Clinical Text Analysis using Visual Analytics for Cancer Patient Cohort Identification

by

Saja Ibrahim Al-Alawneh

Bachelor of Computer Science, Yarmouk University, 2003

Master of Computer Science, Yarmouk University, 2005

Master of Science, University of Pittsburgh, 2018

Submitted to the Graduate Faculty of

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

UNIVERSITY OF PITTSBURGH

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Saja Ibrahim Al-Alawneh

It was defended on

July 22, 2021

and approved by

Dr. Gerald Douglas; Assistant Professor, Department of Biomedical Informatics

Dr. Jonathan Sliverstein; MD, MS, FACS, FACMI Biomedical Informatics

Dr. Adrian Lee, Professor Pharmacology and Chemical Biology

Dissertation Director: Dr. Harry Hochheiser, Associate Professor, Biomedical Informatics

Clinical Text Analysis using Visual Analytics for Cancer Patient Cohort Identification

Saja Ibrahim Al-Alawneh, PhD

University of Pittsburgh, 2021

Due to the complexity nature of cancer patients' records and clinical notes, extracting and summarizing the required data to identify a cohort of interest is a challenge for cancer researchers. DeepPhe pipeline was developed to support cancer cohort identification and hypothesis generation by extracting deep phenotypes from cancer patient' electronic health records using natural language processing, rule-based and machine learning techniques. The pipeline generated high-level summaries from individual mentions as phenotypes and visualized them using a web-based visual analytics interface DeepPhe-Viz to create a longitude representation of cancer process.

In this study, we extended the functionalities of DeepPhe-Viz interface by combining the extracted data from the NLP pipeline and visual analytics to empower researcher's ability to mine and uncover new and challenging insights about cancer population in the textual documents of the EMR data.

To advance the capabilities of the DeepPhe-Viz, first, we implemented an interactive heatmap visualization that viewed high-level representation of all the extracted terms from clinical documents. This feature enabled cancer researchers to investigate the rich contents of the full clinical document text to identify additional key variables such treatment that are extracted using the DeepPhe pipeline. Second, we implemented an interactive Sankey diagram visualization to aggregate all the transitions of predefined episodes of care for a cohort of cancer

patient which represent temporal events in the clinical documents. This feature is essential for gaining a deeper understanding of different patterns and trends of cancer treatment mentioned in EMR.

Finally, we evaluated the usability of DeepPhe-Viz interface to identify a cohort of patients and to drill down to more details about patients. User studies results including qualitive and quantitative feedback indicated the usefulness and feasibility of the DeepPhe-Viz interface to the cancer investigators to conduct cancer retrospective studies using EMR data

Table of Contents

List of Tables

List of Figures

List of Abbreviations

EMR: **E**lectronic **M**edical **R**ecord

NLP: **N**atural **L**anguage **P**rocessing

cTAKES: **c**linical **T**ext **A**nalysis **K**nowledge **E**xtraction **S**ystem

caTIES: **ca**ncer **T**ext **I**nformation **E**xtraction **S**ystem

UPMC: **U**niversity of **P**ittsburgh **M**edical **C**enter

UMLS: **U**nified **M**edical **L**anguage **S**ystem

REX: **R**egenstrief **EX**traction

RADA: **RAD**iology **A**nalysis

eMERGE: **e**lectronic **ME**dical **R**ecord and **GE**nomic

i2b2: **i**nformatics for **i**ntegrating **b**iology and the **b**edside

DeepPhe: **Deep Phe**notype

DeepPhe-Viz: **Deep Phe**notype **Vi**sual**iz**ation

LSP-ML: **L**inguistic **S**tring **P**roject – **M**edical **L**anguage

MedLEE: **Med**ical **L**anguage **E**xtraction and **E**ncoding

HiTEX: **H**ealth **i**nformation **TE**xt e**X**traction

NegEx: **Neg**ation **Ex**tensions

ETL: **E**xtract **T**ransform **L**oad

BioMedICUS: **BioMed**ical **I**nformation **C**ollection and **U**nderstanding **S**ystem

SNOMED-CT: **S**ystematized **NOM**enclature of **MED**icine -**C**linical **T**erms

COAT: **C**linical **O**utcomes **A**ssessment **T**oolkit

EpiDEA: **Epi**lepsy **D**ata **E**xtraction and **A**nnotation

GATE: **G**eneral **A**rchitecture for **T**ext **E**ngineering

LEXIMER: **LEXI**con Mediated **E**ntropy **R**eduction

MedEx: **Med**ication **Ex**traction

UIMA: **U**nstructured **I**nformation **M**anagement **A**rchitecture

EHR: **E**lectronic **H**ealth **R**ecord

FHIR: **F**ast **H**ealthcare **I**nteroperability **R**esources

NCIt: **N**ational **C**ancer **I**nstitute **t**hesaurus

API: **A**pplication **P**rogramming **I**nterface

POS: **P**art **O**f **S**peech

NoSQL: **N**ot **o**nly **S**tructured **Q**uery **L**anguage

SQL: **S**tructured **Q**uery **L**anguage

CoCo: **Co**hort **Co**mparison

CAVA: **C**ohort **A**nalysis via **V**isual **A**nalytics

COQUITO: **CO**hort **QU**eries with an **IT**erative **O**verview

Preface

Earning a PhD is indeed an incredible feeling of personal gratification. However, this journey was not an easy one for me. I did it while being a wife of a very supportive husband, mother of three amazing kids, and a survivor of the fight of my life after cancer diagnosis. I could not achieve this without the support of many people.

First and foremost, I would like to express my deep and sincere gratitude to my research advisor, Dr. Harry Hochheiser, who provided me with valuable advice and support throughout this journey. His vision and motivation have deeply inspired me. He has taught me to carry out the research and to present the research works as clearly as possible. I am extremely appreciative for mentorship, understanding, and encouragement through my PhD project.
I would also like to give a special thanks to my committee member: Dr. Adrian Lee, Dr. Gerry Douglas, and Dr. Jonathan Sliverstein. I am so honored you agreed to be a part of my dissertation.

I gratefully acknowledge the National Library of Medicine and the Biomedical Informatics Training Program at the University of Pittsburgh for their financial support and the opportunity to earn these degrees and Dr. Michael Becich for his guidance and funding support for the last year.

Thank you to all the staff and administrators of DBMI and participants from university of Pittsburgh and UPMC who helped make this work possible. A very special thanks also goes to Toni Porterfield; I appreciate all she did to support this work. Thank you also to my fellow students and friends who have contributed in one way or another: Sanghoon Lee, and Luca Calzoni.

Special thanks to my caring husband Mohammad Al-Bataineh for being my greatest advocate, even when I didn't believe in myself, my wonderful kids Omar and Nisreen and my son Zaid who I had during my PhD journey, he was my sidekick at home when I was working on my PhD project. My family and friends standing beside me with their unconditional support and love. Thank you all for tolerating my ignorance and for helping me keep things in perspective and for the love and patience you showed me during my pursuit of this degree.

Dedication

To my late Father Dr. Ibrahim Al-alawneh. The first who inspired me to be who I am.

