

Modeling and Visualization of Clinical Texts to Enhance Meaningful and User-Friendly Information Retrieval [†]

Jonah Kenei * and Elisha Opiyo

Department of Computing & Informatics, University of Nairobi, Nairobi P.O. Box 30197-00100, Kenya

* Correspondence: jonah.kenei@gmail.com; Tel.: +254-(0)7-2240-0542† Presented at the 2nd International Electronic Conference on Healthcare, 17 February–3 March 2022; Available online: <https://iech2022.sciforum.net/event/IECH2022>.

Abstract: Access to digital health data collections such as clinical notes, discharge summaries, or medical charts has increased in the last few years due to the increased use of electronic health records, which 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 require one to read through to get the required information. This is a challenging task due to lack of suitable techniques to quickly extract the needed information. Information processing tools in the clinical domain that provide support to users in seeking needed information are lacking. The use of data visualization 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 the semantics of various units of a clinical text document to aid doctors in reviewing electronic clinical notes. This is achieved by applying the 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.Modeling and Visualization of Clinical Texts to Enhance Meaningful and User-Friendly Information Retrieval. *Med. Sci. Forum* **2022**, *10*, 9. <https://doi.org/10.3390/IECH2022-12294>

Academic Editor: Tin-Chih Toly Chen

Published: 16 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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 healthcare data, such as sensor data from intensive care units (ICUs), data generated in telemedicine, and longitudinal data from electronic health records (EHRs) and other sources, are opening up new avenues for leveraging data-driven techniques such as machine learning (ML) [1–7] and artificial intelligence techniques on data retrieved from wearable health sensors [8] and data from telemedicine [9], as well as 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 the ease of access to digital clinical data in electronic health records have fueled research efforts aimed at helping physicians make use of the growing digital information. Currently, electronic health records are being used by numerous healthcare facilities, which not only provide huge amount of information available in electronic health records [11] but also presents challenges when using this information [12,13]. This is due to the large volume of clinical narrative texts which need to be read and understood in order to provide effective patient care. Physicians often rely on a patients health history, which makes it difficult to locate information when written

in narrative text [14]. 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 healthcare, generating patients clinical histories with detailed clinical information that supports clinical decision making [15]. Clinical narratives are commonly entered, captured, and stored in electronic health records in digital form [16], which 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], as well as for secondary purposes [29,30], such as clinical research [30]. When it comes to analyzing huge amounts of clinical texts, it becomes too challenging to do so 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 [12,31] makes it challenging for physicians to quickly locate needed information for patient care [32] that is critical to developing an appropriate assessment and plan for the individual patient [33].

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 [34]. However, with the increasing availability of clinical data, data retrieval becomes more difficult, leading to cognitive load and clinician burnout [35]. In most cases, it remains underutilized in clinical practice due to lack of suitable techniques to extract needed information in a timely manner [36]. Currently, there is ready availability of information in electronic health records in narrative text and there is a need for automated techniques to process such texts [37]. Recent research has shown that electronic health records (EHRs) that process, organize, and visualize clinically meaningful information significantly reduce physician cognitive workload [38].

Clinical records are used by doctors to make informed decisions at the point of care [39]. However, as the volume of clinical records along with time constraints inherent in healthcare setting increases, utilizing these records becomes challenging [40] and time-consuming [41]. 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 [42]. 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. Therefore, physicians, as well as researchers, have to spend more time analyzing patients health records [38]. To tackle this problem, data visualization techniques have been employed, to help physicians extract relevant valuable information and to reduce cognitive overload [38].

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 pro-

pose a technique to improve physicians ability to retrieve key concepts relevant to patient at hand.

2. Background

Electronic health records (EHRs) are becoming common in many healthcare establishments, replacing traditional paper records [43]. They are used to create patients health records during clinical encounters with patients in a healthcare facility [43]. The records contain patient demographics, progress notes, and medication history [35], offering important clinical information for care of patients and supporting other functions such as interoperability [44]. The information generated during clinical encounters is stored and maintained in electronic health records in order to take care of patients and follow-up [45]. Therefore, they are sources of important clinical information [46]; however, most of the documentation is in unstructured narrative text, which is time-consuming to review manually [47]. While there are structured patient data in electronic health records, important patient clinical information which describes patient care and management remains buried in narrative text, making it challenging and time-consuming for physicians to review during their usual medical practice [48]. Unstructured data refer to information that does not conform to a predefined model making it difficult to be processed using computer systems. In healthcare, this mostly represented by clinical narratives which constitute the bulk of clinical documentation. These unstructured clinical data are 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, this increases the workload of reviewing it during patients subsequent visits [48].

Substantial documentation in the form of clinical text is captured in electronic health records, often in a notes section [49]. During care episodes, doctors rely on available clinical documentation on which they base their decisions, in order to provide effective patient care. During a typical clinical encounter, physicians create and add to the patients medical records with a variety of clinical information; hence, large amounts of data are generated every time a patient visits the healthcare facility for diagnosis and treatment. Increasing the 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 patients medical history. Therefore, there is a challenge of how to unlock unstructured clinical data to improve patient care. There is a need, therefore, to aid them by providing systems which automate narrative text 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 a need for 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.

