medical sciences forum

MDPI

Proceeding Paper

Modeling and visualization of clinical texts to enhance meaningful and user-friendly information retrieval

Jonah Kenei[1,*], Prof. Elisha Opiyo

University of Nairobi: opiyo@uonbi.ac.ke

* Correspondence: Jonah.kenei@gmail.com; Tel.: (+254-(0)722400542)

† Presented at the the 2nd International Electronic Conference on Healthcare, 17 February–3 March 2022. Available online: https://iech2022.sciforum.net/.

Abstract: Access to digital health data collections such as clinical notes, discharge summaries or medical charts in have increased in the last few years due to increase use of electronic health records that provide instant access to patients' clinical information. The volume and the unstructured nature of these datasets present great challenges in analyses and subsequent applications to healthcare. The growing volume of clinical data generated and stored in electronic health records creates challenges for physicians when reviewing patients' records with the aim of understanding individual patients' health histories. Electronic healthcare records contain large volumes of unstructured data which requires one to read through to get the required information. This is a challenging task due to lack of suitable techniques to quickly extract needed information. Information processing tools in clinical domain that provide support to users in seeking needed information are lacking. The use of data visualizations has been introduced in an attempt to solve this problem; however, no single approach has been widely adopted. In this paper we propose a unique approach for modeling clinical notes using semantics of various units of a clinical text document to aid doctors in reviewing electronic clinical notes. This is achieved by applying supervised machine learning technique to identify and present semantically similar information together, facilitating the identification of relevant information to users.

Keywords: Electronic health record; classification; Clinical notes; visualization

Citation: Kenei, J.; Opiyo, E.; Machii, J. Modeling and visualization of clinical texts to enhance meaningful and user-friendly information retrieval. *Med. Sci. Forum* **2022**, *1*, x. https://doi.org/10.3390/xxxxx

Published: date

1. Introduction

Advances in digital healthcare technologies, such as telemedicine, biosensors and electronic health records are reshaping the future of healthcare delivery. The exponential growth of health-care data, such as sensor data from intensive care units (ICU), data generated in telemedicine, longitudinal data from electronic health records (EHR) and other sources are opening up new avenues for leveraging data-driven techniques such as machine learning (ML)[1, 2, 3, 4, 5, 6, and 7], artificial intelligence techniques on data retrieved from wearable health sensors [8] and data from telemedicine [9] and artificial intelligence (AI) to exploit this data. Electronic health records are becoming commonly used to document and store patient patients' health records. The primary purpose of patients' medical records is to support clinical decision-making and continuity of care by providing readily accessible medical information [10]. The overwhelming growth and ease of access to digital clinical data in electronic health records has fueled research efforts that aim at helping physicians make use of the growing digital information. Now-a-days electronic health records are being used by numerous healthcare facilities which not only provides huge amount of information available in electronic health records [11] but also presents challenges using the same information [12, 13]. This is due to large volume of clinical narrative texts which needs to be read and understood in order to provide effective patient care. Physicians often rely on patient's health history which makes it difficult to locate information in patient's health history [14] written in narrative text. Despite the fact that they were introduced with the hopes of saving time and improving patient care quality, physicians frequently spend more time navigating these records at the expense of interacting with patients. Clinical narratives represent the main form of documentation in health care, generating patients' clinical histories with detailed clinical information that support clinical decision making [15]. Clinical narratives are commonly entered, captured and stored in electronic health records in digital form [16] and is the most preferred method for recording clinical information [17], [18]. Easy access to the available clinical information and the ability to use it are important in providing better patient care [19] and research efforts in the literature are trying to find ways to present clinical data in forms that are easier to use [20,21]. Most of the research efforts in this direction include ways to structure and present information to physicians to aid in their decision-making [22]. Unstructured clinical narratives are continuously being recorded as part of patient care in electronic health records [23]. Healthcare facilities deal with large volumes of unstructured clinical texts such as clinical notes on a daily basis. With the availability of information in electronic health records, analysis of clinical narratives becomes increasingly important as it contains useful information about patients and their health [24] and therefore represents a significant and important source of clinical information. Using narrative text is still the most natural way to express medical information; however, it is still less amenable to computational techniques. In addition, the abundance of patient clinical records being generated has also raised concerns of information overload [25] with potential negative consequences in medical practice [26]. On the other hand, the availability of digital health records provides an opportunity to develop computational tools to extract medical knowledge [27]. The primary use of electronic health records is to support patient care [28] and also for secondary purposes [29, 30] such as clinical research [30]. When it comes to analysing huge amounts of clinical texts, it becomes too challenging to do it manually. Again, manual review of large amounts of documentation is more likely to lead to errors. The increasing amount of information available in electronic health records [31, 32] makes it challenging for physicians to quickly locate needed information for patient care [33] that is critical to developing an appropriate assessment and plan for the individual patient [34].

Using Electronic health records (EHRs) allows healthcare facilities to store and retrieve detailed patient clinical records which can be used by physicians, during care episodes [35]. However, with the increasing availability of clinical data, data retrieval becomes more difficult, leading to cognitive load and clinician burnout [36]. In most cases it remains underutilized in clinical practice due to lack of suitable techniques to timely extract needed information [37]. Currently, there is readily availability of information in electronic health records in narrative text and there is need of automated techniques to process such texts [38]. Recent research has shown that electronic health records (EHRs) that process, organize, and visualize clinically meaningful information significantly reduce physician cognitive workload [39].

Clinical records are used by doctors to make informed decisions at the point of care [40]. However, as the volume of clinical records along with time constraints inherent in healthcare setting increases, utilizing these records becomes challenging [41] and time-consuming [42]. Information stored in clinical documents is difficult to review since it requires more time to read to get the information that forms the basis of clinical decision making. This is even more challenging especially with the prevalence of chronic diseases in our contemporary society where patients are monitored over a long period of time [43]. In such cases, a physician may need to have an overview summary on the progress and changes in the patient health history that have taken place. So, physicians as well as researchers have to spend more time to analyze patients' health records [44]. To tackle this problem, data visualization techniques have been employed, to help physicians extract relevant valuable information and to reduce cognitive overload [44].

In this paper we describe a prototype for visually modeling clinical notes into semantic units with the objective of supporting healthcare delivery. Our objective is to propose a technique to improve physician's ability to retrieve key concepts relevant to patient at hand.

2. Background

Electronic Health Records Systems (EHRs) are becoming common in many healthcare establishments replacing traditional paper records [45]. They are used to create patient's health records during clinical encounters with patients in a healthcare facility [45]. The records contain patient demographics, progress notes, and medication history [36] offering important clinical information for care of patients and supporting other functions such as interoperability [46]. The information generated during clinical encounters is stored and maintained in electronic health records in order to take care of patients and follow-up [47]. Therefore, they are sources of important clinical information [48]; however, most of the documentation is in unstructured narrative text which is time-consuming to review manually [49]. While there are structured patient's data in electronic health records, important patient's clinical information which describes patient's care and management remain buried in narrative text, making it challenging and time-consuming for physicians to review during their usual medical practice [50]. Unstructured data refers to information that does not conform to a predefined model making it difficult to be processed using computer systems. In healthcare, this is mostly clinical narratives which constitute the bulk of clinical documentation. This unstructured clinical data is constantly increasing and the capacity of physicians to read and analyze this data remains the same. Physicians use narrative text to document essential clinical information during clinical encounters; however, it increases the workload of reviewing it during patient's subsequent visits [50].