I hope you would be proud

1.0 Introduction

Retrospective cancer cohort identification is the process of identifying groups of cancer patients from cancer populations who meet often complex criteria to conduct clinical research studies[1]. Currently, patient health information is commonly stored and captured by electronic medical records (EMR) in structured and unstructured (free text notes) formats electronically[2]. These electronic data contain extensive details about patient's information, history, and records of cancer care. Cancer patients go through multiple cycle of diagnosis to treatment over several encounters and observations that generate a rich set of clinical notes. Extraction details about the patients in a usable form requires manual abstraction process from the clinical notes. This process can be extremely time-consuming and challenging due to the volume and the complexity of the clinical notes. Furthermore, this process is inflexible and inefficient because curation is generally performed with a specific set of goals and revising these goals often requires re-curation of the data where researchers need to re-review manually the clinical notes to identify patients who meet the new criteria. Furthermore, for some patients who experience cancer recurrence or undergo different combinations of treatment hundreds or thousands of notes can be generated along the cancer journey[3]. As the number of clinical documents grows larger, it becomes increasingly difficult for cancer researchers to track the connections between disparate facts or concepts through patient cancer journey timeline and make sense of it all[4]. Some information about cancer patients is stored in structured form while other details are stored in clinical notes narratives. These details including comorbidities, treatment histories and response and medications are not easily accessible using computational approaches.

To better assist translational investigators in performing retrospective cancer research, automated methods are needed to extract and view the rich data in the free clinical text from electronic medical records (EMR). Clinical natural language processing (NLP) has been widely used to resolve these challenges [5]. Clinical NLP pipeline focuses on developing various methods for semantic processing and analysis of clinical texts and can thus be applied to a variety of applications[2]. Currently, a number of clinical NLP systems including Clinical Text Analysis and Knowledge Extraction System (cTAKES)[6] and Cancer Tissue Information Extraction System (caTIES)[7] have been successfully developed and implemented in various medical domains. While cTAKES focuses on extracting clinical terms from the free text of the EMRs, caTIES focuses on extracting information from surgical pathology reports.

Other systems used NLP in various domains to explore and identify keywords in clinical text. Sohn et al[8] used a complex NLP pipeline along with clinical knowledge and heuristics approaches to identify complex cases of drug side effects. They also developed another system that leverages NLP using cTAKES along with decision trees to extract side effect keyword features from psychiatry and psychology patients EMR records[8]. Bejan et al[9] used MetaMap[10] to identify pneumonia cases by extracting UMLS concepts unigrams and bigrams from clinical reports and applied a statistical feature selection technique to rank the relevance of the concepts. The REX[11] system applied a combination of rule-based system and regular expression to identify patients with pancreatic cancer. The RADA[12] system demonstrates the utilization of predefined rules and domain dictionary to extract medical concepts and attributes from radiology reports.

Previous initiatives have been developed to tackle the need for identifying cohorts of patients using structured and unstructured data to perform retrospective studies. The eMERGE[13] and the i2b2[14] systems demonstrated scalable computational framework using natural language

processing to identify cohorts patients with non-cancer phenotypes and identified patients with a specific existing phenotype[15,16]. However, there is still a need for a generalizable computational infrastructure that extracts mention of cancer related terms and links these mentions into high-level summaries. These summaries utilized multiple level representations of cancer mentions by aggregating these mentions to high-level phenotype summaries. This modeling approach uses multiple levels of abstraction to facilitate the generation of longitudinal phenotype and treatment data for cancer patients [16].

To address this need, DeepPhe, a computational pipeline, was developed using advanced natural language processing, knowledge engineering methods and machine learning algorthims[17] to automate the extraction process of detailed phenotype information from cancer EMR, then generate high-level summaries of documents from clinical mentions and observations, and finally classify clinical documents into different episodes of care using important events stated in the clinical documents[16]. The DeepPhe system also includes DeepPhe-Viz, an interactive phenotype exploring visualization interface to support cohort identification and analysis for cancer research[18]. DeepPhe-Viz provides a preliminary visualization of cancer patients' information including tumor summary, cancer stage, diagnosis, age, laterality, and other details. However, many information that are extracted using DeepPhe pipeline are not visualized using DeepPhe-Viz. This information provides additional information to support the navigation of the rich and computable representations of cancer patient trajectories.

In this project, we built upon the DeepPhe NLP platform and expanded the DeepPhe-Viz visualization features to develop visual representations of (1) information within the free text of the clinical documents to provide visual summary of clinical documents contents for all the patients in the cohort and (2) model the temporal events of different episodes of care for cancer

cohort using visual analytics approaches to identify different patterns of care for all patients in the cohort.

1.1 Dissertation Overview

In Chapter 2, I present an overview of the literature on clinical NLP, the structure of DeepPhe cancer Phenotype. I then review current visual analytics tools and approaches that are developed for diverse tasks, and I also discuss prior work on combing visual analytics and NLP in healthcare including DeepPhe-Viz. Chapters 3 demonstrates the hypothesis, the aims of the study, and significance and Innovation of the project. Chapter 4 presents the new proposed design for the DeepPhe-Viz tool. Chapter 5 describes the user studies conducted with translational cancer investigators to evaluate the revised version of the DeepPhe-Viz with the new integrated features. Chapter 6 summarizes the results of the user study including the task time completion, task correctness and the qualitative assessment. Finally, chapter 7 includes a discussion of the completed work, define the limitations, suggest future directions, and present final conclusions.

2.0 Background

This chapter provides a review of the preceding work and literature relevant to this dissertation. Section 2.1 reviews the prior work on the clinical NLP systems, including the integration of machine learning, and rule-based approaches. Section 2.2 provides an overview of the DeepPhe-NLP pipeline, including the goals, infrastructure, and different levels of data extraction and summarization. Section 2.3 describes how the DeepPhe-Viz visual analytics tools were developed, and explains the structure of the DeepPhe-Viz. Section 2.4 reviews the different visual analytics tools, in terms of the performed tasks, the type and source of data used by these systems

2.1 Clinical Natural Language Processing

The widespread adoption of electronic medical records (EMRs) has led to the collection of rich data documenting the cancer patient's care journey[18].This EMR data includes structure and unstructured data type and much of these data, especially narrative reports are "locked" within free text (unstructured) provide rich and important insights about different aspects of the cancer diagnosis, treatment, and prognosis.

Given the rate at which unstructured clinical information has been created, automated solutions utilizing NLP are needed to analyze this text and generate structured representations of the data. This data processing is useful in applications such as clinical decision support, cohort identification, as well as public health surveillance and quality assurance[19].

The clinical NLP architecture, according to Friedman and Elhadad[20], contains two main components: background knowledge and a framework that integrates NLP tools and modules (Figure 1). The background knowledge consists of trained corpora, domain models, domain knowledge, and linguistic knowledge while the framework contains methods, tools, systems, and applications.