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 [50], and widespread problems due to lack of suitable techniques to extract needed clinical information [36] have been noted. In addition to a lack of appropriate techniques to support retrieval of needed clinical information, the problem of information overload [51] is contributing to the difficulty of using this information.

Patients clinical records are needed for a variety of reasons. Physicians usually examine a patients medical record in order to get information that will allow them to make in-

formed 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 [52]. The need for new computational techniques for the growing volume of clinical data in digital form [12] and mostly in unstructured narrative form [53] cannot be underestimated, e.g., the need for information retrieval from clinical notes in order to provide effective medical care [54] and generating clinical summaries from clinical texts [55] with important information relevant to a particular patient or accurate task-specific clinical summaries [49]. 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 need the attention of physicians. The use of visualization techniques has the potential to aid such tasks, thus improving health care [56]. On the basis of the above, we believe that there is a need for the best approach to ease 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, thus 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, thus supporting them in making clinical decisions. The tool is particularly designed to provide relevant information for physicians with respect to the patient at hand. As a starting point, we consider clinical notes written using the SOAP (subjective, objective, assessment, and plan) documentation format. However, a future goal is to explore how such a 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, the supervised machine learning technique was applied to clinical narrative datasets.

3. Related Works

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

3.1. Topic Modeling

Extensive study has been performed 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. It is 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 datasets is becoming increasingly popular. However, there has been little effort to adapt these models to clinical practice. In the literature, there are a number of topic models that are commonly used. These include LDA (Latent Dirichlet Allocation) [57], LSA (Latent Semantic Analysis) [58], and PLSA (Probabilistic Latent Semantic Analysis) [59]. For modeling clinical notes, the majority of previous methods used latent topic models for various applications. For example, the authors in [60] made use of topic modeling to explore electronic health records. There are several other research works reported in the literature that used topic models to solve the problem of finding themes in electronic health records. Examples include mining cancer clinical notes [61], comparing patients notes to the subjects discovered [62], grouping discharge summaries into hierarchical concepts [63], and identifying the most relevant subjects [64]. These publications, however, do not address the identification of the most common issues on which they focus.

3.2. Data Visualization Modeling

Data visualization is becoming increasingly important for analyzing large volumes of complex data [65]. There are many techniques that have been proposed for text visualization in the clinical domain [20,66], as well as in the general domain. This is largely driven by the need to improve the efficacy and utility of the collected data in electronic health records [20]. In clinical domain, the visualization strategies are meant to aid in understanding clinical data [66]. In the general domain, there has been extensive work in automated text visualization, such as visualization of news [67–69]. There are several research works that have been carried out and many authors have proposed a variety of visualization techniques [70]. One of the most prevalent techniques uses the original concept of TimeLine [71]. Examples falling under this category include Lifelines [72], Lifelines2 [73,74] KNAVE II [75], CLEF Visual Navigator [76], and AsbruView [77]. The focus of these techniques is to visualize clinical data as a function of time. Thus, we can refer to as time-based visualizations graphical representations of data collected over time. Other techniques include LifeFlow [78] and EventFlow [79]. Unlike the above techniques, these two techniques do not use a timeline but represent an ordered series of events and outcomes chronologically [71]. Another new concept is representing and visualizing patient clinical history as a visual map [80] to enhance navigation and analysis. In this technique, clinical semantic groups are visualized as a map [80] to organize and visualize personal history. This 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 [81], which was used to visualize concepts from history of present illness notes in [82]. Another technique is the use of tag clouds [83] where words are seen by their size, depending on their frequency. Both word clouds and tag clouds are used to provide visual representation of text content by displaying words considered important in a document. They are mainly applied in textual visualizations.

3.3. Information Retrieval

There has lately been increased interest in using text segments in information retrieval rather than the whole document. In such cases, information retrieval needs to match relevant texts with a given query. Most research works have dealt with the problem of matching the query content with the whole document. However, there are some attempts that focused 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 their queries. This is particularly useful when documents are long, and some segments are relevant to user needs. Many works that have adopted this approach such as [84,85]. In the clinical domain, physicians chart notes are divided into sections that identify different information facets that make it easier to retrieve information [86]. 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 [87–89] looked into the problem of segmenting clinical texts; however, none 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 [90]. 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, improve 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 and discharge summaries. One of the common routines in medical practice is looking for information in clinical documentation [91]. This is a difficult task since most clinical information is in unstructured narrative text documents [91–97].

It is convenient for doctors to document clinical encounters in narrative text, as it provides complete descriptions that are not possible to obtain using structured form [98], thus resulting in clinical text documents that need to be read while looking for information [99]. However, it is widely acknowledged to be a laborious task to look for information in clinical text documents [91,100]. Reading and going through numerous clinical documents in its entirety considering the time constraints doctors face during clinical encounters [101] 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 for automatic methods to extract 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 up important information may be time-consuming. There is need for taking pieces of texts, classifying them into important information classes, and displaying them in respective groups.

6. Proposed Technique

Clinical charts document a patient's clinical history with different types of information. Headings and subheadings are occasionally used in clinical documentation to indicate the organization of clinical documents. However, many clinical texts are long with very little structural demarcation; in such a case, modeling into multiple facets can be useful. In this paper, we consider 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 on the basis of 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 to 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 subheadings or smaller semantically coherent chunks.

