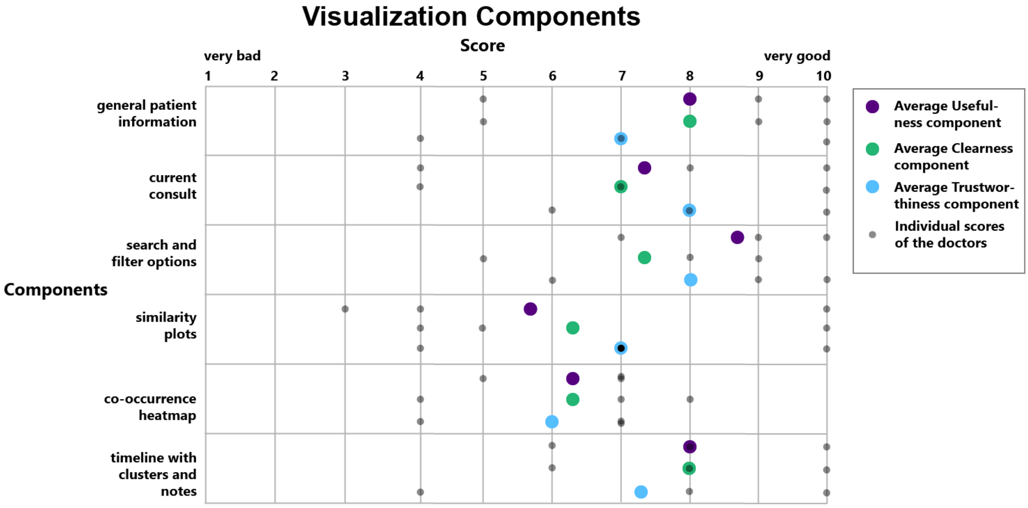


Fig. 9. Component ranking scores. Scores: 1 = very bad; 10 = very good.

to help with the diagnosis process (clinical decision support principle [63]). They proposed to improve this based on the relative frequencies with the population and by filtering out low co-occurrences, concepts not related to their specialty, and outdated notes. Although, this should be done very carefully because the doctor needs to be aware of which information is hidden (related to the error prevention—information suppression principle [63]).

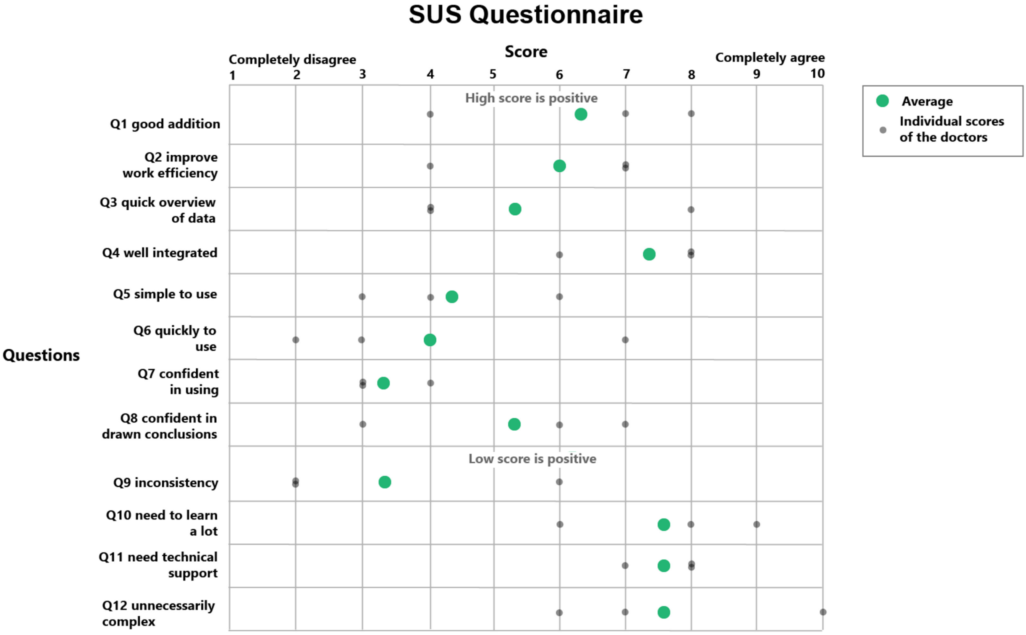


Fig. 10. System usability scale.