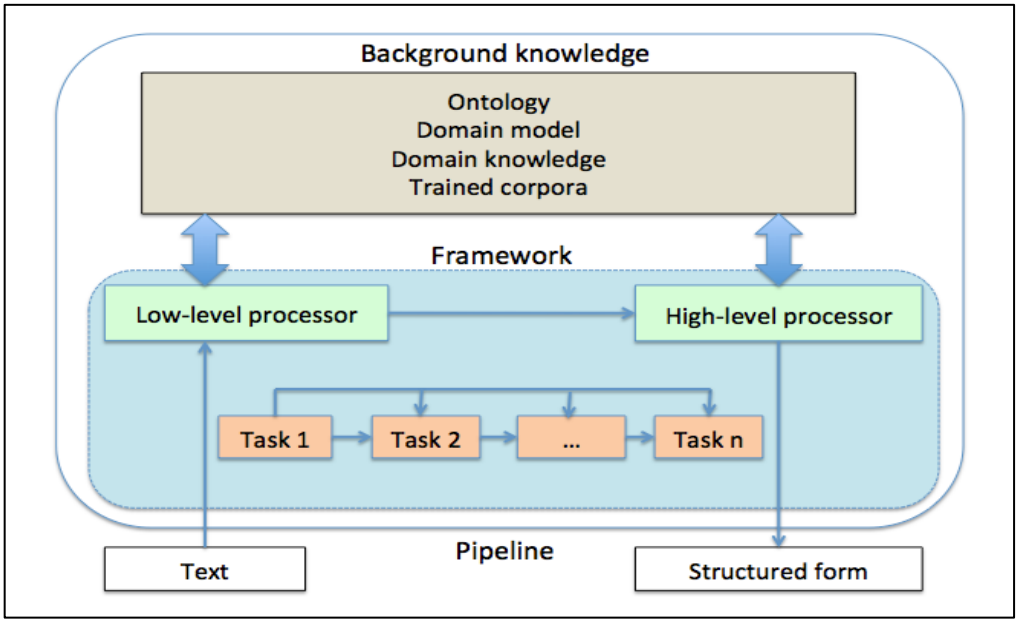


Figure 1. Clinical NLP pipeline architecture. *Two main parts: background knowledge and framework* [19].

Table 1 provides a list of some examples of clinical NLP systems that have been developed previously with brief descriptions. Early NLP systems including LSP-ML[21, 22] (the Linguistic String Project (LSP) – Medical Language Processor) and MedLEE[5] (Medical Language Extraction and Encoding System), as well as more recent open source systems such as cTAKES[6] and Health Information Text Extraction (HiTEX)[23, 24] have proposed automated methods to extract information from free text, including clinical narratives using both rule-based and hybrid approaches. MetaMap[25] focused on extracting concepts from clinical notes and

recognized the different contextual and semantic information about these concepts such as negated concepts using additional algorithms including NegEx[26].

Other clinical NLP systems extracted concepts, detected assertion and classified relations from clinical text[27]. For example, ClinRead, a rule-based NLP system, was used to categorize metastatic prostate cancer patients into four different levels of pain from clinical records[28]. Skeppstedt et al.[29] extracted mentions of finding, disorder, body structure and drugs from clinical narratives by constructing an annotated corpus and then using named entity recognition methods. Other clinical NLP systems focused on extracting data from clinical narratives to build a cohort of patients. While Botsis et al.[30] utilized EMR data for a retrospective creation of a pancreatic cancer cohort, Xu et al.[31] used structured and unstructured EMR data to automatically detect patient with colon cancer using the Extract Transform Load (ETL) workflow. Furthermore, Kreuzthaler et al.[27] used regular expression and rule-based NLP system to extract information from clinical narratives into a structured template and to build a cohort study on metabolic syndrome.

Table 1. Examples of clinical NLP systems

System Name	Brief Description	Evaluation document type
BioMedICUS [32]	Identify family history statements in clinical notes using rules generated by a stochastic gradient descent classifier	History notes Physical notes
caTIES[7]	A GATE-based system for annotating pathology reports with terms from NCI Thesaurus	Pathology reports
ChartIndex[33]	Identify noun phrases in radiology or pathology reports and map them to SNOMED-CT or UMLS terms by adopting Stanford Parser and MetaMap pipelines	Surgical pathology reports for extracting anatomic site and findings/diagnosis
ClearForest[34]	Identify certain diagnoses along with dates and laterality using rules in breast cancer pathology reports	Pathology reports
ClinREAD[35]	Categorize the pain status in patients with metastatic prostate cancer into different groups using rule-based approach	Cancer patients record
COAT[36]	A framework that combines rule-based and machine learning to process clinical note and extract Gleason score, tumor stage, and surgical margin status from pathology reports	pathology reports
EpiDEA[37-39]	Extract and structure data from Epilepsy patient clinical notes using an extension of cTAKES and incorporating the Epilepsy Seizure Ontology band	Manually annotated discharge summaries
GATE[40, 41]	An extensive open-source framework for text processing and text extraction pipeline with additional components	Extract mini mental state exam scores and dates from mental health records
HITEx[19]	Utilize the GATE infrastructure to extract normalized diagnostic and family history terms from clinical notes and discharge summaries	Discharge summaries
LEXIMER[42, 43]	Detect clinical finding and recommendation in CT and MR imaging reports using machine learning approach	Radiology reports
MedLEE[5]	An early rule-based system for structuring different all types of clinical notes	Clinical documents
MedEx[19, 44]	An UIMA-based system that extracts medication mentions	Discharge summaries
SymText MPlus[45]	Extract and normalize findings from radiography reports using Bayesian network-based semantic grammar	Radiography reports to detect pneumonia

The previous works mentioned above represented different systems that applied traditional NLP techniques to extract cancer and non-cancer phenotypes from both structure and unstructured clinical notes. Some of these systems annotated different temporal mentions and relations among concepts and construct summaries into dashboards from the structured part of the EHR[46].

However, these systems were used to extract individual mention and didn't support any aggregation process across multiple level of extracting, collecting and synthesizing mentions into high-level summaries from observation level using both structured and unstructured parts of EMR. Moreover, these systems didn't support cross-document inference and apply domain knowledge driven rules to link similar observations across documents over time. This step is important to form longitudinal patient summaries that include details of diagnoses, treatments, and temporal relationships. Finally, there is still a need to identify and model key events that are stated in clinical document indicating temporally representation of diseases progression phases including diagnosis, treatment, and follow-up.

2.2 DeepPhe NLP Pipeline

Deep Phenotype platform was developed to extract cancer phenotypes directly from medical records with the goal of supporting retrospective cancer translational research and precision medicine. The DeepPhe was motivated by previous software developments including eMERGE, THYME, TIES and Apache cTAKES to extract terms from clinical text, generate high-level summaries of cancer phenotypes [47] and display the deep phenotype information using an interactive visual analytic tool DeepPhe-Viz. DeepPhe platform addressed the need of translational cancer researchers to mine EMR data and explore the information within the

narrative clinical notes. DeepPhe and DeepPhe-Viz are an open-source where the components and more detailed technical information about the DeepPhe are available at https://github.com/DeepPhe/DeepPhe-Release.

The development of the DeepPhe NLP pipeline was informed using different goals and requirements related to retrospective cancer research and cohort identification process (Table 2)[16]. These requirements were established to guide the subsequent system design and information model including (1) the essential coverage of cancer domain concepts using proper terminology by incorporating FHIR modeling[48] and NCI-Thesaurus ontology[49]. The DeepPhe NLP pipeline extracted number of cancer related terms from EMRs (e.g., cancer type, tumor, laterality, cancer stage, and biomarkers)[16] to define cancer cohorts and identify relationships between these concepts (e.g., is-a hierarchy relationship between concepting) (R1-R3). Another requirement that DeepPhe pipeline applied is the inclusion of both the structured and unstructured portions of EMR data as an input source (R6). This requirement supported the investigation of depth and broad insights about cancer populations and answer complex research questions that could be found in the unstructured free clinical text.