A lot of documentation in form of clinical text is captured in electronic health records, often in a notes section [51]. During care episodes, doctors rely on available clinical documentation on which they base their decisions on, in order to provide effective patient care. During a typical clinical encounter, physicians create and add to the patients' medical record with a variety of clinical information which means large amount of data gets generated every time a patient visits the healthcare facility for diagnosis and treatment. Increasing volume of clinical data mostly in unstructured form can lead to information overload for healthcare practitioners. As the volume of clinical data grows large, it becomes increasingly difficult to browse and review a patient medical history. Therefore, there is a challenge of how to unlock unstructured clinical data to improve patient care. There is need therefore to aid them by providing systems which automate narrative texts processing. To provide high-quality and safe care, physicians must be able to distill and easily use the available clinical information. To facilitate, the use of the available clinical information in electronic health records, there is need of organizing and presenting key patient information in a convenient way. The goal of this project is to create a classification algorithm and supporting visualization system for automatically presenting medical information in order to assist physicians.

2.1 Clinical significance

Health information has become readily available and accessible through computers, and technology is becoming an integral part of healthcare ecosystem. Much of the health information that was previously available only in paper records is now available in digital form and directly accessible to healthcare professionals. Consequently, physicians often navigate vast amounts of health information on their own, typically with little support on how to retrieve the available information. Again, the already available information remains not fully utilized [52] and widespread problems due to lack of suitable techniques to extract needed clinical information [53] have been noted. In addition to lack of appropriate techniques to support retrieval of needed clinical information, the problem of information overload [54] is contributing to the difficulty of using this information.

Patients' clinical records are needed for a variety of reasons. Physicians usually examine a patient medical record in order to get information that will allow them to make informed decisions regarding a particular patient or case. This entails gleaning the complete medical history, spotting important information, noting trends, cause-effect relationships, or reviewing past medical history [55]. The need for new computational techniques for growing volume of clinical data in digital form [56] and mostly in unstructured narrative form [57] cannot be under estimated for example: need for information retrieval from clinical notes in order to provide effective medical care [58], generating clinical summaries from clinical texts [59] with important information relevant to a particular patient or accurate task-specific clinical summaries [60]. In our contemporary society, we are faced with the challenge of chronic diseases, which accumulates a large volume of patient data collected over a long period of time that needs attention of physicians. The use of visualization techniques has the potential of aiding such tasks thus improving health care [61]. Based on the above, we believe that, there is need of best approach for easing the burden of using the available clinical information in electronic health records which is mostly available in unstructured narrative form. Without this, physicians are vulnerable to acting on inaccurate or incomplete health information and jeopardizing healthcare decisions.

Therefore, our main objective was to design and develop a visualization tool to help physicians retrieve and visualize unstructured narrative texts in electronic health records by providing easy means to retrieve and visually review such datasets and, thus support them in making clinical decisions. The tool is particularly designed to provide relevant information for physicians with respect to a patient at hand. As a starting point, we consider clinical notes written using SOAP (Subjective, Objective, Assessment and Plan) documentation format. But a future goal is to explore how such visualization technique can also be extended to other documentation formats. The goal of this paper was to propose a technique for organizing and visualizing clinical narrative documents into predefined semantic groups. For this purpose, supervised machine learning technique was applied on clinical narrative data sets.

3. Related Works

The literature relevant to our work is divided into three distinct study areas: topic modeling, data visualization techniques and information retrieval.

3.1. Topic modelling

A lot of study has been done in the subject of topic modeling because of the vast quantity of text documents that are becoming available. Topic modeling is an unsupervised learning approach for discovering topics in a collection of documents. They are often used to extract the main topics that represent the information covered by a given text document, thus tackling information discovery challenges. Application of topic models to clinical narrative data sets is becoming increasingly popular. However, there's been little effort to adapt these models to clinical practice. From the literature, there are a number of topic models that are commonly used. This includes LDA (Latent Dirichlet Allocation) [62].LSA (Latent Semantic Analysis) [63] and PLSA (Probabilistic Latent Semantic Analysis) [64]. For modelling clinical notes, the majority of previous methods have used latent topic models for various applications. For example, the authors in [65] made use of topic modeling to explore electronic health records. There are several other research works reported in the literature that uses topic models to solve the problem of finding themes in electronic health records. Examples include mining cancer clinical notes [66], comparing patients' notes to the subjects discovered [67], grouping discharge summaries into hierarchical concepts [68] and identifying the most relevant subjects [69]. These publications, however, do not address the identification of the most common issues on which they focus.

3.2. Data visualization modelling

Data visualization is becoming increasingly important for analyzing large volumes of complex data [70]. There are many techniques that have been proposed for text visualization in the clinical domain [71, 72] as well as in the general domain. This is largely driven by the need to improve the efficacy and utility of collected data in electronic health records [71]. In clinical domain, the visualization strategies are meant to aid in understanding clinical data [72]. In the general domain, there has been a lot of work in automated text visualization, such as visualization of news [73-75]. There are several research works that have been carried out and many authors have proposed a variety of visualization techniques [76]. One of the most prevalent techniques uses the original concept of TimeLine [77]. Examples falling under this category include; Lifelines [78], Lifelines2 [79, 80] KNAVE II [81], CLEF Visual Navigator [82] and AsbruView [83]. The focus of these techniques is to visualize clinical data based on time. Thus, we can refer to as Time-based visualizations graphical representations of data collected over time. Other techniques include LifeFlow [84] and EventFlow [85]. Unlike the above techniques, these two techniques do not use a timeline but represent an ordered series of events and outcomes chronologically [86]. Another new concept is representing and visualizing patient clinical history as a visual map [87] to enhance navigation and analysis. In this technique clinical semantic groups are visualized as a map [87] to organize and visualize personal history. It transforms and organizes clinical text documents into semantic groups to provide healthcare providers with a single view of a patient medical history. Visualizing semantic units of clinical texts is a nascent approach to visualizing clinical narrative texts. Other techniques include word cloud [88] which was has been used to visualize concepts from history of present illness notes in [89]. Another technique is use of tag clouds [90] where words are seen by their size, depending on their frequency. Both word cloud and Tag cloud are used to provide visual representation of text content by displaying words considered important in a document. They are mainly applied in textual visualizations.

3.2. Information retrieval

There has lately been increased interest in using text segments in information retrieval rather the whole document. In such cases, information retrieval needs to match relevant texts with a given query. Most research works have been dealing with the problem of matching the query content with the whole document. However, there are some attempts that are focusing on how to partition a document into relevant segments of a document from which users can issue queries i.e. providing the user with the relevant facets of information that are relevant to his or her queries. This is particularly useful when documents are long, and some segments are relevant to user needs. There are works that have adopted this approach such as [91, 92] and many other works. In the clinical domain, physicians' chart notes are divided into sections that identify different information facets that make it easier to retrieve information [93]. This is achieved using clinical documentation formats such as SOAP (Subjective, Objective, Assessment and Plan) where each section is indicated by a section header that corresponds to one of the four SOAP data elements. Retrieving information in these sections allows one to create searches that are specific to a particular section rather than the entire document. Many studies such as in [94-96] have looked into the problem of segmenting clinical texts; however none have looked into whether it improves information retrieval performance.
The benefit of segmentation is that it organizes clinical texts so that information can be found quickly. Making the most clinically relevant data in the medical record easier to find and more readily available is critical.

4. Motivation

Clinicians now have easier access to information thanks to the growing use of electronic medical records [97]. A solution that can help physicians organize and manage patient data in a way that makes it easier for them to use the information available and hence improves efficiency is needed. In this paper, we look at how to visualize a patient's medical history that is provided in unstructured text so that physicians may get a quick summary. Data visualization has become more useful for reviewing and exploring vast volumes of healthcare data. As a result, in recent decades, the number of data visualization tools has grown.

5. Problem description

Large volumes of clinical data are available to users in electronic health records in the form of unstructured narrative texts such as clinical notes, discharge summaries, etc. One of the common routines in medical practice is looking for information in clinical documentation [98]. This is a difficult task since most clinical information is in unstructured narrative text documents [98-105].