6.1. Overview of Our Approach

Because physicians frequently review patients' clinical documentation made in the past, the goal of this study was to propose a novel technique for semantic modeling of clinical texts to support physicians in finding information in electronic clinical texts, as well as improving the accuracy of the retrieved information. In this section, we describe the proposed technique in detail. Our objective is to address the problem of visually organizing clinical text documents to help physicians review clinical text documents by modeling semantic classes of a patient's 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 precursor 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 four-dimensional semantic space. In addition to the idea of clusters, the cluster map needs to be organized such that the relationships between clusters are shown.

6.2. Design Requirements

On the basis of interviews and workshops with doctors, as well as a literature search, several tasks and design needs were determined. We were only interested in how doctors review information in electronic health records. As clinical decisions are often based on a patient's medical history, the relevant data elements we are interested in modeling are elements that describe patient's clinical events that occur during clinical encounters, as well as the clinical documentation format 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 (4) 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 visualize and display using a cluster map.

These requirements are summarized as follows:

- R1: Facilitate review of clinical text documents and make it easier for physicians to browse various types of information.
- R2: Visually present SOAP clinical notes sections in a cluster map, facilitating selective access of information.
- R3: Visually distinguish different semantic groups of information using different colors.
- R4: Group clinical texts with respect to SOAP documentation format.
- R5: Show relationships between different clusters of information.

The cluster map graphically presents document classes with the relationship between these classes.

6.3. SOAP Documentation Format

As mentioned in the previous section, the SOAP documentation format is made up of four sections: subjective, objective, assessment, and plan. The subjective part of SOAP is usually the background information of the patient which is required for understanding their current state. Objective is measurable and quantifiable information which can be analyzed. Assessment results from differential 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 <https://www.kaggle.com/datasets/tboyle10/medicaltranscriptions> (accessed on 31 January 2022) and contain a large collection of transcribed medical reports. This dataset was originally obtained from <https://mtsamples.com> (accessed on 31 January 2022). Table 1 shows the elements and descriptions of SOAP clinical notes.

Table 1. SOAP documentation format.

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, and series of events that happened in between
Objective	Quantifiable or measurable data obtained from past records and examinations, screening, and tests
Assessment	Possible diagnosis provided by the practitioners or the staff treating the patient
Plan	Treatment strategies, actions to be taken, and follow-up plans

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 sought 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 asked to complete a questionnaire to evaluate the prototype. For the assessment of clinical charts, a select group of 12 doctors were chosen. Patients charts with complicated illnesses and various comorbidities were chosen for this investigation.

7.2. Dataset

The clinical charts used in this paper were originally obtained from mimic3.org, which gives access to a large collection of transcribed medical reports. This dataset comprises 5000 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 doctors 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; therefore, we considered modeling the information in each section. Although these parts can be further divided into subsections, we only look at the four main aspects.

7.3. 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 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 practitioners. 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, in which users can look for particular information.

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.

7.4. Text Classification

Usually, humans organize information into groups or categories. Artificial intelligence follows the same principles using two broad types of algorithms: clustering and classification. In this paper, we adopted the classification algorithm to group clinical texts into various semantic groups inherent in clinical documentation. We relied on an a priori reference SOAP documentation structure that divides 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 multiclass 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. On the basis of 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 belongs to.

Clinical sentences in SOAP document are classified using a variant of a recurrent neural network known as long short-term memory network (LSTM). In a 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. A summary of our steps is presented below.

1. Tokenization—a collection of patient clinical text documents $D = \{d_1, d_2, \dots, d_n\}$ is split into a set of sentences $S = \{s_1, s_2, \dots, s_n\}$. 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 use word embedding to generate the required feature vectors for each sentence. Word embedding results in input features.
3. Input layer—these feature vectors are then used as input into the embedding layer of the 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 the 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 obtained from the abovementioned 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 was labeled with the help of medical professionals. The dataset was split into 80% for training and 10% for testing. Using the trained neural network, the sentences were classified into the four classes (subjective, objective, assessment, and plan) that were found to be relevant and useful clinical information in a clinical chart.

7.5. Cluster Map Generation

Our objective was to generate clusters of information containing similar sentences according to classification results. Therefore, a cluster should have sentences correctly classified in 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 labels leads to the need for 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 indicates 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 labeled with the same class so that their relevance can be immediately recognized. Figure 1 shows an example of a cluster map derived from clinical notes, with four clusters representing distinct semantic classes of information.