Table 2. DeepPhe cancer information model requirement[16]

	Requirement	Description
R1	Appropriate terminology	Use accepted terminologies and vocabularies whenever possible
R2	Cancer-specific content	Provide expressivity necessary to develop appropriately detailed descriptions of cancer treatment and progression
R3	Available tooling	Align with existing APIs, schemata, validators, etc.
R4	Community-driven modeling	Use community contributions and critiques to improve models
R5	Compatibility with existing NLP infrastructure	Facilitate interaction with existing NLP tools and type systems.
R6	Combinations of structured and unstructured data	Support the combination of structured data represented in EMRs with unstructured details extracted from clinical texts.
R7	Multi-level modeling	Support summarization of data across multiple levels of abstraction, ranging from instances/mentions to documents, episodes (collections of records indicating a distinct phase in disease progression such as diagnosis or treatment), and high-level summaries of cancers and tumors.
R8	Provenance	Preserve and expose linkages between abstracted models and source data

Figure 2 illustrates the components of the DeepPhe pipeline according to Friedman and Elhadad framework[20]. The domain model and knowledge of the DeepPhe NLP system were established after interviews with translational cancer researchers, while existing ontology including NCI-Thesaurus [16] was utilized for the linguistic knowledge. The training corpus was obtained from the UPMC breast cancer dataset. The DeepPhe NLP extended the cTAKES [6], a widely used clinical NLP pipeline for processing unstructured documents to extract different mentions from unstructured text and store them in a graph database. cTAKES was built upon the Unstructured Information Management Architecture (UIMA)[50] framework (R5) (Figure 3-A). cTAKES pipeline consists of components that combine rule-based and machine learning techniques to extract information from the clinical narrative. The components of the cTAKES pipeline are executed in sequence to incrementally add to the extraction process. The components

of cTAKES are sentence boundary detector, tokenizer, normalizer, part-of-speech (POS) tagger, shallow parser and named entity recognition annotator (Figure 3-B).

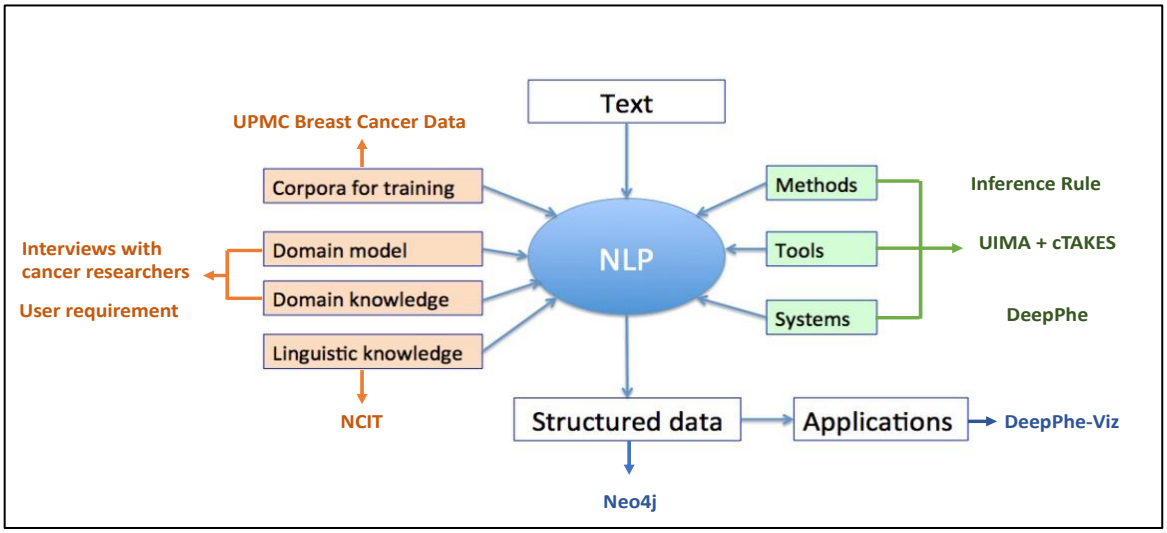


Figure 2. The Components of clinical DeepPhe NLP system according to Friedman and Elhadad framework[20].*[Text in the boxes] represents Friedman and Elhadad framework[20].. [Text without boxes] represent the components of clinical DeepPhe NLP system.*

Cancer EMR data consists of multiple documents with large numbers of individual facts and many details with concepts that are mentioned many times across many documents throughout cancer treatment progress. Due to that, the interpretation and conveying the key points in the cancer clinical notes is a difficult process. Therefore, higher-level summaries that hide details and enable consideration of patients in terms of key details (patient longitudinal summary) are highly necessitated. DeepPhe supported modeling at different level of granularity ranging from individual mentions to summaries from multiple cancer clinical documents to form longitude patients' summaries (R7). These summaries were generated using inference rules developed by domain experts to regulate the process of taking individual observations from the mention level and

transform them into high level summaries at the phenotype level (Figure 4). These rules were developed using the Semantic Web Rule Language[16, 51], and the Drools system[16,52].

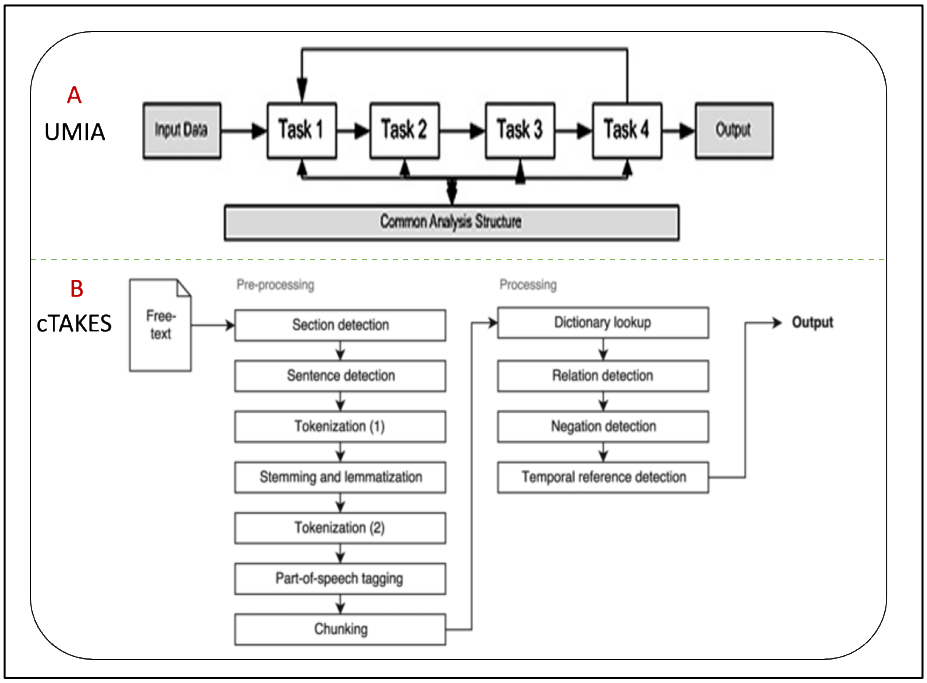


Figure 3. Description of the UIMA framework and the cTAKES pipeline. *(A) UIMA framework start from the input task and then a series of subsequent tasks where the output of one task is the input to the next task. Then the output of each task stored in a common analysis structure. Finally the output of the overall tasks is at the end of the tasks chain [53] (B): Components of cTAKES pipeline[54]including section and sentence detection, tokenizer, POS and chunking process.*