Second, participants thought it was well integrated, and there was a low inconsistency (scores of 7.3 and 3.3, respectively, see Figure 10, the conceptual model—integration principle [63]). Also, linking the information in the similarity plots, heatmap, and timeline by highlighting was considered useful (visual design—highlighting principle [63]). Third, the participants also experienced the organization structure and the interactions of the tool as positive and logical.

D3: *"All the EHR systems look very hard in the beginning. But I think people can learn it quickly due to its consistent and organized structure."*

Overall, the participants stated that extracting disease, drugs, and treatments from the text, displaying their relations, and linking them to the notes is beneficial because it is impossible to read all the notes during the short patient consultations (related to the clinical decision support principle [63]). Furthermore, *MediCoSpace* can reduce bias to certain diagnoses (D3), give an overview of the patient narrative (D4), and facilitate communication and collaboration between different specialties (D5):

"Every specialty lives on their island, and there are complaints that fit multiple specialties. Here your idea could help."

Furthermore, the tool could improve the accuracy of the found information (D3), the speed of finding information (D4), the communication between colleagues (D3, D5), and the reliability (D3).

D3: *"This interface is more reliable because all the data is in one place."*

In conclusion, the participants recognized the problem and thought that our tool could help them after implementing some small improvements–despite feeling a bit overwhelmed by the amount of information.

7 DISCUSSION

In this article, we researched a novel approach to medical decision-support for doctors for diagnosing and treating patients. The *MediCoSpace* tool applies visual analytics to show the relations between diseases, treatments, and drugs and link them back to the textual notes from the EHR. The interviewed medical professionals recognized the problems identified in Figure 1 and the need to solve these problems. In general, they found that the idea of showing relations between disease, drugs, and treatments of certain time periods for a single patient is beneficial (requirements R1 and R2) and could improve their diagnosing workflow. Yet, doctors could not form a complete conceptual model of our tool because the visualizations needed less information. This could be accomplished after carefully pruning relations for a simplified view. Therefore, it was not possible yet to form well-substantiated hypotheses. Doctors, however, deemed comparing a single patient to a population of similar patients (requirement R4), linking the relations back to the notes (requirement R5), and saving interesting findings (requirement R6) beneficial to their workflow. The doctors mentioned that our tool could reduce personal biases and stimulate communication between different medical specialties, which were not our initial goals but positive outcomes.

7.1 Limitations

There are several limitations to our study. First, only five doctors were available (due to COVID-19). Moreover, each user evaluation only lasted one hour, again due to the doctors' busy schedules. This resulted in a very short onboarding time. Combined with the complexity and visual density (richness of all the displayed details) of the tool, the doctors mostly could not complete the given scenario and form a complete diagnosis. Currently, following the doctor through a complete workflow cycle would take too long and would require access to their patient's EHR.

7.2 Future Work

In general, more research is needed into using data-centric approaches for medical decision-making for diagnosis and treatment planning. While researchers are making big steps in providing medical data sources and making rich information available to doctors, a crucial counterpart, tools, especially of the visual analytics kinds, are equally important pieces in a larger puzzle. Looking at *MediCoSpace*, small things, such as pre-filtering, can reduce the amount of information displayed in the heatmaps. Also, pop-ups could help point out interesting information in the heatmaps.