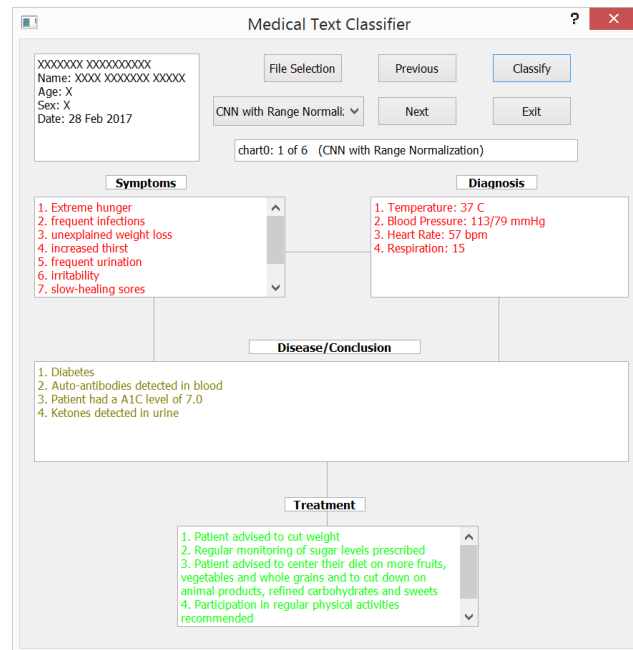


Figure 1. Sample cluster map.

8. Evaluation

The prototype was demonstrated to be effective in producing information groups that closely match human-generated subtopics from text documents. Tasks, such as information retrieval, should benefit from such a model. To validate the model for practical use, there was a 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 with 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) [102,103] was adopted. It consists of 10 questions evaluated on a five-point scale of level of agreement as shown in Table 2 below. To evaluate the prototypes 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), and 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 SUS scores for questions and then multiplying the resulting sum by 2.5 to obtain the overall SUS score [102]. This score usually ranges between 0 and 100, where 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 [103]. As a rule of thumb, a score above 70 indicates good usability, while a lower score indicates poor usability and that the system needs more improvement. It is a reliable, low-cost scale used for evaluating system usability [104].

Table 2. SUS questionnaire.

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

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 3 displays the results of the SUS score assessment.

Table 3. SUS results.

Participant	Questions										Sus Score
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	
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
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
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
10	4	1	2	2	4	3	3	2	3	2	65
11	4	2	4	2	4	2	2	2	4	2	70
12	3	2	3	1	3	1	4	1	4	1	77.5
Average											73.96

The final SUS score was 73.96, which indicates good usability on the basis of the above results. According to the SUS score scale [105], a SUS value of 73.96 is regarded as a good usability rating as shown in Figure 2 below.

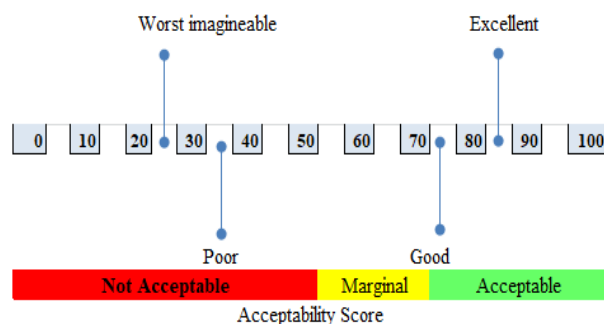


Figure 2. SUS score scale.

10. Discussion

The improved availability and accessibility of patient care data creates new possibilities for data use and reuse with the potential for improving the quality, safety, and efficiency of clinical work [106]. In this paper, we 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 is only 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, users can 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 patients 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 [106]. Additionally, this can contribute to decreasing cognitive load on the physicians part, reducing the time required to complete tasks and, thus, giving more time for face-to-face interaction with patients. This is in line with findings from earlier studies such as [38], which showed 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, requiring physicians to review them. When completing clinical tasks, electronic health records have been shown to increase physician cognitive workload [107], which, according to cognitive load theory, can lead to cognitive overload [108]. By organizing and visually presenting information, users can synthesize data into meaningful information.

11. Conclusions

We created a model to assist doctors in effectively gleaning insights into 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 by several factors. One 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 for changing the format of narrative texts. The structuring and visualization principles behind SOAP medical record structures were 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 great 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.

Author Contributions: Conceptualization, J.K.; methodology, J.K.; software, J.K.; validation, J.K.; formal analysis, J.K.; investigation, J.K.; resources, J.K.; data curation, J.K.; writing—original draft preparation, J.K.; writing—review and editing, E.O.; visualization, J.K.; supervision, E.O.; project administration, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Publicly available dataset was used in this study. The dataset (Medical transcription notes) is openly available at <https://www.kaggle.com/datasets/tboyle10/medicaltranscriptions> (accessed on 31 January 2022) [109].