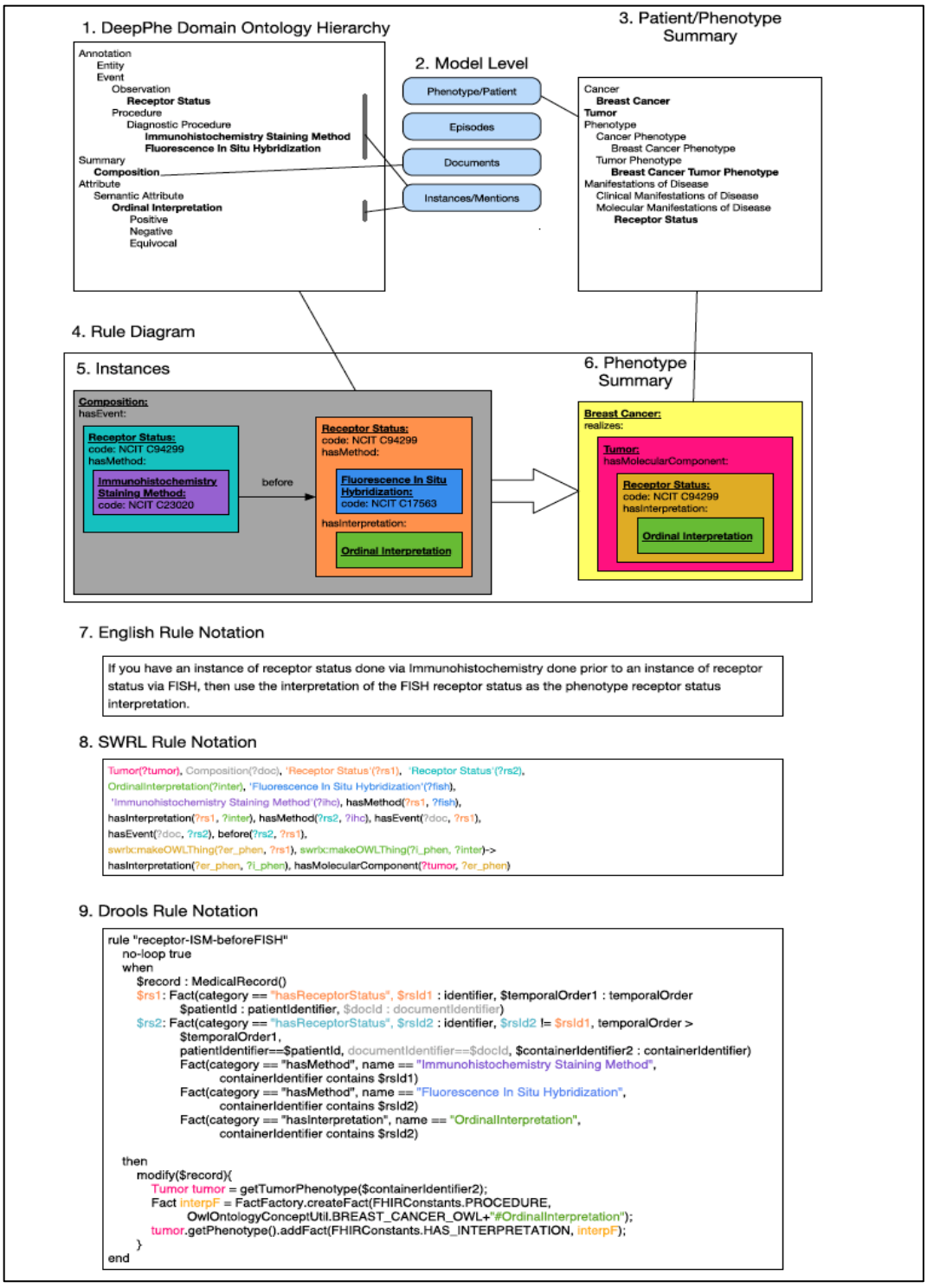


Figure 4. Example of the inference rules in the DeepPhe pipeline. *Starting with the domain ontology and the model level to extract and summarize mentions into patients' summaries. Using inference rules to determine the phenotype summary for multiple instances of the phenotype* [16].

Figure 5 shows DeepPhe different levels of extraction and summarization of breast, ovarian, and malignant melanoma cancer patients' EMR data. During the first level, the DeepPhe NLP pipeline extracted individual mentions from EMR documents (e.g., Lab, procedure, and symptom). This level represents the basic building block for a higher-level summarization. In the second level, DeepPhe pipeline aggregated and summarized the extracted mentions and their relationships found in each document. In addition to that, DeepPhe utilized high-level cross-document coreference[55] by grouping different mentions of the same entity throughout the clinical documents[56].

DeepPhe also modeled the extractions of temporal events by classifying documents in the same phase of disease progression using supervised machine learning algorithm into predefined episodes of care. These episodes are pre-diagnostic, diagnostic, medical decision making, treatment, follow-up and unknown episodes.

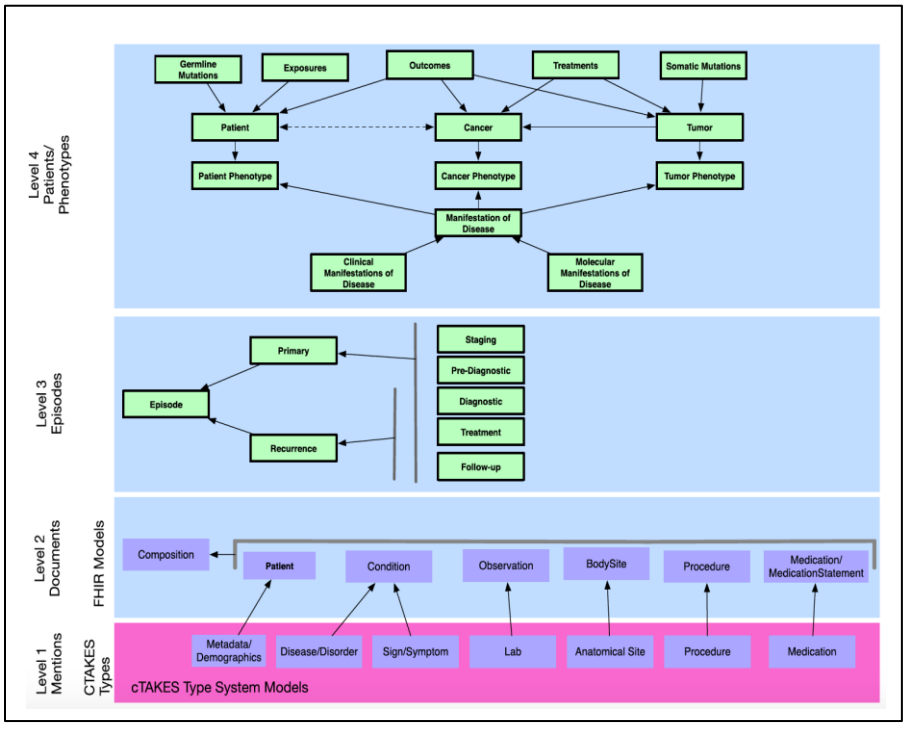


Figure 5 .DeepPhe levels of extraction and summarization[16]