On the side of text analysis and processing pipelines, we see tangible future improvements: First, more standardization in recording EHR notes would help to better extract concepts. Doctors use abbreviations, forget punctuation and make spelling mistakes, which is something to take into account in the data preparation. In the future, this could also be done interactively. For example, certain areas of concern could be flagged by an algorithm and medical technicians could correct this. Second, the current relations can also be extended to facilitate interdisciplinary patient treatment by highlighting specific multiple-specialty relations. Furthermore, linking to external data sources can help increase the confidence in doctors' diagnoses and potentially make treatment plans more reliable. For example, *MediCoSpace* can link highlighted relations to the latest scientific articles to provide up-to-date information. On the visualization side, more research is needed to validate if displaying the similarity between disease, drugs, and treatments is useful (requirement R3).

Moreover, additional research is needed in data security; when a new patient or population notes are added, the updated data structure needs to be stored and accessed safely. From a usability point of view, we need to ensure that no hazardous errors occur. Healthcare technology ISO standards [32] could help with the above and a summative usability test [63] could help to evaluate the production-ready tool.

In conclusion, we present our visual analytics tool, not as a final answer to a growing problem space, but to open many other directions to explore in this research area.

8 CONCLUSION

This article describes a novel way to provide decision-support for doctors to diagnose and devise treatments for patients by extracting and showing relations between disease, drug, and treatments of a single patient over time. This is coupled with the ability to visually compare individual patients to a population of similar patients by means of medical concepts extracted from their EHR information. We identified the doctor's workflow and limitations of current systems; based on these and interviews, we researched how visual analytics could help. Our tool, *MediCoSpace*, both in the design process and validation with medical professionals provides us with insights into the information need of doctors about these relations and showed the potential to improve the communication between doctors, reduce personal bias and get a more holistic view of the patient. *MediCoSpace* is a contribution to a growing field of data-driven machine-learning supported medical decision-making approaches that aim at improving medical care and easing the pressure on worldwide healthcare systems.

The visualization JavaScript files can be found here: https://github.com/SannevdLinden/MediCoSpace.

ACKNOWLEDGMENTS

We want to thank Fabian Viehmann; the data analysis section is based on his master graduation project. Moreover, we want to thank all the participants for their time and valuable feedback.

REFERENCES

[1] Julia Adler-Milstein, Jonathan H. Chen, and Gurpreet Dhaliwal. 2021. Next-generation artificial intelligence for diagnosis: From predicting diagnostic labels to "wayfinding". *JAMA* 326, 24 (2021), 2467–2468. DOI: https://doi.org/10.1001/jama.2021.22396

[2] AHRQ. 2022. *Clinical Classifications Software (CCS) for ICD-9-CM Fact Sheet*. Retrieved Jan 14, 2022 from https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccsfactsheet.jsp.

[3] Allen institute of artificial intelligence. 2020. *Scispacy*. Retrieved Apr 9, 2021 from https://allenai.github.io/scispacy/.

[4] Erin P. Balogh, Bryan T. Miller, John R. Ball, and editors. 2015. Committee on diagnostic error in health care; Board on health care services; institute of medicine; The national academies of sciences, engineering, and medicine. In *Proceedings of the Improving Diagnosis in Health Care*.

[5] Sergio Baranzini, Sui Huang, Sharat Israni, and Mike Keiser. 2022. *WHAT is SPOKE?* Retrieved Jan 11, 2022 from https://spoke.ucsf.edu/.

[6] Vijayaraghavan Bashyam, William Hsu, Emily Watt, Alex A. T. Bui, Hooshang Kangarloo, and Ricky K. Taira. 2009. Problem-centric organization and visualization of patient imaging and clinical data. *Radiographics* 29, 2 (2009), 331–343. DOI: https://doi.org/10.1148/rg.292085098

[7] Olivier Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Research* 32, suppl_1 (2004), D267–D270. DOI: https://doi.org/10.1093/nar/gkh061

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. DOI: https://doi.org/10.1191/1478088706qp063

[9] Cynthia Brewer, Mark Harrower, Ben Sheesley, Andy Woodruff, and David Heyman. 2013. *COLORBREWER 2.0 Color Advice for Cartography*. Retrieved October 11, 2021 from https://colorbrewer2.org/#.