Acknowledgments: The authors wish to thank the reviewers for their helpful comments. Their suggestions greatly improved the paper. We also thank the participants for providing insight into many clinical tasks and their input in designing the prototype. Cohen also 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. In Proceedings of the 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, 1–3 March 2018; pp. 1–5. [CrossRef]
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. [CrossRef]
3. Mustafa, A.; Rahimi Azghadi, M. Automated Machine Learning for Healthcare and Clinical Notes Analysis. *Computers* **2021**, *10*, 24. [CrossRef]
4. Muralidhar, E.S.; Gowtham, T.S.; Jain, A.; Padmaveni, K. Development of Health Monitoring Application using Machine Learning on Android Platform. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICES), Coimbatore, India, 10–12 June 2020; pp. 1076–1085. [CrossRef]
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. [CrossRef]
6. Singh, R.; Sharma, N. Machine Learning based Medical Information Analysis, Estimations and Approximations over Present Health Research Domain. In Proceedings of the 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2–4 July 2020; pp. 704–708. [CrossRef]
7. Massaro, A.; Maritati, V.; Savino, N.; Galiano, A. Neural Networks for Automated Smart Health Platforms oriented on Heart Predictive Diagnostic Big Data Systems. In Proceedings of the 2018 AEIT International Annual Conference, Bari, Italy, 3–5 October 2018; pp. 1–5. [CrossRef]
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. In Proceedings of the 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, 3–5 June 2020; pp. 718–723. [CrossRef]
9. Massaro, A.; Galiano, A.; Scarafile, D.; Vacca, A.; Frassanito, A.; Melaccio, A.; Solimando, A.; Ria, R.; Calamita, G.; Bonomo, M.; et al. Telemedicine DSS-AI Multi Level Platform for Monoclonal Gammopathy Assistance. In Proceedings of the 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Bari, Italy, 1 June–1 July 2020; pp. 1–5. [CrossRef]
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. *Eur. J. Gen. Pract.* **2017**, *23*, 257–260. [CrossRef]
11. Rostamzadeh, N.; Abdullah, S.S.; Sedig, K. Visual Analytics for Electronic Health Records: A Review. *Informatics* **2021**, *8*, 12. [CrossRef]
12. Wanderer, J.P.; Nelson, S.E.; Ehrenfeld, J.M.; Monahan, S.; Park, S. Clinical Data Visualization: The Current State and Future Needs. *J. Med. Syst.* **2016**, *40*, 1–9. [CrossRef]

13. Kuhn, T.; Basch, P.; Barr, M.; Yackel, T.; Medical Informatics Committee of the American College of Physicians. Clinical Documentation in the 21st Century: Executive Summary of a Policy Position Paper from the American College of Physicians. *Ann. Intern. Med.* **2015**, *162*, 301–303. [[CrossRef](#)]
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), Laguna Hills, CA, USA, 4–6 February 2016; pp. 321–324.
15. Spasic, I.; Nenadic, G. Clinical Text Data in Machine Learning: Systematic Review. *JMIR Med. Inform.* **2020**, *8*, e17984. [[CrossRef](#)]
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, Minneapolis, MN, USA, 3–6 September 2009; pp. 5905–5908.
17. Baron, R.J. Doctors Stories: The Narrative Structure of Medical Knowledge. Literature and Medicine. *Johns Hopkins Univ. Press* **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. *J. Am. Med. Inform. Assoc.* **2008**, *15*, 54–64. [[CrossRef](#)]
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. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 141–145. [[CrossRef](#)]
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; Now Publishers, Inc.: Delft, The Netherlands, 2013.
21. Lesselroth, B.J.; Pieczkiewicz, D.S. *Data Visualization Strategies for the Electronic Health Record*; Nova Science Publishers: New York, NY, USA, 2011.
22. De Oliveira, J.M.; da Costa, C.A.; Antunes, R.S. Data Structuring of Electronic Health Records: A Systematic Review. *Health Technol.* **2021**, *11*, 1219–1235. [[CrossRef](#)]
23. Venkataraman, G.R.; Pineda, A.L.; Bear Dont Walk, O.J., IV; 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. [[CrossRef](#)]
24. Lin, W.; Ji, D.; Lu, Y. Disorder Recognition in Clinical Texts Using Multi-Label Structured SVM. *BMC Bioinform.* **2017**, *18*, 75. [[CrossRef](#)]
25. Pivovarov, R.; Elhadad, N. Automated Methods for the Summarization of Electronic Health Records. *J. Am. Med. Inform. Assoc.* **2015**, *22*, 938–947. [[CrossRef](#)]
26. Caban, J.J.; Gotz, D. Visual Analytics in Healthcare—Opportunities and Research Challenges. *J. Am. Med. Inform. Assoc.* **2015**, *22*, 260–262. [[CrossRef](#)]
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 *Pacific Symposium on Biocomputing 2020*; World Scientific: Hackensack, NJ, USA, 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 Med. Res. Methodol.* **2019**, *19*, 155. [[CrossRef](#)]
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. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 1–9. [[CrossRef](#)]
30. Liu, S.; Wang, Y.; Wen, A.; Wang, L.; Hong, N.; Shen, F.; Bedrick, S.; Hersh, W.; Liu, H. 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 Med. Inform.* **2020**, *8*, e17376. [[CrossRef](#)]
31. Murdoch, T.B.; Detsky, A.S. The Inevitable Application of Big Data to Health Care. *JAMA* **2013**, *309*, 1351–1352. [[CrossRef](#)]
32. Jarabek, B.; Mink, P.; Winden, T.; Bork, L.; Elison, J.T.; Finley, G.; Giaquinto, R.; Hultman, G.M.; Lindemann, E.A.; McEwan, R.; et al. *Discovery and Visualization of New Information from Clinical Reports in the EHR*; AHRQ Grant Final Progress Report—Grant No.: R01HS022085; Agency for Healthcare Research and Quality: Rockville, MD, USA, 2019.
33. Lauster, C.D.; Srivastava, S.B. *Fundamental Skills for Patient Care in Pharmacy Practice*; Jones & Bartlett Publishers: Burlington, MA, USA, 2013.
34. 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*, 1–40. [[PubMed](#)]
35. 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. *J. Am. Med. Inform. Assoc.* **2021**, *28*, 899–906. [[CrossRef](#)]
36. 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. *Appl. Netw. Sci.* **2019**, *4*, 1–23. [[CrossRef](#)]
37. Li, Y.; Lipsky Gorman, S.; Elhadad, N. Section Classification in Clinical Notes Using Supervised Hidden Markov Model. In Proceedings of the 1st ACM International Health Informatics Symposium, Arlington, VA, USA, 11–12 November 2010; pp. 744–750.
38. Pollack, A.H.; Pratt, W. Association of Health Record Visualizations with Physicians Cognitive Load When Prioritizing Hospitalized Patients. *JAMA Netw. Open* **2020**, *3*, e1919301. [[CrossRef](#)] [[PubMed](#)]