It is convenient for doctors to document clinical encounters in narrative text, as it provides complete descriptions that is not possible to obtain using structured form [106] thus resulting in clinical text documents that needs to be read while looking for information [107]. However, it is widely acknowledged to be a laborious task to look for information in clinical text documents [98], 108]. Reading and going through numerous clinical documents in its entirety considering the time constraints doctors face during clinical encounters [109] is a challenge. One solution to this problem is to provide selective reading of pieces of texts rather than reading the entire text document. It is more convenient for users to look for particular information by browsing through categories rather than searching the whole information space.

The need of automatic methods for extracting relevant clinical information from large clinical text documents requires a method for organizing information and presenting it visually. For example, during clinical encounters with patients, the clinical documentation of previous encounters is very important information for decision making. Reading the entire patient clinical history and picking importance information may take a lot of time. There is need of taking pieces of texts and classifying them into important information classes and displaying them in respective groups.

6. Proposed technique

Clinical charts document a patient clinical history with different types of information. Headings and sub-headings are occasionally used in clinical documentation to indicate the organization of clinical documents. However, many clinical texts are long with very little structural demarcation, and in such a case, modeling into multiple facets can be useful. In this paper we considered the problem of subdividing narrative text documents into semantically coherent units that represent subtopics. In this case, the natural solution is to organize information into groups based on common themes and give these groups meaningful names. To achieve this, there is need to first label strings of texts (sentences or phrases) to enable us categorize information by means of labels. The SOAP documentation section names are used as labels which serve as a basis for recognizing important information facets. Subtopic structure is sometimes marked in technical texts by headings and sub-headings. o smaller semantically coherent chunks

*6.1. **Overview of our approach***

Because physicians frequently review patients' clinical documentation that they have made in the past, the goal of this study was to propose a novel technique for semantic modelling of clinical texts to support physicians in finding information in electronic clinical texts as well as the accuracy of the retrieved information. In this section we describe the propose technique in detail. Our objective is to address the problem of visually organizing clinical text documents to help physicians review clinical text documents by modelling semantic classes of a patient medical history. In particular, we would like to provide a means which retrieves and visualizes different facets of information in a long narrative text document. We propose text classification as a pre-cursor to creating a visual cluster map that organizes a document in terms of basic facets of clinical information, each of which is called a cluster. The cluster map is organized as a 4-dimensional semantic space. In addition to the idea of clusters, a cluster map needs to be organized such that, the relationships between clusters are shown.

6.2. *Design Requirements*

Based on interviews and workshops with doctors, as well as a literature research, several tasks and design needs were determined. We were only interested on how doctors review information in electronic health records. As clinical decisions are often based on patient's medical history, the relevant data elements we are interested to model are elements that describe patients' clinical events that occur during clinical encounters, and clinical documentation format that is used to document these events. Thus, we considered patient SOAP clinical notes which consist of four main types of descriptions: (1) Subjective, (2) Objective, (3) Assessment and (3) Plan. We need to classify these data elements in a given clinical text document and map similarly classified texts to corresponding semantic groups which can then be used to visualized and display using a cluster map.

These requirements are summarized as follows:

i. R1: Facilitate review of clinical text documents and make it easier for physicians to browse various types of information.
ii. R2: Visually present SOAP clinical notes sections in a cluster map facilitating selective access of information.
iii. R3: Visually distinguish different semantic groups of information using different colours.
iv. R4: Grouping clinical texts with respect to SOAP documentation format.
v. R5: Showing relationships between different clusters of information.

The cluster map graphically presents document classes with the relationship be-tween these classes.

6.3. *SOAP Documentation Format*

As mentioned in the previous section, SOAP documentation format is made up of four sections; Subjective, Objective, Assessment and Plan. Subjective part of SOAP is usually the background information of the patient which is required for understanding the current state of patient. Objective is measurable and quantifiable information which can be analyzed. Assessment is defined as the diagnosis based on the differential number of diagnosis. Plan is defined as the actions that need to be taken including any follow up checkup and treatment actions. We obtained the dataset for this work from the mtsamples.com, a website which provides a large collection of transcribed medical reports. The table shows the elements and descriptions of SOAP clinical notes.

SOAP Sections	Description
Subjective	Background information that is relevant for knowing the current state of the patient.It may include ; Family history, Daily habits, Current medications, allergies, series of events that happens in between
Objective	Quantifiable or measurable data obtained from past records and examinations, screening, tests

Assessment	Possible diagnosis provided by the practitioners or the staff treating the patient
Plan	Treatment strategies, actions to be taken, follow up plans

Table 1.SOAP documentation format.

7. Method

7.1. Sampling strategy and selection of participants

For the model evaluation, we used a purposive sampling technique to recruit research participants. The participants were approached in person and asked whether they were interested in participating. Participation was entirely optional. During the process, the following criteria for inclusion and exclusion were developed:

1. Professional doctors who are actively utilizing any type of electronic health system and capturing patient health data using any type of EHR were required to participate in the evaluation.
2. Participants who did not match the aforementioned inclusion criteria were not allowed to participate in the study.

As a result, individuals were sought to complete a questionnaire to evaluate the prototype. For the assessment of clinical charts, a select group of twelve doctors was chosen. Patients' charts with complicated illnesses and various comorbidities were chosen for this investigation.

7.1. Dataset

The clinical charts used in this paper were obtained from mtsamples.com which gives access to a large collection of transcribed medical reports. This dataset comprises 5,000 sample medical transcription reports. It is a useful dataset that has been used in many medical NLP research works.

We obtained SOAP clinical notes which contained a set of observations organized into four SOAP format sections. The SOAP description of these sections is as follows;

1) Subjective – Description of information such as symptoms, behaviors, and past medical information.
2) Objective - Description of the doctor's observations from physical examinations and previously ordered tests.
3) Assessment - Description of the potential problem(s) and related synthesis of the information from Subjective and Objective sections.
4) Plan- Description of how the problem will be addressed or description of further investigation.

All these sections are relevant to physicians and therefore we considered modeling all these information in each of the section. Although these parts can be further divided into subsections, we will just look at the four main aspects and ignore the sub-sections.

7.2 Design process

Clinical notes provide useful information that aids in the development of a more thorough understanding of a patient. Our goal is to figure out how to model the information in clinical notes that is frequently seen in a clinical report. We used an iterative design approach to design our prototype, which included cycles of defining the context and needs, brainstorming ideas, building a prototype, and testing it with users. The prototype application was developed in cooperation with medical practioners. There were initial meetings aimed at obtaining a list of needs for the prototype, as well as follow-up sessions targeted at gathering input, which might include new prospective features or a shift in approach in previously developed functionalities.

Our dataset contains a description of patients' clinical histories that must be segmented into predefined facets of information. In general, the proposed method entails determining pieces of text that describe a similar information facet and organizing them into clusters from which users can look for information in particular clusters.

The system was designed with two main components, the classification component, which is responsible with classifying sentences to various classes, and the visualization component, which provides the user with information in a visual map.

7.3. *Text classification*

Usually, humans organize information into groups or categories. Artificial Intelligent follows the same principles using two broad types of algorithms; clustering and classification. In this paper we adopted classification algorithm to group clinical texts into various semantic groups inherent in clinical documentation. We relied on a priori reference SOAP documentation structure that divide the space of all possible data points into a set of classes(Subjective, Objective, Assessment and Plan). In this section, our objective is to be able to categorize clinical text sentences into one of these classes. This is a multi-class classification problem.

In this section we designed a classifier to classify sentences in a given clinical document. In this paper, we used clinical notes in SOAP documentation format. Based on input from doctors, we defined four semantic classes of information in a SOAP clinical document which is usually information of interest to practitioners: Subjective, Objective, Assessment and Plan. Each sentence in the corpus must be classified as belonging to one of these four categories. In the task, given a sentence narrative, the model attempts to predict which class the sentence is about.