DeepPhe NLP differs from other NLP systems that were developed to extract various cancer data from EMR using NLP[57], including cancer diagnosis[58], tumor description (size, grade, stage)[59], cancer procedure[60], assessment of cancer care[61], and cancer treatment in three important ways: first, DeepPhe' s information model uses different levels of abstractions and temporal relations including inference rules and cross-documents coreference to support a longitudinal representations of patients' cancer journeys [7, 8, 11]. Second, DeepPhe used machine learning to classify clinical documents within distinct disease progression phase into predefined episodes of care to model cancer care patterns. Third, other systems focused on extracting specific aspect or phenotype about cancer patients from cancer clinical document. For example MedTAP[59] system extracting tumor characteristic from pathology report. While DeepPhe extracts many types of data from cancer patients' EMR including diagnosis, treatment, tumor characteristics (laterality, stages, biomarkers) and patient demographics using different types of clinical notes (pathology, radiology, progress notes, discharge summary, and clinical notes) and link the extracted data to the original clinical documents' sources.

The output of the DeepPhe pipeline is stored into a neo4j graph database (Figure 6). Graph databases represent entities as nodes and relationships between entities as links. Both nodes and links can have attributes. Relationships can be directed or undirected. Each node in the graph could have multiple attributes. Relationships between nodes are represented by a directed edge, type, start node and end node. Neo4j is an open-source NoSQL native graph database that utilizes a declarative query language like SQL, called cypher query language. Neo4J offers different drivers for various programming languages such Java, JavaScript, and Python (R3).

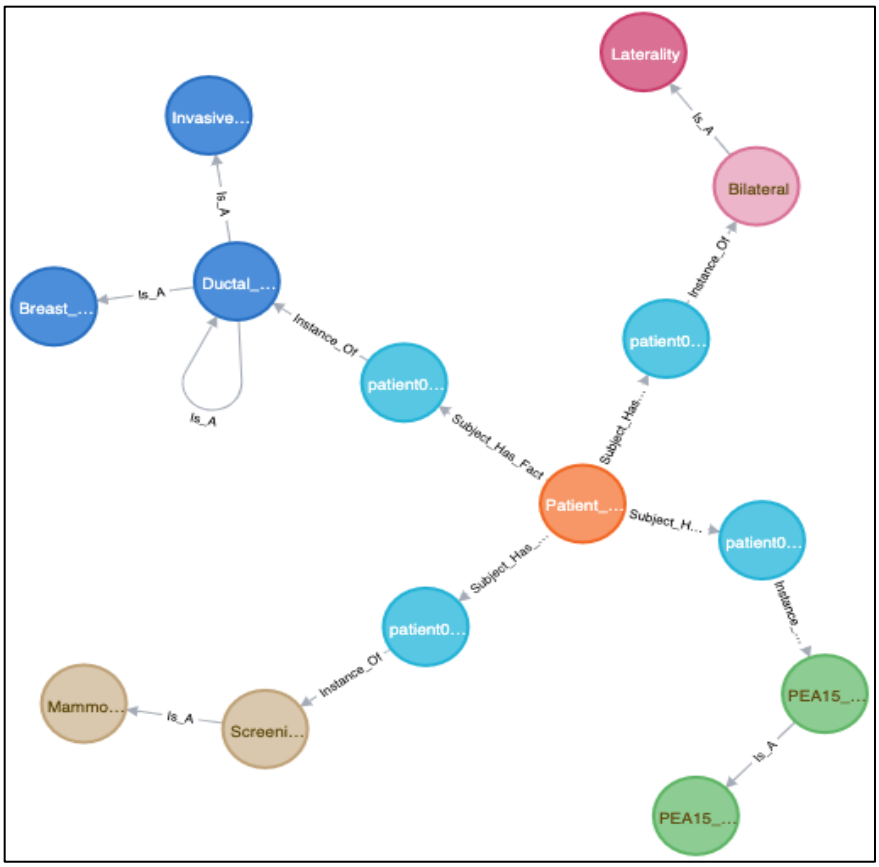


Figure 6. Portion of the clinical terms and relationship stored in neo4j.

Figure 7 shows the different types of relationships between nodes, some of these relationships are defined according to how each node relate to other nodes. Examples of different relationships between nodes including Subject*_Has_Fact*, *Subjcet_Has_Note*, *Cancer_Has_tumor* relationships. The *IS-A* and *Instance_Of* relationships were identified by including of the NCI metathesaurus[49] to represent underlying cancer and biomedical concepts. Figure 7 illustrates an example of the different relationships in neo4j where patient node and fact node relationship (*Subject_Has_Fact*) was inferred from the documents using the different level of summarization. The relationships between the mention, concept, and class node including the *Instance_Of* and *Is_A* hierarchy were modeled using NCI metathesaurus

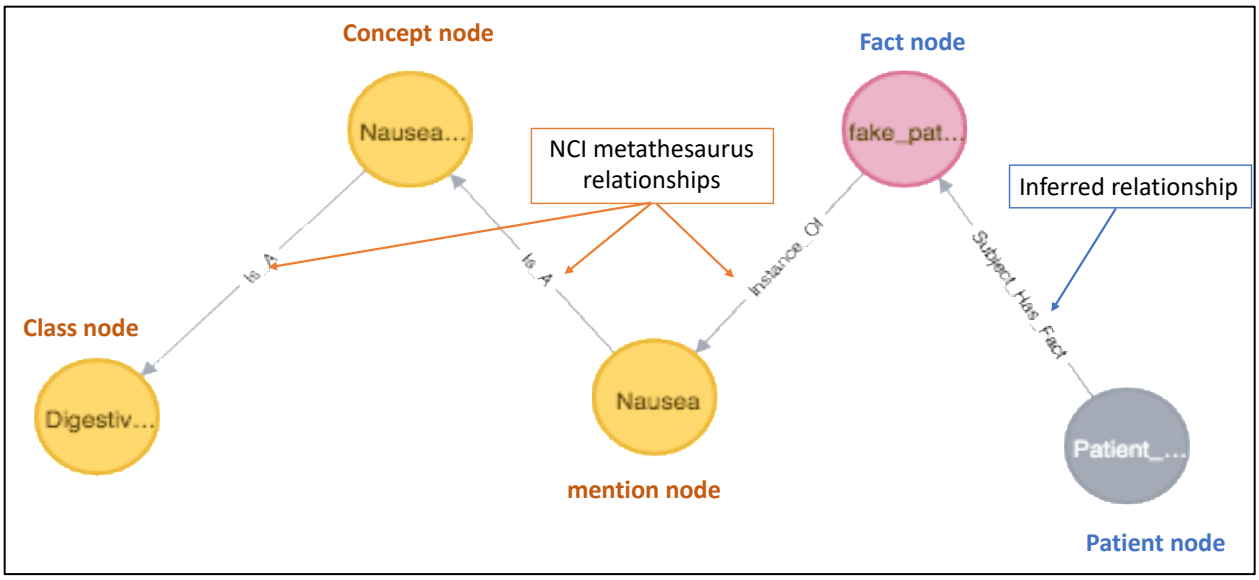


Figure 7. Example of the relationships between nodes stored in neo4j*. Inferred relationship and NCI metathesaurus relationship.*

2.3 DeepPhe-Viz

To visualize the raw output of the DeepPhe NLP pipeline, visual analytics approaches were used to unlock the value of high-dimensional data, support clinical decision-making[62-64] and assist the investigative analysis of different data types. DeepPhe-Viz, a web-based visualization tool, has been developed to visualize the results of DeepPhe pipeline and to assist the visualization process of longitude cancer patient using visual analytics (Figure 8).