39. 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 Technol. Interact.* **2020**, *4*, 7. [[CrossRef](#)]
40. 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 2018 CHI Conference on Human Factors in Computing Systems, New York, NY, USA, 21–26 April 2018; pp. 1–13.
41. Shah, A.D.; Martinez, C.; Hemingway, H. The Freertext Matching Algorithm: A Computer Program to Extract Diagnoses and Causes of Death from Unstructured Text in Electronic Health Records. *BMC Med. Inform. Decis. Mak.* **2012**, *12*, 88. [[CrossRef](#)]
42. Deng, Y.; Denecke, K. Visualizing Unstructured Patient Data for Assessing Diagnostic and Therapeutic History. In *e-Health—For Continuity of Care*; IOS Press: Amsterdam, The Netherlands, 2014; Volume 205, pp. 1158–1162.
43. 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 Aff.* **2005**, *24*, 1103–1117. [[CrossRef](#)]
44. Nguyen, L.; Bellucci, E.; Nguyen, L.T. Electronic Health Records Implementation: An Evaluation of Information System Impact and Contingency Factors. *Int. J. Med. Inform.* **2014**, *83*, 779–796. [[CrossRef](#)]
45. Pai, M.M.; Ganiga, R.; Pai, R.M.; Sinha, R.K. Standard Electronic Health Record (EHR) Framework for Indian Healthcare System. *Health Serv. Outcomes Res. Methodol.* **2021**, *21*, 339–362. [[CrossRef](#)]
46. Kubben, P.; Dumontier, M.; Dekker, A. *Fundamentals of Clinical Data Science*; Springer: Cham, Switzerland, 2019.
47. Wang, Z.; Shah, A.D.; Tate, A.R.; Denaxas, S.; Shawe-Taylor, J.; Hemingway, H. Extracting Diagnoses and Investigation Results from Unstructured Text in Electronic Health Records by Semi-Supervised Machine Learning. *PLoS ONE* **2012**, *7*, e30412. [[CrossRef](#)]
48. Liang, J.J.; Tsou, C.-H.; Poddar, A. A Novel System for Extractive Clinical Note Summarization Using EHR Data. In Proceedings of the 2nd Clinical Natural Language Processing Workshop, Minneapolis, MN, USA, 7 June 2019.
49. 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. *Annu. Rev. Public Health* **2016**, *37*, 61–81. [[CrossRef](#)]
50. Jensen, P.B.; Jensen, L.J.; Brunak, S. Mining Electronic Health Records: Towards Better Research Applications and Clinical Care. *Nat. Rev. Genet.* **2012**, *13*, 395–405. [[CrossRef](#)]
51. Ning, X.; Fan, Z.; Burgun, E.; Ren, Z.; Schleyer, T. Improving Information Retrieval from Electronic Health Records Using Dynamic and Multi-Collaborative Filtering. *PLoS ONE* **2021**, *16*, e0255467. [[CrossRef](#)]
52. Kosara, R.; Miksch, S. Visualization Methods for Data Analysis and Planning in Medical Applications. *Int. J. Med. Inform.* **2002**, *68*, 141–153. [[CrossRef](#)]
53. 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 Emerg. Care* **2015**, *31*, 536–541. [[CrossRef](#)]
54. Chou, S.; Chang, W.; Cheng, C.-Y.; Jehng, J.-C.J.; Chang, C. An Information Retrieval System for Medical Records & Documents. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 1474–1477.
55. 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. *J. Asthma* **2018**, *55*, 1068–1076. [[CrossRef](#)]
56. 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 *Symposium of the Austrian HCI and Usability Engineering Group*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 301–320.
57. Blei, D.M.; Ng, A.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
58. Deerwester, S.C.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R.A. Indexing by Latent Semantic Analysis. *J. Assoc. Inf. Sci. Technol.* **1990**, *41*, 391–407. [[CrossRef](#)]
59. Hofmann, T. Probabilistic Latent Semantic Analysis. In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden, 30 July–1 August 1999.
60. 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*, 131–147.
61. Chan, K.R.; Lou, X.; Karaletsos, T.; Crosbie, C.; Gardos, S.M.; Artz, D.; Rättsch, G. An Empirical Analysis of Topic Modeling for Mining Cancer Clinical Notes. In Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, USA, 7–10 December 2013; pp. 56–63.
62. Arnold, C.W.; El-Saden, S.M.; Bui, A.A.T.; Taira, R.K. Clinical Case-Based Retrieval Using Latent Topic Analysis. *AMIA Annu. Symp. Proc.* **2010**, *2010*, 26–30.
63. Perotte, A.; Wood, F.; Elhadad, N.; Bartlett, N. Hierarchically Supervised Latent Dirichlet Allocation. In *Advances in Neural Information Processing Systems 24*; Curran Associates, Inc.: Red Hook, NY, USA, 2011; pp. 1–9.
64. Coroiu, A.M.; Călin, A.D.; Nu, M. Topic Modeling in Medical Data Analysis. In Case Study Based on Medical Records Analysis. In Proceedings of the 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 19–21 September 2019; pp. 1–5.
65. Liu, X.; Alharbi, M.; Best, J.; Chen, J.; Diehl, A.; Firat, E.E.; Rees, D.; Wang, Q.; Laramée, R.S. Visualization Resources: A Starting Point. In Proceedings of the 2021 25th International Conference Information Visualisation (IV), Sydney, Australia, 5–9 July 2021; pp. 160–169.