Clinical sentences in SOAP document are classified using a variant of Recurrent Neural Network known as Long Short Term Memory network (LSTM). In SOAP note, each clinical sentence belongs to a certain semantic class depending upon its meaning and corresponds to a section in a SOAP documentation format. The summary of our steps are listed below;

1. Tokenization - A collection of patient clinical text documents $\mathrm{D} = \{\mathrm{d}_1, \mathrm{d}_2, \dots \dots \mathrm{d}_\mathrm{n}\}$ is split into a set of sentences $\mathrm{S} = \{\mathrm{s}_1, \mathrm{s}_2, \dots \dots \mathrm{s}_\mathrm{n}\}$. Our objective is to classify these sentences into a predefined set of classes.
2. Feature generation- After tokenization, a feature vector for our deep learning classifiers is required. We used word embedding to generate the required feature vectors for each sentence. Word embedding results as input features
3. Input layer - These feature vectors are then used as input into the embedding layer of neural network i.e. word embedding results are used as input features.
4. Embedding layer output - The output generated from the embedding layers is fed into the next fully connected layer (dense layer) of neural network.
5. Output layer - A relevant class label (Subjective, Objective, Assessment, and Plan) is assigned to each sentence at the output layer.

The dataset which obtained from the above mentioned site was used for the classifier. However, since we adopted supervised learning which requires labeled data, sentences from clinical reports in the dataset were manually chosen randomly and classified into four classes. The model was trained using the training dataset which had been labeled with the help of medical professionals. The dataset was split into two; 80% for training, and 10% for testing. Using the trained Neural Network, the sentences are classified into the four classes (Soap, subjective, Assessment and Plan) that were found to be relevant and useful clinical information in a clinical chart.

7.4. *Cluster map generation*

Our objective is to generate clusters of information which contains similar sentences according to classification results. Therefore, a cluster should have sentences that have been correctly classified to the same class. The classified sentences are grouped according to their label and visualized in a map layout to depict the semantic classes of information. After classifying sentences with appropriate labels, we now have a bunch of sentences. The existence of some sentences with similar class label leads to the need of placing them into a specific group. Sentences that are in the same group discuss similar information while sentences in different groups discuss dissimilar information.

Every single sentence has a label (class), which tells the type of group it belongs to. A group in this case is a container (cluster) for a given number of sentences. It has a type. The cluster map is then used to display classed sentences, with each cluster consisting of a collection of sentences that have been labeled with the same class so that their relevance can be immediately recognized. Figure 5 shows an example of a cluster map derived from clinical notes, with four clusters representing distinct semantic classes of information.

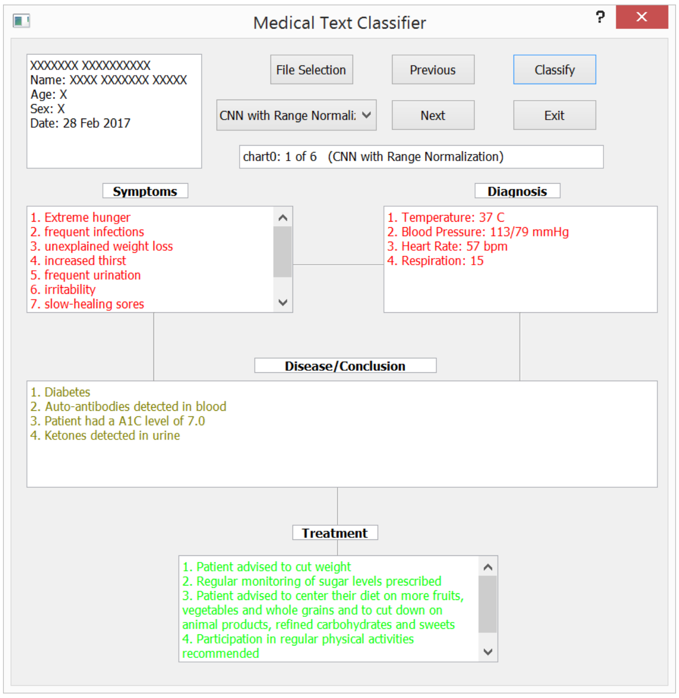

Figure 1. Sample Cluster map.

8. Evaluation

The prototype has been demonstrated to be effective in producing information groups that closely matches human generated subtopic from text documents. Tasks, such as information retrieval, should benefit from such a model. To validate the model for practical use, there is need to evaluate it to ascertain if it addresses the needs of physicians. A user study was conducted in order to assess the usability of the proposed prototype. Evaluators were exposed to the prototype and the system usability (SUS) questionnaire was administered to assess its usability. The evaluation process was conducted the objective of determining the usability using the System Usability Scale (SUS). Twelve physicians were recruited to evaluate the perceived usability of the proposed system.

To evaluate the usability of the prototype, the System Usability Scale (SUS) [110], [111] was adopted. It is made up of ten questions that are evaluated on a 5-point scale of level of agreement. To evaluate the prototype's usability, we conducted a user study with 12 physicians who used the prototype to review medical transcription reports. Participants were asked to score the level of agreement with 10 questions using a five-point Likert scale: Strongly agree (5), Agree (4), Neutral (3), Disagree (2), Strongly Disagree (1). The SUS-score for individual questions is obtained by subtracting 1 from odd questions (Response – 1) and subtracting 5 from even questions (5 – Response). The final score is obtained by summing the questions SUS scores and then multiplying the resulting sum by 2.5 to obtain the overall SUS score [110]. This score usually ranges between 0 and 100 and a higher score indicates good usability. The final SUS score gives an overall usability measurement, according to ISO 9241–11, which is made up of three characteristics; effectiveness, efficiency and satisfaction [111]. As a rule of thumb, a score above 70 has good usability; a lower score means poor usability and the system needs more improvement. It is a reliable, low-cost scale used for evaluating systems usability [112].

	QUESTION	Strongly disagree 1	2	3	4	Strongly agree 5
1	I think that I would like to use system frequently	☐	☐	☐	☐	☐
2	I found the visualization artefact unnecessarily complex	☐	☐	☐	☐	☐
3	I thought the visualization artefact was easy to use	☐	☐	☐	☐	☐
4	I think that I would need the support of a technical person to be able to use this system	☐	☐	☐	☐	☐
5	I found the various functions in this system were well integrated.	☐	☐	☐	☐	☐
6	I thought there was too much inconsistency in this system.	☐	☐	☐	☐	☐
7	I would imagine that most people would learn to use this system very quickly	☐	☐	☐	☐	☐
8	I found the system very cumbersome to use	☐	☐	☐	☐	☐
9	I felt very confident using the system.	☐	☐	☐	☐	☐
10	I needed to learn a lot of things before I could get going with this system	☐	☐	☐	☐	☐

Table 2. SUS questionnaire.

9. Results

The SUS Score was calculated using the results of the questionnaire, which were obtained from 12 respondents, and the results of the questionnaire were calculated using the above formula. Table 1.2 displays the results of the SUS score assessment.

PARTICIPANT	QUESTIONS										
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS SCORE
1	5	1	5	1	4	1	4	1	4	2	90.0
2	3	2	4	2	5	2	5	2	4	3	75.0
3	5	2	3	2	4	2	4	1	5	1	82.5
4	4	1	4	2	4	3	3	1	4	2	75.0
5	5	2	4	1	3	1	4	2	4	1	82.5
6	2	2	3	3	2	2	3	3	3	2	52.5
7	4	2	2	3	4	2	4	2	2	1	65.0
8	4	2	3	2	5	1	2	2	3	3	67.5
9	5	2	4	1	4	2	4	1	4	1	85.0
10	4	1	2	2	4	3	3	2	3	2	65.0
11	4	2	4	2	4	2	2	2	4	2	70.0
12	3	2	3	1	3	1	4	1	4	1	77.5
										AVERAGE	73.96

Table 3. SUS results.

The final SUS-score is 73.96, which indicates good usability, based on the above results. According to the SUS score scale [113], a SUS value of 73.96 is regarded as a good usability rating.