[10] Alex A. T. Bui, Denise R. Aberle, and Hooshang Kangarloo. 2007. TimeLine: Visualizing integrated patient records. *IEEE Transactions on Information Technology in Biomedicine* 11, 4 (2007), 462–473. DOI: https://doi.org/10.1109/TITB.2006.884365

[11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* Jill Burstein, Christy Doran, and Thamar Solorio (Eds.), Association for Computational Linguistics, 4171–4186. DOI: https://doi.org/10.18653/v1/n19-1423

[12] Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Hannaneh Hajishirzi, and Noah Smith. 2020. Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. arXiv:2002.06305. Retrieved from https://arxiv.org/abs/2002.06305.

[13] Enthought. 2015. *A Better Default Colormap for Matplotlib | SciPy 2015 | Nathaniel Smith and StÃľfan van der Walt.* Retrieved October 11, 2021 from https://www.youtube.com/watch?v=xAoljeRJ3lU.

[14] Epic. 2022. *Epic with the Patient at Heart.* Retrieved Jan 14, 2022 from https://www.epic.com/.

[15] Hossein Estiri and Kari Stephens. 2017. DQe-v: A database-agnostic framework for exploring variability in electronic health record data across time and site location. *eGEMs* 5, 1 (2017), 1–10. DOI: https://doi.org/10.13063/2327-9214.1277

[16] William R. H. Evans. 2018. Dare to think rare: Diagnostic delay and rare diseases. *The British Journal of General Practice* 68, 670 (2018), 224–225. DOI: https://doi.org/10.3399/bjgp18X695957

[17] Sarah Faisal, Ann Blandford, and Henry W. W. Potts. 2013. Making sense of personal health information: Challenges for information visualization. *Health Informatics Journal* 19, 3 (2013), 198–217. DOI: https://doi.org/10.1177/1460458212465213

[18] Hamish Fraser, Enrico Coiera, and David Wong. 2018. Safety of patient-facing digital symptom checkers. *The Lancet* 392, 10161 (2018), 2263–2264. DOI: https://doi.org/10.1016/S0140-6736(18)32819-8

[19] Stephen Gilbert, Alicia Mehl, Adel Baluch, Caoimhe Cawley, Jean Challiner, Hamish Fraser, Elizabeth Millen, Maryam Montazeri, Jan Multmeier, Fiona Pick, Claudia Richter, Ewelina Türk, Shubhanan Upadhyay, Vishaal Virani, Nicola Vona, Paul Wicks, and Claire Novorol. 2020. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. *BMJ Open* 10, 12 (2020), e040269. DOI: https://doi.org/10.1136/bmjopen-2020-040269

[20] Benjamin S. Glicksberg, Boris Oskotsky, Phyllis M. Thangaraj, Nicholas Giangreco, Marcus A. Badgeley, Kipp W. Johnson, Debajyoti Datta, Vivek A. Rudrapatna, Nadav Rappoport, Mark M. Shervey, Riccardo Miotto, Theodore C. Goldstein, Eugenia Rutenberg, Remi Frazier, Nelson Lee, Sharat Israni, Rick Larsen, Bethany Percha, Li Li, Joel T. Dudley, Nicholas P. Tatonetti, and Atul J. Butte. 2019. PatientExploreR: An extensible application for dynamic visualization of patient clinical history from electronic health records in the OMOP common data model. *Bioinformatics* 35, 21 (2019), 4515–4518. DOI: https://doi.org/10.1093/bioinformatics/btz409

[21] David Gotz and Harry Stavropoulos. 2014. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1783–1792. DOI: https://doi.org/10.1109/TVCG.2014.2346682

[22] David Gotz and Krist Wongsuphasawat. 2012. Interactive intervention analysis. In *Proceedings of the American Medical Informatics Association Annual Symposium.* 274–280.