66. Ledesma, A.; Bidargaddi, N.; Strobel, J.; Schrader, G.; Nieminen, H.; Korhonen, I.; Ermes, M. Health Timeline: An Insight-Based Study of a Timeline Visualization of Clinical Data. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 170. [[CrossRef](#)] [[PubMed](#)]
67. Grobelnik, M.; Mladenic, D. Visualization of News Articles. *Informatica* **2004**, *28*, 375–380.
68. Imai, T.; Nakamura, K.; Ohmameuda, T. Visualization of Similar News Articles with Network Analysis and Text Mining. In Proceedings of the 2015 IEEE 4th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 27–30 October 2015; pp. 151–152.
69. Doshi, K.; Gokhale, S.; Mamtara, H.; Bide, P. Analytics and Visualization of Trends in News Articles. In Proceedings of the 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), Mumbai, India, 20–21 December 2019; pp. 1–9.
70. Roque, F.; Slaughter, L.; Tkatenko, A. A Comparison of Several Key Information Visualization Systems for Secondary Use of Electronic Health Record Content. In Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, Los Angeles, CA, USA, 5 June 2010; pp. 76–83.
71. Powsner, S.M.; Tufte, E.R. Graphical Summary of Patient Status. *Lancet* **1994**, *344*, 386–389. [[CrossRef](#)] [[PubMed](#)]
72. Plaisant, C.; Milash, B.; Rose, A.; Widoff, S.; Shneiderman, B. LifeLines: Visualizing Personal Histories. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 13–18 April 1996; pp. 221–227.
73. 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 Trans. Vis. Comput. Graph.* **2009**, *15*, 1049–1056. [[CrossRef](#)] [[PubMed](#)]
74. 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 Jt. Summits Transl. Sci. Proc.* **2013**, *2013*, 168–172.
75. 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. In Proceedings of the Working Conference on Advanced Visual Interfaces, Gallipoli, Italy, 25–28 May 2004; pp. 171–174.
76. Hallett, C. Multi-Modal Presentation of Medical Histories. In Proceedings of the 13th International Conference on Intelligent User Interfaces, Gran Canaria, Spain, 13–16 January 2008; pp. 80–89.
77. Miksch, S.; Kosara, R.; Shahar, Y.; Johnson, P.D. AsbruView: Visualization of Time-Oriented, Skeletal Plans. In Proceedings of the Artificial Intelligence Planning Systems, Pittsburgh, PA, USA, 7–10 June 1998; pp. 11–18.
78. Wongsuphasawat, K.; Guerra Gómez, J.A.; Plaisant, C.; Wang, T.D.; Taieb-Maimon, M.; Shneiderman, B. LifeFlow: Visualizing an Overview of Event Sequences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 1747–1756.
79. Meyer, T.E.; Monroe, M.; Plaisant, C.; Lan, R.; Wongsuphasawat, K.; Coster, T.S.; Gold, S.; Millstein, J.; Shneiderman, B. Visualizing Patterns of Drug Prescriptions with Eventflow: A Pilot Study of Asthma Medications in the Military Health System. In Proceedings of the Workshop on Visual Analytics in HealthCare, Washington, DC, USA, 16 November 2013; pp. 1–4.
80. Kenei, J.K.; Opiyo, E.T.O.; Oboko, R.O. Visualizing Semantic Structure of a Clinical Text Document. *Eur. J. Electr. Eng. Comput. Sci.* **2020**, *4*, 6. [[CrossRef](#)]
81. Sellars, B.B.; Sherrod, D.R.; Chappel-Aiken, L. Using Word Clouds to Analyze Qualitative Data in Clinical Settings. *Nurs. Manag.* **2018**, *49*, 51–53. [[CrossRef](#)]
82. 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? *Appl. Clin. Inform.* **2019**, *10*, 189–198. [[CrossRef](#)]
83. Foldes, D. Using Tag Clouds as a Tool for Patients Medical History Visualization and Record Retrieval. Ph.D. Thesis, Concordia University, Montréal, QC, Canada, 2015.
84. Llopis, F.; Ferrández, A.; Vicedo, J.L. Text Segmentation for Efficient Information Retrieval. In *Proceedings of the Third International Conference, CICLing 2002, Mexico City, Mexico, 17–23 February 2002*; Lecture Notes in Computer Science; Gelbukh, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2002; Volume 2276, pp. 373–380.
85. Salton, G. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1989.
86. Edinger, T.; Demner-Fushman, D.; Cohen, A.M.; Bedrick, S.; Hersh, W.R. Evaluation of Clinical Text Segmentation to Facilitate Cohort Retrieval. *AMIA Annu. Symp. Proc.* **2017**, *2017*, 660–669.
87. Ganesan, K.; Subotin, M. A General Supervised Approach to Segmentation of Clinical Texts. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27–30 October 2014; pp. 33–40.
88. Tepper, M.; Capurro, D.; Xia, F.; Vanderwende, L.; Yetisgen-Yildiz, M. Statistical Section Segmentation in Free-Text Clinical Records. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC12), Istanbul, Turkey, 23–25 May 2012; pp. 2001–2008.
89. Mowery, D.L.; Wiebe, J.; Visweswaran, S.; Harkema, H.; Chapman, W.W. Building an Automated SOAP Classifier for Emergency Department Reports. *J. Biomed. Inform.* **2012**, *45*, 71–81. [[CrossRef](#)]
90. Bashyam, V.; Hsu, W.; Watt, E.; Bui, A.A.T.; Kangaroo, H.; Taira, R.K. Problem-Centric Organization and Visualization of Patient Imaging and Clinical Data. *Radiographics* **2009**, *29*, 331–343. [[CrossRef](#)]