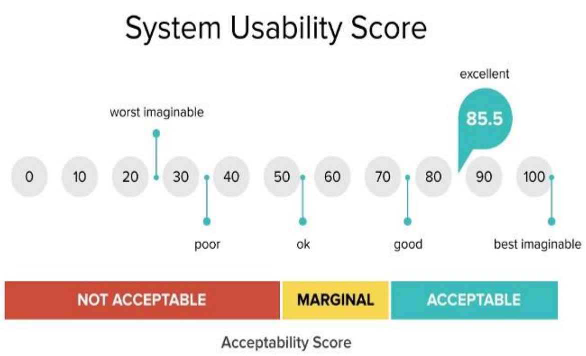

Figure 2. SUS score scale (Adapted from 10up.com).

10. Discussion

The improved availability and accessibility of patient care data, has created new possibilities for data use and reuse with the potential to improving the quality, safety, and efficiency of clinical work [114]. In this paper we have presented how clinical narrative texts can be classified and presented using a visual cluster map. Such an approach can help users sift through large quantities of text documents. In our approach we were primarily concerned with the general characterization of a clinical text document into semantic clusters of information, enabling the user to rapidly focus on a subset of potentially needed information. The user will only be left with the effort of reading a particular cluster of information rather than reading the whole document. This can be useful for lengthy documents. This was informed by the fact that users usually judge the importance of a piece of text by simply looking at its title and then deciding whether to read or not. By categorizing texts into semantic clusters and assigning descriptive semantic labels will allow users to instantly view the content of text and decide on which information cluster to focus on. This can be useful in tasks such as chart biopsy which involves getting a general overview of the patient by selectively examining parts of a patient's health record with the objective of getting specific data about a particular patient or appraising oneself with a patient and the care that a patient has received [114]. Additionally, this can contribute to decreasing cognitive load on physician part, reducing time required to complete tasks, thus giving more time for face-to-face interaction with patients. This is in line with findings from earlier studies such as in [115] which have shown that usability enhancements within EHRs can reduce cognitive load. With the prevalence of electronic health records in contemporary medical practice, large quantities of data are generated which requires physicians to review them. When completing clinical tasks, electronic health records have been shown to increase physician cognitive workload [116], which, according to cognitive load theory, can lead to cognitive overload [117]. By organizing and visually presenting information users can be help to synthesize data into meaningful information.

11. Conclusion

We created a model to assist doctors in effectively gleaning insights and identifying vital information from clinical text documents. Clinical documentation of patient encounters is important for providing patient care. The increasing use of electronic health records is impeded because of several reasons. One of the reasons is the continued use of unstructured narrative text which is inherent in clinical documentation. In this paper we illustrated how clinical notes can be modeled into easily accessible facets of information, without the need of changing the format of narrative texts. The structuring and visualization principles behind SOAP medical record structures are presented. A clinical document is classified into four SOAP elements: Subjective, Objective, Assessment and Plan. A map is presented in which each of these elements can be viewed. In this paper we reviewed various approaches to visualizing clinical text documents. The review of existing concepts of health data visualization may provide a useful theoretical framework for future research on how clinical data visualization can best be used to support medical practice. Text classification and visualization were explored as ways of helping physicians review clinical texts. Text visualization has a lot of potential for extracting relevant information from narrative clinical notes, which can help doctors make better decisions. This will go a long way toward harnessing data from electronic health records to improve treatment while also assisting physicians in doing so.

Funding: This research received no external funding

Acknowledgments: The authors wish to thank the reviewers for their helpful comments. Their suggestion greatly improved the paper. We also thank the participants for providing insight on many clinical tasks and their input in designing the prototype. Cohen deserves our thanks for proofreading the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Maheshwari, R.; Moudgil, K.; Parekh, H.; Sawant, R. A Machine Learning Based Medical Data Analytics and Visualization Research Platform. *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 2018, pp. 1-5, doi: 10.1109/ICCTCT.2018.8550953.
2. López-Martínez, F.; Núñez-Valdez, E.R.; García-Díaz, V.; Bursac, Z. A Case Study for a Big Data and Machine Learning Platform to Improve Medical Decision Support in Population Health Management. Algorithms 2020, 13, 102. https://doi.org/10.3390/a13040102
3. Mustafa, A.; Rahimi Azghadi, M. Automated Machine Learning for Healthcare and Clinical Notes Analysis. Computers 2021, 10, 24. https://doi.org/10.3390/computers10020024
4. Muralidhar, E. S.; Gowtham, T. S.; Jain, A.; Padmaveni, K. Development of Health Monitoring Application using Machine Learning on Android Platform. 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 1076-1085, doi: 10.1109/ICCES48766.2020.9137969.
5. Chowdhury, A.; Rosenthal, J.; Waring, J.; Umeton, R. Applying Self-Supervised Learning to Medicine: Review of the State of the Art and Medical Implementations. Informatics 2021, 8, 59. https://doi.org/10.3390/informatics8030059
6. Singh, R.; Sharma, N. Machine Learning based Medical Information Analysis, Estimations and Approximations over Present Health Research Domain. 2020 International Conference on Computational Performance Evaluation (ComPE), 2020, pp. 704-708, doi: 10.1109/ComPE49325.2020.9200045.
7. Massaro, A.; Maritati, V.; Savino, N.; Galiano, A. Neural Networks for Automated Smart Health Platforms oriented on Heart Predictive Diagnostic Big Data Systems. 2018 AEIT International Annual Conference, 2018, pp. 1-5, doi: 10.23919/AEIT.2018.8577362.
8. Massaro, A.; Ricci, G.; Selicato, S.; Raminelli, S.; Galiano, A. Decisional Support System with Artificial Intelligence oriented on Health Prediction using a Wearable Device and Big Data. 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, 2020, pp. 718-723, doi: 10.1109/MetroInd4.0IoT48571.2020.9138258.
9. Massaro, A.; Galiano, A.; Scarafile, D.; Vacca, A.; Frassanito, A.; Melaccio, A.; Solimando, A.; Ria, R.; Calamita, G.; Bonomo, M.; Vacca, F. Telemedicine DSS-AI Multi Level Platform for Monoclonal Gammopathy Assistance. *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2020, pp. 1-5, doi: 10.1109/MeMeA49120.2020.9137224.
10. Tange, H.; Nagykaldi, Z.; De Maeseneer, J. Towards an Overarching Model for Electronic Medical-Record Systems, Including Problem-Oriented, Goal-Oriented, and Other Approaches. European Journal of General Practice 2017, 23, 257–260.
11. Rostamzadeh, N.; Abdullah, S.S.; Sedig, K. Visual Analytics for Electronic Health Records: A Review. In Proceedings of the Informatics; Multidisciplinary Digital Publishing Institute, 2021; Vol. 8, p. 12.
12. Wanderer, J.P.; Nelson, S.E.; Ehrenfeld, J.M.; Monahan, S.; Park, S. Clinical Data Visualization: The Current State and Future Needs. Journal of medical systems 2016, 40, 1–9.
13. Kuhn, T.; Basch, P.; Barr, M.; Yackel, T.; Physicians*, M.I.C. of the A.C. of Clinical Documentation in the 21st Century: Executive Summary of a Policy Position Paper from the American College of Physicians. Annals of internal medicine 2015, 162, 301–303.
14. Grasso, C.T.; Joshi, A.; Siegel, E. Visualization of Pain Severity Events in Clinical Records Using Semantic Structures. In Proceedings of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC); IEEE, 2016; pp. 321–324.
15. Spasic, I.; Nenadic, G.; others Clinical Text Data in Machine Learning: Systematic Review. JMIR medical informatics 2020, 8, e17984.
    https://doi.org/10.2196/17984
16. Apostolova, E.; Channin, D.S.; Demner-Fushman, D.; Furst, J.; Lytinen, S.; Raicu, D. Automatic Segmentation of Clinical Texts. In Proceedings of the 2009 annual international conference of the IEEE engineering in medicine and biology society; IEEE, 2009; pp. 5905–5908.
17. Baron, R.J. Doctors' Stories: The Narrative Structure of Medical Knowledge. Princeton University Press. Literature and Medicine 1992, 11, 321–324.
18. Johnson, S.B.; Bakken, S.; Dine, D.; Hyun, S.; Mendonça, E.; Morrison, F.; Bright, T.; Van Vleck, T.; Wrenn, J.; Stetson, P. An Electronic Health Record Based on Structured Narrative. Journal of the American Medical Informatics Association 2008, 15, 54–64.
19. Osheroff, J.A.; Teich, J.M.; Middleton, B.; Steen, E.B.; Wright, A.; Detmer, D.E. A Roadmap for National Action on Clinical Decision Support. Journal of the American medical informatics association 2007, 14, 141–145.
20. Rind, A.; Wang, T.D.; Aigner, W.; Miksch, S.; Wongsuphasawat, K.; Plaisant, C.; Shneiderman, B. Interactive Information Visualization to Explore and Query Electronic Health Records. Foundations and Trends in Human-Computer Interaction 2013, 5, 207–298.
21. Lesselroth, B.J.; Pieczkiewicz, D.S. Data Visualization Strategies for the Electronic Health Record; Nova Science Publishers, Inc., 2011;
22. de Oliveira, J.M.; da Costa, C.A.; Antunes, R.S. Data Structuring of Electronic Health Records: A Systematic Review. Health and Technology 2021, 11, 1219–1235.
23. Venkataraman, G.R.; Pineda, A.L.; Bear Don't Walk IV, O.J.; Zehnder, A.M.; Ayyar, S.; Page, R.L.; Bustamante, C.D.; Rivas, M.A. FasTag: Automatic Text Classification of Unstructured Medical Narratives. PLoS one 2020, 15, e0234647.
24. Lin, W.; Ji, D.; Lu, Y. Disorder Recognition in Clinical Texts Using Multi-Label Structured SVM. BMC bioinformatics 2017, 18, 1–11.
25. Pivovarov, R.; Elhadad, N. Automated Methods for the Summarization of Electronic Health Records. Journal of the American Medical Informatics Association 2015, 22, 938–947.
26. Caban, J.J.; Gotz, D. Visual Analytics in Healthcare–Opportunities and Research Challenges. Journal of the American Medical Informatics Association 2015, 22, 260–262.