DeepPhe-Viz retrieved the output of DeepPhe NLP that was stored into neo4j. Several modules were embedded at the back end of the DeepPhe-Viz system and were called using different routes. The JAVA API was implemented to query the extracted data from the graph database and a middle layer between the JAVA API and the front end of the DeepPhe-Viz system was implemented using neo4j API and node.js modules to prepare and preprocess the retrieved data.

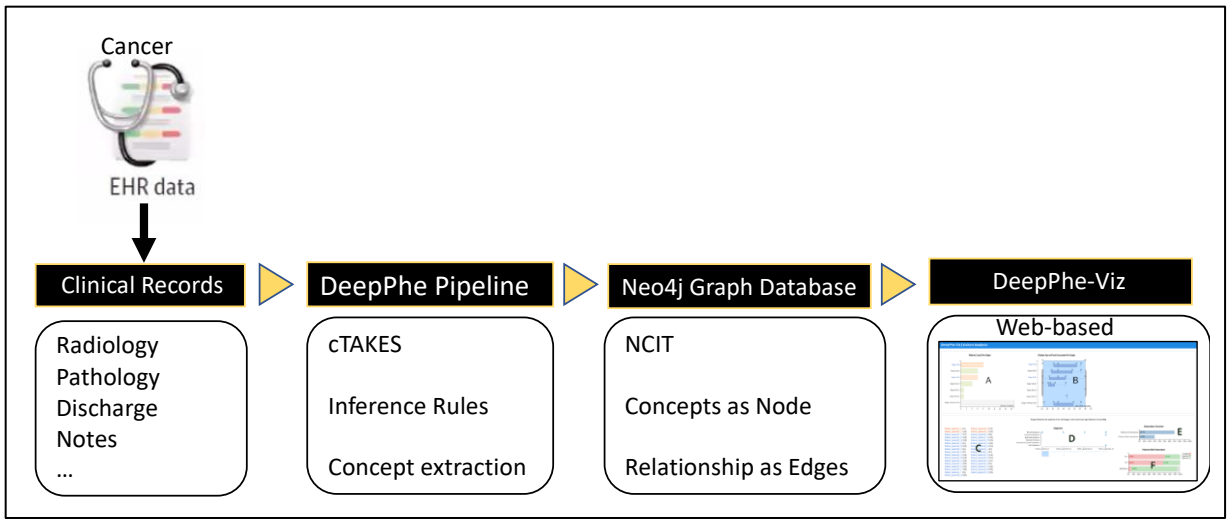


Figure 8. The process for extracting clinical terms and visualizing the results using visual analytics.

DeepPhe-Viz was developed after the completion of several qualitative inquiry studies with cancer domain experts to gain insights about the challenges in cancer retrospective research[18, 47]. These interviews were conducted with cancer researchers in the early stages of developing the DeepPhe artifacts and were used to identify different user stories, goals and information needs that informed design aspects including DeepPhe information model, the DeepPhe pipeline, and the DeepPhe-Viz visual analytics features. These qualitative inquiries informed the potential roles of the participants who have wide domain expertise in breast cancer, ovarian data, and melanoma from the University of Pittsburgh and the Magee Women's Research Institute and the Hillman cancer institute. The interviews focus on the user goals and challenges in retrospective cancer research, information needs and demonstrations[16, 47]. The results of the interview notes and the recordings were used to inform the development of a set of user requirements (Table 3) and possible design suggestions for DeepPhe-Viz. The current version of this tool provided the basic functionalities to address key users' requirements; however other requirements highlighted in Table 3 still are not addressed[18].

Table 3. All the required functionalities and the associated user stories for developing DeepPhe-Viz tool*. The highlighted rows represent the requirements that still need to be addressed* [18]

Requirement	User stories	Barriers
Overview	Explore distribution of available data	
Temporal	Note time points or intervals corresponding to changes in clinical practice or other contextual factors that might influence results; Examine trends in lab values; Review sequences of care episodes; Ask temporal queries. Determine length of overall and progression-free survival	NLP improvement Extended DeepPhe-Viz interfaces
Text	Refer to original clinical text	
Provenance	Identify original text used to extract/infer observations	
Multrecords	Use multiple types of records to bolster interpretations	
QA	Verify accuracy and relevance of extracted data; Check for consistency within documents set, and between documents and unstructured or registry data	
Stage	Identify cancer stage	
Genomics	Determine biomarker status and method used to make determination	
Treatments	Identify treatments provided, number of cycles, doses, treatments offered but not taken, reasons for discontinuing treatments; Relate start and end dates of treatments to outcomes; Understand response to treatment/progression, including failure to respond, recurrence/no benefit, metastasis, developed resistance	Extended DeepPhe-Viz interfaces
Filters	Filter cohorts based on categorical or temporal values; Focus patient exploration based on relevant constraints	
Uncertainty	Communicate both inherent ambiguity in notes and confidence in extracted information	NLP improvement
Search	Conduct searches based on clinical criteria and temporal relationships	Extended DeepPhe-Viz interfaces
Comparison	Compare outcomes across selected cohorts	Extended DeepPhe-Viz interfaces
Criteria	Report criteria used to identify or compare cohorts	Extended DeepPhe-Viz interfaces

DeepPhe-Viz also provides multiple granularities of patient and cohort views and provides a link to the original clinical text[18].

- Cohort level: provides an overview of all the patients in the cohort with different characteristics including:
    - Number of patients in each cancer stage: Figure 9-A demonstrations patients' counts for each cancer stage.
    - Distribution of first Encounter Age for patient in each Stage: Figure 9-B illustrates the usage of box-whisker plots to summarize the distribution of patient age of diagnosis across all cancer stages.
    - List of the target Patients: Figure 9-C shows a list of patients that is grouped by their age of first encounter and serves as link to transfer to the patient level view to explore more details for the individual patient profile.
    - Diagnosis chart: Figure 9-D demonstrates a summary of all target patients that are grouped under diagnosis.
    - Overview of the biomarkers: Figure 9-E shows the distribution of the percentage of target patients with and without biomarkers
    - Overview of the patients with biomarkers: Figure 9-F presents the percentage of patients who have positive, negative, and unknown biomarkers using stacked bar chart.