[23] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao, and Hoifung Poon. 2021. Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare* 3, 1 (2021), 23 pages. DOI: https://doi.org/10.1145/3458754

[24] Catalina Hallett. 2008. Multi-modal presentation of medical histories. In *Proceedings of the 13th International Conference on Intelligent User Interfaces.* 80–89. DOI: https://doi.org/10.1145/1378773.1378785

[25] Boran Hao, Henghui Zhu, and Ioannis Paschalidis. 2020. Enhancing clinical BERT embedding using a biomedical knowledge base. In *Proceedings of the 28th International Conference on Computational Linguistics.* International Committee on Computational Linguistics, 657–661. DOI: https://doi.org/10.18653/v1/2020.coling-main.57

[26] Jamie S. Hirsch, Jessica S. Tanenbaum, Sharon Lipsky Gorman, Connie Liu, Eric Schmitz, Dritan Hashorva, Artem Ervits, David Vawdrey, Marc Sturm, and Noémie Elhadad. 2015. HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association* 22, 2 (2015), 263–274. DOI: https://doi.org/10.1136/amiajnl-2014-002945

[27] Matthew Honnibal. 2022. *Industrial Strength Natural Language Processing.* Retrieved Apr 9, 2021 from https://spacy.io/.

[28] William Hsu, Ricky K. Taira, Suzie El-Saden, Hooshang Kangarloo, and Alex A. T. Bui. 2012. Context-based electronic health record: Toward patient specific healthcare. *IEEE Transactions on Information Technology in Biomedicine* 16, 2 (2012), 228–234. DOI: https://doi.org/10.1109/TITB.2012.2186149

[29] Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. 2019. ClinicalBERT: Modeling clinical notes and predicting hospital readmission. (CHIL 2020 Workshop). Retrieved Mar 20, 2021 from https://github.com/kexinhuang12345/clinicalBERT.

[30] Cinyoung Hur, JeongA Wi, and YoungBin Kim. 2020. Facilitating the development of deep learning models with visual analytics for electronic health records. *International Journal of Environmental Research and Public Health* 17, 22 (2020), 8303. DOI: https://doi.org/10.3390/ijerph17228303

[31] Fahd Husain, Rosa Romero-Gómez, Emily Kuang, Dario Segura, Adamo Carolli, Lai Chung Liu, Manfred Cheung, and Yohann Paris. 2021. A multi-scale visual analytics approach for exploring biomedical knowledge. In *Proceedings of the 2021 IEEE Workshop on Visual Analytics in Healthcare.* IEEE, 30–35. DOI: https://doi.org/10.1109/VAHC53616.2021.00010

[32] ISO. 2022. 35.240.80 - IT applications in health care technology. Retrieved June 10, 2022 from https://www.iso.org/ics/35.240.80/x/.

[33] Lotte Groth Jensen and Claus Bossen. 2016. Factors affecting physicians' use of a dedicated overview interface in an electronic health record: The importance of standard information and standard documentation. *International Journal of Medical Informatics* 87, 2016 (2016), 44–53. DOI: https://doi.org/10.1016/j.ijmedinf.2015.12.009

[34] Zhuochen Jin, Shunan Guo, Nan Chen, Daniel Weiskopf, David Gotz, and Nan Cao. 2020. Visual causality analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 1343–1352. DOI: https://doi.org/10.1109/TVCG.2020.3030465

[35] Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-Wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo A. Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. *Scientific Data* 3, 1 (2016), 1–9. DOI: https://doi.org/10.1038/sdata.2016.35

[36] Josua Krause, Adam Perer, and Harry Stavropoulos. 2015. Supporting iterative cohort construction with visual temporal queries. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 91–100. DOI: https://doi.org/10.1109/TVCG.2015.2467622

[37] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 260–270. DOI: https://doi.org/10.18653/v1/N16-1030

[38] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (2020), 1234–1240. DOI: https://doi.org/10.1093/bioinformatics/btz682