27. Chen, I.Y.; Agrawal, M.; Horng, S.; Sontag, D. Robustly Extracting Medical Knowledge from EHRs: A Case Study of Learning a Health Knowledge Graph. In Proceedings of the PACIFIC SYMPOSIUM ON BIOCOMPUTING 2020; World Scientific, 2019; pp. 19–30.
28. Pomares-Quimbaya, A.; Kreuzthaler, M.; Schulz, S. Current Approaches to Identify Sections within Clinical Narratives from Electronic Health Records: A Systematic Review. BMC medical research methodology 2019, 19, 1–20.
29. Safran, C.; Bloomrosen, M.; Hammond, W.E.; Labkoff, S.; Markel-Fox, S.; Tang, P.C.; Detmer, D.E. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association 2007, 14, 1–9. https://doi.org/10.1197/jamia.M2273
30. Liu, S.; Wang, Y.; Wen, A.; Wang, L.; Hong, N.; Shen, F.; Bedrick, S.; Hersh, W.; Liu, H.; others Implementation of a Cohort Retrieval System for Clinical Data Repositories Using the Observational Medical Outcomes Partnership Common Data Model: Proof-of-Concept System Validation. JMIR medical informatics 2020, 8, e17376.
31. Murdoch, T.B.; Detsky, A.S. The Inevitable Application of Big Data to Health Care. Jama 2013, 309, 1351–1352.
32. Wanderer, J.P.; Nelson, S.E.; Ehrenfeld, J.M.; Monahan, S.; Park, S. Clinical Data Visualization: The Current State and Future Needs. Journal of medical systems 2016, 40, 1–9.
33. Jarabek, B., Mink, P., Winden, T., Bork, L., Elison, J.T., Finley, G., Giaquinto, R., Hultman, G.M., Lindemann, E.A., McEwan, R., Rogers, J., Sarda, G., & Sun, D. Discovery and Visualization of New Information from Clinical Reports in the EHR - Final Report.2019
34. Lauster, C.D.; Srivastava, S.B. Fundamental Skills for Patient Care in Pharmacy Practice; Jones & Bartlett Publishers, 2013;
35. Silow-Carroll, S.; Edwards, J.N.; Rodin, D. Using Electronic Health Records to Improve Quality and Efficiency: The Experiences of Leading Hospitals. Issue Brief (Commonw Fund) 2012, 17, 40.
36. Semanik, M.G.; Kleinschmidt, P.C.; Wright, A.; Willett, D.L.; Dean, S.M.; Saleh, S.N.; Co, Z.; Sampene, E.; Buchanan, J.R. Impact of a Problem-Oriented View on Clinical Data Retrieval. Journal of the American Medical Informatics Association 2021, 28, 899–906.
https://doi.org/10.1093/jamia/ocaa332
37. Altuncu, M.T.; Mayer, E.; Yaliraki, S.N.; Barahona, M. From Free Text to Clusters of Content in Health Records: An Unsupervised Graph Partitioning Approach. Applied network science 2019, 4, 1–23.
38. Li, Y.; Lipsky Gorman, S.; Elhadad, N. Section Classification in Clinical Notes Using Supervised Hidden Markov Model. In Proceedings of the Proceedings of the 1st ACM International Health Informatics Symposium; 2010; pp. 744–750.
39. Pollack, A.H.; Pratt, W. Association of Health Record Visualizations with Physicians' Cognitive Load When Prioritizing Hospitalized Patients. JAMA network open 2020, 3, e1919301–e1919301.
40. Rostamzadeh, N.; Abdullah, S.S.; Sedig, K. Data-Driven Activities Involving Electronic Health Records: An Activity and Task Analysis Framework for Interactive Visualization Tools. Multimodal Technologies and Interaction 2020, 4, 7.
41. Sultanum, N.; Brudno, M.; Wigdor, D.; Chevalier, F. More Text Please! Understanding and Supporting the Use of Visualization for Clinical Text Overview. In Proceedings of the Proceedings of the 2018 CHI conference on human factors in computing systems; 2018; pp. 1–13.
42. Shah, A.D.; Martinez, C.; Hemingway, H. The Freetext Matching Algorithm: A Computer Program to Extract Diagnoses and Causes of Death from Unstructured Text in Electronic Health Records. BMC medical informatics and decision making 2012, 12, 1–13.
43. Deng, Y.; Denecke, K. Visualizing Unstructured Patient Data for Assessing Diagnostic and Therapeutic History. In e-Health–For Continuity of Care; IOS press, 2014; pp. 1158–1162.
44. Pollack, A.H., ; Pratt, W. Association of Health Record Visualizations With Physicians' Cognitive Load When Prioritizing Hospitalized Patients. JAMA network open 2020, 3, e1919301–e1919301.
45. Hillestad, R.; Bigelow, J.; Bower, A.; Girosi, F.; Meili, R.; Scoville, R.; Taylor, R. Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, and Costs. Health affairs 2005, 24, 1103–1117.
46. Nguyen, L.; Bellucci, E.; Nguyen, L.T. Electronic Health Records Implementation: An Evaluation of Information System Impact and Contingency Factors. International journal of medical informatics 2014, 83, 779–796.
47. Pai, M.M.; Ganiga, R.; Pai, R.M.; Sinha, R.K. Standard Electronic Health Record (EHR) Framework for Indian Healthcare System. Health Services and Outcomes Research Methodology 2021, 21, 339–362.
48. Kubben, P.; Dumontier, M.; Dekker, A. Fundamentals of Clinical Data Science. 2019. https://doi.org/10.1007/978-3-319-99713-1_1
49. Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting Diagnoses and Investigation Results from Unstructured Text in Electronic Health Records by Semi-Supervised Machine Learning. 2012. PLoS ONE 7(1): e30412. https://doi.org/10.1371/journal.pone.0030412
50. Liang, J.J.; Tsou, C.-H.; Poddar, A. A Novel System for Extractive Clinical Note Summarization Using EHR Data. Proceedings of the 2nd Clinical Natural Language Processing Workshop 2019.
51. Casey, J.A.; Schwartz, B.S.; Stewart, W.F.; Adler, N.E. Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. Annual review of public health 2016, 37, 61–81.
52. Jensen, P.B.; Jensen, L.J.; Brunak, S. Mining Electronic Health Records: Towards Better Research Applications and Clinical Care. Nature Reviews Genetics 2012, 13, 395–405.
53. Altuncu, M.T.; Mayer, E.; Yaliraki, S.N.; Barahona, M. From Free Text to Clusters of Content in Health Records: An Unsupervised Graph Partitioning Approach. Applied Network Science 2019, 4.
54. Ning X.; Fan Z.; Burgun E.; Ren Z, Schleyer T. Improving Information Retrieval from Electronic Health Records Using Dynamic and Multi-Collaborative Filtering. 2019 IEEE International Conference on Healthcare Informatics (ICHI) 2019, 1–3.. https://doi.org/10.1371/journal.pone.0255467
55. Kosara, R.; Miksch, S. Visualization Methods for Data Analysis and Planning in Medical Applications. International journal of medical informatics 2002, 68 1-3, 141–153.
56. Wanderer, J.P.; Nelson, S.E.; Ehrenfeld, J.M.; Monahan, S.; Park, S. Clinical Data Visualization: The Current State and Future Needs. Journal of Medical Systems 2016, 40, 1–9.