91. Choi, Y.J.; Byun, J.; Berkovich, S. Cross-Search Technique and Its Visualization of Peer-to-Peer Distributed Clinical Documents. *Trans. Eng. Comput. Technol.* **2004**, *3*, 59–63.
92. Fan, Z.; Burgun, E.; Schleyer, T.; Ning, X. Improving Information Retrieval from Electronic Health Records Using Dynamic and Multi-Collaborative Filtering. In Proceedings of the 2019 IEEE International Conference on Healthcare Informatics (ICHI), Xian, China, 10–13 June 2019; pp. 1–3. [CrossRef]
93. Furlow, B. Information Overload and Unsustainable Workloads in the Era of Electronic Health Records. *Lancet Respir. Med.* **2020**, *8*, 243–244. [CrossRef]
94. Wei, X.; Eickhoff, C. Embedding Electronic Health Records for Clinical Information Retrieval. *arXiv* **2018**, arXiv:1811.05402.
95. 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 Med. Inform.* **2019**, *7*, e12239. [CrossRef]
96. Kuhn, L.; Eickhoff, C. Implicit Negative Feedback in Clinical Information Retrieval. *arXiv* **2016**, arXiv:1607.03296. [CrossRef]
97. 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). *JMIR Med. Inform.* **2018**.
98. Tayefi, M.; Ngo, P.D.; Chomutare, T.; Dalianis, H.; Salvi, E.; Budrionis, A.; Godtliessen, F. Challenges and Opportunities beyond Structured Data in Analysis of Electronic Health Records. *Wiley Interdiscip. Rev. Comput. Stat.* **2021**, *13*, e1549. [CrossRef]
99. 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. In Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Ottawa, ON, Canada, 12–13 June 2018; pp. 1–6.
100. Holzinger, A.; Schantl, J.; Schroettner, M.; Seifert, C.; Verspoor, K. Biomedical Text Mining: State-of-the-Art, Open Problems and Future Challenges. In *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*; Lecture Notes in Computer Science 8401; Holzinger, A., Jurisica, I., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 271–300.
101. Sultanum, N.; Singh, D.; Brudno, M.; Chevalier, F. Doccurate: A Curation-Based Approach for Clinical Text Visualization. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 142–151. [CrossRef]
102. Peres, S.C.; Pham, T.; Phillips, R.G. Validation of the System Usability Scale (SUS). *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2013**, *57*, 192–196. [CrossRef]
103. Brooke, J.B. SUS—A retrospective. *J. Usability Stud.* **2013**, *8*, 29–40.
104. Brooke, J. Chapter SUS—A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*, 1st ed.; Taylor & Francis: London, UK, 1996; pp. 189–194.
105. Bangor, A.; Kortum, P.T.; Miller, J.T. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Stud. Arch.* **2009**, *4*, 114–123.
106. Hilligoss, B.; Zheng, K. Chart Biopsy: An Emerging Medical Practice Enabled by Electronic Health Records and Its Impacts on Emergency Department-Inpatient Admission Handoffs. *J. Am. Med. Inform. Assoc. JAMIA* **2013**, *20*, 260–267. [CrossRef] [PubMed]
107. 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 Netw. Open* **2019**, *2*, e191709. [CrossRef]
108. Tawfik, A.A.; Kochendorfer, K.M.; Saporova, D.; Al Ghenaimi, S.; Moore, J.L. Using Semantic Search to Reduce Cognitive Load in an Electronic Health Record. In Proceedings of the 2011 IEEE 13th International Conference on e-Health Networking, Applications and Services, Columbia, MO, USA, 13–15 June 2011; pp. 181–184.
109. Medical Transcriptions. Available online: <https://www.kaggle.com/datasets/tboyle10/medicaltranscriptions> (accessed on 31 January 2022).