[39] Jingjing Li, Kai Larsen, and Ahmed Abbasi. 2020. TheoryOn: A design framework and system for unlocking behavioral knowledge through ontology learning. *MIS Quarterly* 44, 4 (2020), 1733–1772. DOI: https://doi.org/10.25300/MISQ/2020/15323

[40] Rui Li, Changchang Yin, Samuel Yang, Buyue Qian, and Ping Zhang. 2020. Marrying medical domain knowledge with deep learning on electronic health records: A deep visual analytics approach. *Journal of Medical Internet Research* 22, 9 (2020), e20645. DOI: https://doi.org/10.2196/20645

[41] Daniel Loureiro and Alípio Mário Jorge. 2020. MedLinker: Medical entity linking with neural representations and dictionary matching. *Advances in Information Retrieval* 12036 (2020), 230–237. DOI: https://doi.org/10.1007/978-3-030-45442-5_29

[42] Lena Mamykina, Stuart Goose, David Hedqvist, and David Beard. 2004. CareView: Analyzing nursing narratives for temporal trends. In *Proceedings of the CHI'04 extended abstracts on Human Factors in Computing Systems*. 1147–1150. DOI: https://doi.org/10.1145/985921.986010

[43] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. 2018. UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software* 3, 29 (2018), 861. DOI: https://doi.org/10.21105/joss.00861

[44] Sunil Mohan and Donghui Li. 2019. Medmentions: A large biomedical corpus annotated with umls concepts. In *Proceedings of the 2019 Conference on Automated Knowledge Base Construction (AKBC'19)*. DOI: 10.24432/C5G59C

[45] David E. Newman-Toker, Kathryn M. McDonald, and David O. Meltzer. 2013. How much diagnostic safety can we afford, and how should we decide? A health economics perspective. *BMJ Quality and Safety* 22, Suppl 2 (2013), ii11–ii20. DOI: https://doi.org/10.1136/bmjqs-2012-001616

[46] Chris North. 2006. Toward measuring visualization insight. *IEEE Computer Graphics and Applications* 26, 3 (2006), 6–9. DOI: https://doi.org/10.1109/MCG.2006.70

[47] J. Marc Overhage and David McCallie Jr. 2020. Physician time spent using the electronic health record during outpatient encounters: A descriptive study. *Annals of Internal Medicine* 172, 3 (2020), 169–174. DOI: https://doi.org/10.7326/M18-3684

[48] Heekyong Park and Jinwook Choi. 2012. V-model: A new innovative model to chronologically visualize narrative clinical texts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 453–462. DOI: https://doi.org/10.1145/2207676.2207739

[49] Rimma Pivovarov and Noémie Elhadad. 2015. Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association* 22, 5 (2015), 938–947. DOI: https://doi.org/10.1093/jamia/ocv032

[50] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J. Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Yi Zhang, Gerardo Flores, Gavin E. Duggan, Jamie Irvine, Quoc Le, Kurt Litsch, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, Dana Ludwig, Samuel L. Volchenboum, Katherine Chou, Michael Pearson, Srinivasan Madabushi, Nigam H. Shah, Atul J. Butte, Michael D. Howell, Claire Cui, Greg S. Corrado, and Jeffrey Dean. 2018. Scalable and accurate deep learning for electronic health records. *npj Digital Medicine* 1 (2018), Article No. 18. DOI: https://doi.org/10.1038/s41746-018-0029-1

[51] Adam Rule, Steven Bedrick, Michael F. Chiang, and Michelle R. Hribar. 2021. Length and redundancy of outpatient progress notes across a decade at an academic medical center. *JAMA Network Open* 4, 7 (2021), e2115334–e2115334. DOI: https://doi.org/10.1001/jamanetworkopen.2021.15334

[52] Alberto Santos, Ana R. Colaço, Annelaura B. Nielsen, Lili Niu, Philipp E. Geyer, Fabian Coscia, Nicolai J. Wewer Albrechtsen, Filip Mundt, Lars Juhl Jensen, and Matthias Mann. 2020. Clinical knowledge graph integrates proteomics data into clinical decision-making. *bioRxiv* (2020), 1–35. `DOI:`https://doi.org/10.1101/2020.05.09.084897