57. Kimia, A.A.; Savova, G.K.; Landschaft, A.; Harper, M.B. An Introduction to Natural Language Processing: How You Can Get More From Those Electronic Notes You Are Generating. Pediatric emergency care 2015, 31 7, 536–541.
58. Chou, S.; Chang, W.; Cheng, C.-Y.; Jehng, J.-C.J.; Chang, C. An Information Retrieval System for Medical Records & Documents. 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2008, 1474–1477.
59. Colorafi, K.; Moua, L.; Shaw, M.R.; Ricker, D.; Postma, J.M. Assessing the Value of the Meaningful Use Clinical Summary for Patients and Families with Pediatric Asthma. Journal of Asthma 2018, 55, 1068–1076.
60. Liang, J.J.; Tsou, C.-H.; Poddar, A. A Novel System for Extractive Clinical Note Summarization Using EHR Data. Proceedings of the 2nd Clinical Natural Language Processing Workshop 2019.
61. Rind, A.; Aigner, W.; Miksch, S.; Wiltner, S.; Pohl, M.; Turic, T.; Drexler, F. Visual Exploration of Time-Oriented Patient Data for Chronic Diseases: Design Study and Evaluation. In Proceedings of the USAB; 2011.
62. Blei, D.M.; Ng, A.; Jordan, M.I. Latent Dirichlet Allocation. J. Mach. Learn. Res. 2003, 3, 993–1022.
63. Deerwester, S.C.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R.A. Indexing by Latent Semantic Analysis. Journal of the Association for Information Science and Technology 1990, 41, 391–407.
64. Hofmann, T. Probabilistic Latent Semantic Analysis. In Proceedings of the UAI; 1999.
65. Duarte, D.; Puerari, I.; Bianco, G.D.; Lima, J.F. Exploratory Analysis of Electronic Health Records Using Topic Modeling. J. Inf. Data Manag. 2020, 11.
66. Chan, K.R.; Lou, X.; Karaletsos, T.; Crosbie, C.; Gardos, S.M.; Artz, D.; Rätsch, G. An Empirical Analysis of Topic Modeling for Mining Cancer Clinical Notes. 2013 IEEE 13th International Conference on Data Mining Workshops 2013, 56–63.
67. Arnold, C.W.; El-Saden, S.M.; Bui, A.A.T.; Taira, R.K. Clinical Case-Based Retrieval Using Latent Topic Analysis. AMIA ... Annual Symposium proceedings. AMIA Symposium 2010, 2010, 26–30.
68. Perotte, A.J.; Wood, F.D.; Elhadad, N.; Bartlett, N. Hierarchically Supervised Latent Dirichlet Allocation. In Proceedings of the NIPS; 2011.
69. Coroiu, A.M.; Calin, A.D.; Nutu, M. Topic Modeling in Medical Data Analysis. Case Study Based on Medical Records Analysis. 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM) 2019, 1–5.
70. Liu, X.; Alharbi, M.; Best, J.; Chen, J.; Diehl, A.; Firat, E.E.; Rees, D.; Wang, Q.; Laramee, R.S. Visualization Resources: A Starting Point. 2021 25th International Conference Information Visualisation (IV) 2021, 160–169.
71. Rind, A.; Wang, T.D.; Aigner, W.; Miksch, S.; Wongsuphasawat, K.; Plaisant, C.; Shneiderman, B. Interactive Information Visualization to Explore and Query Electronic Health Records. Found. Trends Hum. Comput. Interact. 2013, 5, 207–298.
72. Ledesma, A.; Bidargaddi, N.P.; Strobel, J.E.; Schrader, G.; Nieminen, H.; Korhonen, I.; Ermes, M. Health Timeline: An Insight-Based Study of a Timeline Visualization of Clinical Data. BMC Medical Informatics and Decision Making 2019, 19.
73. Grobelnik, M.; Mladenic, D. Visualization of News Articles. Informatica (Slovenia) 2004, 28, 375–380.
74. Imai, T.; Nakamura, K.; Ohmameuda, T. Visualization of Similar News Articles with Network Analysis and Text Mining. 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE) 2015, 151–152.
75. Doshi, K.; Gokhale, S.; Mamtora, H.; Bide, P.J. Analytics and Visualization of Trends in News Articles. 2019 International Conference on Advances in Computing, Communication and Control (ICAC3) 2019, 1–9.
76. Roque, F.S.; Slaughter, L.A.; Tkatsenko, A. A Comparison of Several Key Information Visualization Systems for Secondary Use of Electronic Health Record Content. In Proceedings of the Louhi@NAACL-HLT; 2010.
77. Powsner, S.M.; Tufte, E.R. Graphical Summary of Patient Status. The Lancet 1994, 344, 386–389.
78. Plaisant, C.; Milash, B.A.; Rose, A.; Widoff, S.; Shneiderman, B. LifeLines: Visualizing Personal Histories. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 1996.
79. Wang, T.D.; Plaisant, C.; Shneiderman, B.; Spring, N.; Roseman, D.; Marchand, G.; Mukherjee, V.; Smith, M.S. Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison. IEEE Transactions on Visualization and Computer Graphics 2009, 15.
80. Manning, J.D.; Marciano, B.E.; Cimino, J.J. Visualizing the Data – Using Lifelines2 to Gain Insights from Data Drawn from a Clinical Data Repository. AMIA Summits on Translational Science Proceedings 2013, 2013, 168–172.
81. Goren-Bar, D.; Shahar, Y.; Galperin-Aizenberg, M.; Boaz, D.; Tahan, G. KNAVE II: The Definition and Implementation of an Intelligent Tool for Visualization and Exploration of Time-Oriented Clinical Data. Proceedings of the working conference on Advanced visual interfaces 2004.
82. Hallett, C. Multi-Modal Presentation of Medical Histories. In Proceedings of the IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces; 2008
83. Miksch, S.; Kosara, R.; Shahar, Y.; Johnson, P.D. AsbruView: Visualization of Time-Oriented, Skeletal Plans. In Proceedings of the AIPS; 1998.
84. Wongsuphasawat, K.; Gómez, J.A.G.; Plaisant, C.; Wang, T.D.; Taieb-Maimon, M.; Shneiderman, B. LifeFlow: Visualizing an Overview of Event Sequences. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2011.
85. Monroe, M.; Meyer, T.E.; Plaisant, C.; Lan, R.; Wongsuphasawat, K.; Coster, T.S.; Gold, S.; Millstein, J.A.; Shneiderman, B. Visualizing Patterns of Drug Prescriptions with EventFlow: A Pilot Study of Asthma Medications in the Military Health System.; 2013.
86. Ledesma, A.; Bidargaddi, N.P.; Strobel, J.E.; Schrader, G.; Nieminen, H.; Korhonen, I.; Ermes, M. Health Timeline: An Insight-Based Study of a Timeline Visualization of Clinical Data. BMC Medical Informatics and Decision Making 2019, 19.
87. Kenei, J.K.; Opiyo, E.T.O.; Oboko, R.O. Visualizing Semantic Structure of a Clinical Text Document. European Journal of Electrical Engineering and Computer Science, 4(6). https://doi.org/10.24018/ejece.2020.4.6.256; 2020.
88. Sellars, B.B.; Sherrod, D.R.; Chappel-Aiken, L. Using Word Clouds to Analyze Qualitative Data in Clinical Settings. Nursing Management (Springhouse) 2018.