[53] Purvi Saraiya, Chris North, and Karen Duca. 2005. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics* 11, 4 (2005), 443–456. `DOI:`https://doi.org/10.1109/TVCG.2005.53

[54] Chaitanya Shivade. 2019. *MedNLI - A Natural Language Inference Dataset For The Clinical Domain*. Retrieved Apr. 3, 2021 from https://physionet.org/content/mednli/1.0.0/.

[55] Fabian Sperrle, Mennatallah El-Assady, Grace Guo, Rita Borgo, Duen Horng Chau, Alex Endert, and Daniel Keim. 2021. A survey of human-centered evaluations in human-centered machine learning. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 543–568. `DOI:`https://doi.org/10.1111/cgf.14329

[56] Nicole Sultanum, Michael Brudno, Daniel Wigdor, and Fanny Chevalier. 2018. More text please! Understanding and supporting the use of visualization for clinical text overview. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13. `DOI:`https://doi.org/10.1145/3173574.3173996

[57] Cong Sun, Zhihao Yang, Lei Wang, Yin Zhang, Hongfei Lin, and Jian Wang. 2021. Biomedical named entity recognition using BERT in the machine reading comprehension framework. *Journal of Biomedical Informatics* 118, 2021 (2021), 103799. `DOI:`https://doi.org/10.1016/j.jbi.2021.103799

[58] Sanne van der Linden, Jarke J. van Wijk, and Mathias Funk. 2021. Multiple scale visualization of electronic health records to support finding medical narratives. In *Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine*. `DOI:`https://doi.org/10.2312/vcbm.20211351

[59] Gautam Verma, Alexander Ivanov, Francis Benn, Anil Rathi, Nathaniel Tran, Ashwad Afzal, Parag Mehta, and John F. Heitner. 2020. Analyses of electronic health records utilization in a large community hospital. *PloS One* 15, 7 (2020), e0233004. `DOI:`https://doi.org/10.1371/journal.pone.0233004

[60] Qiru Wang and Robert S. Laramee. 2022. EHR STAR: The state-of-the-art in interactive EHR visualization. In *Proceedings of the Computer Graphics Forum*. Wiley Online Library, 69–105. `DOI:`https://doi.org/10.1111/cgf.14424

[61] WHO. 2022. *Anatomical Therapeutic Chemical (ATC) Classification*. Retrieved Jan 14, 2022 from https://www.who.int/tools/atc-ddd-toolkit/atc-classification.

[62] WHO. 2022. *International Statistical Classification of Diseases and Related Health Problems (ICD)*. Retrieved Jan 14, 2022 from https://www.who.int/classifications/classification-of-diseases.

[63] Michael E. Wiklund, Jonathan Kendler, Limor Hochberg, and Matthew B. Weinger. 2015. Technical basis for user interface design of health IT. *Grant/Contract Reports (NISTGCR), National Institute of Standards and Technology, Gaithersburg, MD*. `DOI:`https://doi.org/10.6028/NIST.GCR.15-996

[64] Krist Wongsuphasawat, John Alexis Guerra Gómez, Catherine Plaisant, Taowei David Wang, Meirav Taieb-Maimon, and Ben Shneiderman. 2011. LifeFlow: Visualizing an overview of event sequences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1747–1756. `DOI:`https://doi.org/10.1145/1978942.1979196

[65] Tianyi Zhang, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Elena Glassman. 2021. Interactive cohort analysis and hypothesis discovery by exploring temporal patterns in population-level health records. In *Proceedings of the 2021 IEEE Workshop on Visual Analytics in Healthcare*. IEEE, 14–18. `DOI:`https://doi.org/10.1109/VAHC53616.2021.00007

Received 14 January 2022; revised 29 June 2022; accepted 3 September 2022