89. Rousseau, J.F.; Ip, I.K.; Raja, A.S.; Valtchinov, V.I.; Cochon, L.; Schuur, J.D.; Khorasani, R. Can Automated Retrieval of Data from Emergency Department Physician Notes Enhance the Imaging Order Entry Process? Applied Clinical Informatics 2019, 10, 189–198. https://doi.org/10.1055/s-0039-1679927
90. Foldes, D. Using Tag Clouds as a Tool for Patients' Medical History Visualization and Record Retrieval.; 2015.
91. Llopis, F.; Rodríguez, A.F.; González, J.L.V. Text Segmentation for Efficient Information Retrieval. In Proceedings of the CICLing; 2002.
92. Salton, G. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.; 1989.
93. Edinger, T.; Demner-Fushman, D.; Cohen, A.M.; Bedrick, S.; Hersh, W.R. Evaluation of Clinical Text Segmentation to Facilitate Cohort Retrieval. AMIA ... Annual Symposium proceedings. AMIA Symposium 2017, 2017, 660–669.
94. Ganesan, K.A.; Subotin, M. A General Supervised Approach to Segmentation of Clinical Texts. 2014 IEEE International Conference on Big Data (Big Data) 2014, 33–40.
95. Tepper, M.; Capurro, D.; Xia, F.; Vanderwende, L.; Yetisgen-Yildiz, M. Statistical Section Segmentation in Free-Text Clinical Records. In Proceedings of the LREC; 2012.
96. Mowery, D.L.; Wiebe, J.; Visweswaran, S.; Harkema, H.; Chapman, W.W. Building an Automated SOAP Classifier for Emergency Department Reports. Journal of biomedical informatics 2012, 45 1, 71–81.
97. Bashyam, V.; Hsu, W.; Watt, E.; Bui, A.A.T.; Kangarloo, H.; Taira, R.K. Problem-Centric Organization and Visualization of Patient Imaging and Clinical Data. Radiographics : a review publication of the Radiological Society of North America, Inc 2009, 29 2, 331–343.
98. Choi, Y.J.; Byun, J.; Berkovich, S.Y. Cross-Search Technique and Its Visualization of Peer-to-Peer Distributed Clinical Documents.; 2007.
99. Fan, Z.; Burgun, E.; Schleyer, T.K.; Ning, X. Improving Information Retrieval from Electronic Health Records Using Dynamic and Multi-Collaborative Filtering. 2019 IEEE International Conference on Healthcare Informatics (ICHI) 2019, 1–3.
100. Furlow, B. Information Overload and Unsustainable Workloads in the Era of Electronic Health Records. The Lancet. Respiratory medicine 2020, 8 3, 243–244.
101. Pollack, A.H.; Pratt, W. Association of Health Record Visualizations With Physicians' Cognitive Load When Prioritizing Hospitalized Patients. JAMA Network Open 2020, 3.
102. Wei, X.; Eickhoff, C. Embedding Electronic Health Records for Clinical Information Retrieval. ArXiv 2018, abs/1811.05402.
103. Sheikhalishahi, S.; Miotto, R.; Dudley, J.T.; Lavelli, A.; Rinaldi, F.; Osmani, V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. JMIR Medical Informatics 2019, 7.
104. Kuhn, L.; Eickhoff, C. Implicit Negative Feedback in Clinical Information Retrieval. ArXiv 2016, abs/1607.03296.
105. Sheikhalishahi, S.; Miotto, R.; Dudley, J.T.; Lavelli, A.; Rinaldi, F.; Osmani, V. Information Extraction from Clinical Notes: A Systematic Review for Chronic Diseases (Preprint).; 2018.
106. Tayefi, M.; Ngo, P.D.; Chomutare, T.; Dalianis, H.; Salvi, E.; Budrionis, A.; Godtliebsen, F. Challenges and Opportunities beyond Structured Data in Analysis of Electronic Health Records. Wiley Interdisciplinary Reviews: Computational Statistics 2021.
107. Ruan, W.; Appasani, N.; Kim, K.; Vincelli, J.; Kim, H.; Lee, W.-S. Pictorial Visualization of EMR Summary Interface and Medical Information Extraction of Clinical Notes. 2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) 2018, 1–6.
108. Holzinger, A.; Schantl, J.; Schroettner, M.; Seifert, C.; Verspoor, K.M. Biomedical Text Mining: State-of-the-Art, Open Problems and Future Challenges. In Proceedings of the Interactive Knowledge Discovery and Data Mining in Biomedical Informatics; 2014. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-662-43968-5_16
109. Sultanum, N.; Singh, D.; Brudno, M.; Chevalier, F. Doccurate: A Curation-Based Approach for Clinical Text Visualization. IEEE Transactions on Visualization and Computer Graphics 2019, 25, 142–151.
110. Peres, S.C.; Pham, T.; Phillips, R.G. Validation of the System Usability Scale (SUS). Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2013, 57, 192–196.
111. Brooke, J B. SUS — a retrospective. Journal of Usability Studies, Vol. 8, Issue 2, 2013 pp. 29–40.
112. Brooke, J.B. SUS: A "Quick and Dirty" Usability Scale.; 1996. In Jordan, P.W.; Thomas, B.; McClelland, I.; Weerdmeester, B.A. Usability Evaluation In Industry.; 1996.
113. Bangor, A.; Kortum, P.T.; Miller, J.T. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies archive 2009, 4, 114–123.
114. Hilligoss, B.; Zheng, K. Chart Biopsy: An Emerging Medical Practice Enabled by Electronic Health Records and Its Impacts on Emergency Department-Inpatient Admission Handoffs. Journal of the American Medical Informatics Association : JAMIA 2013, 20 2, 260–267.
115. Mazur, L.; Mosaly, P.; Moore, C.R.; Marks, L.B. Association of the Usability of Electronic Health Records With Cognitive Workload and Performance Levels Among Physicians. JAMA Network Open 2019, 2.
116. Pollack, A.H.; Pratt, W. Association of Health Record Visualizations With Physicians' Cognitive Load When Prioritizing Hospitalized Patients. JAMA Network Open 2020, 3.
117. Tawfik, A.A.; Kochendorfer, K.M.; Saparova, D.; Ghenaimi, S.A.; Moore, J.L. Using Semantic Search to Reduce Cognitive Load in an Electronic Health Record. 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services 2011, 181–184.