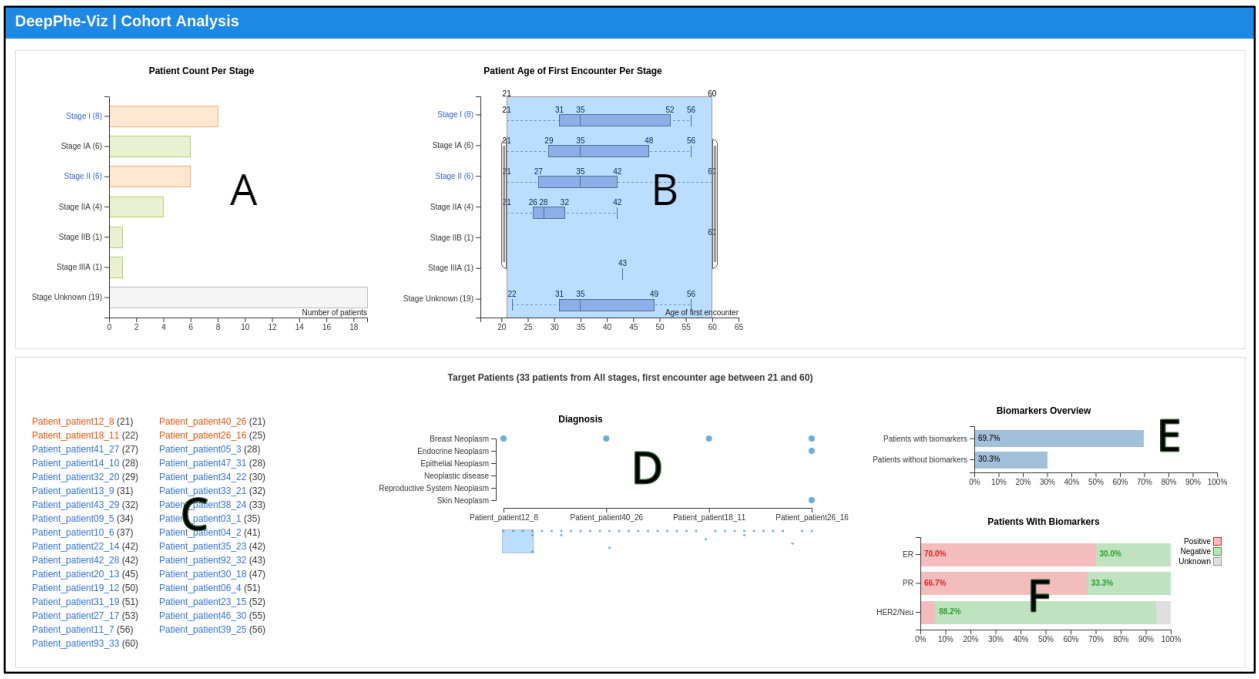


Figure 9. ***Cohort level in the DeepPhe-Viz[18]***. *(A): Number of patients in each cancer stage. (B): Distribution of first Encounter Age for patient in each Stage. (C): List of the target Patients. (D): Diagnosis chart. (E): Overview of the biomarkers. (F): Overview of the patients with biomarkers.*

- Patient level: provides a way to support fine-grained explanation of an individual patient's profile using multiple views of a selected target patient form the cohort level patient list
  - Patient personal information: Figure 10-A view the different demographic characteristic about the patient
  - Cancer and Tumor summary: Figure 10-B shows the list view stacks of different tumor/cancer characteristic as a default
  - Timeline View of reports and episodes: Figure 10-C illustrates a temporal view of different clinical reports and episode of care of the selected patient.
  - Clinical note: Figure 10-D shows the original text that was used to extract/ the information or infer all the observations via domain rules in the cohort level and patient level.

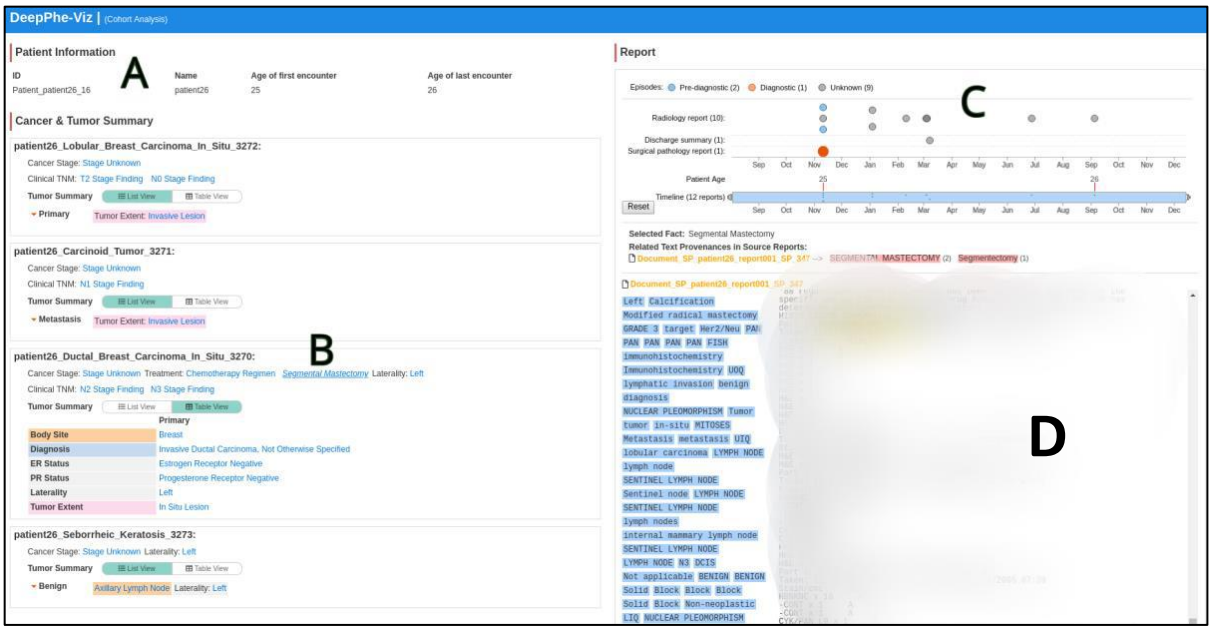


Figure 10. Patient level in the DeepPhe-Viz[18] *(A): Patient personal information. (B): Cancer and Tumor summary. (C): Timeline View of reports and episodes. (D): Clinical note (Text obscured to protect patient privacy)*

2.4 Visual Analytics

Visual analytics is described as "the science of analytical reasoning facilitated by interactive visual interfaces"[65]. Visual Analytics tools are developed to identify opportunities to support exploration, insight finding, and comprehensible visualization of information in different domains. Recent work has shown the effectiveness of using visual analytics approaches for outcome analysis, comparing cohorts, and events sequences[66]. Moreover, visual analytics tools enhanced precision and personalized medicine through providing prediction to make discoveries and uncovered hidden truths from large collections of data and information[67, 68]. The visual analytics utilized interactive visual interfaces to facilitate analytical reasoning and knowledge gain from the data[66]. Visual analytics techniques incorporated different domains from data management, data mining, cognitive science and human interaction to assist the reasoning and

interpretation of convoluted data[69]. Figure 11 shows an abstract overview of the visual analytics process. The component of this process started with data preparation and transformation to get the required representation of the data. Then mapped the data into the visualization interface or build an automatic model from the original data using data mining techniques. Using the interactions between the generated models, visualization interface and the user feedback supported the knowledge discovery and uncovered new aspects about the original data[70] .

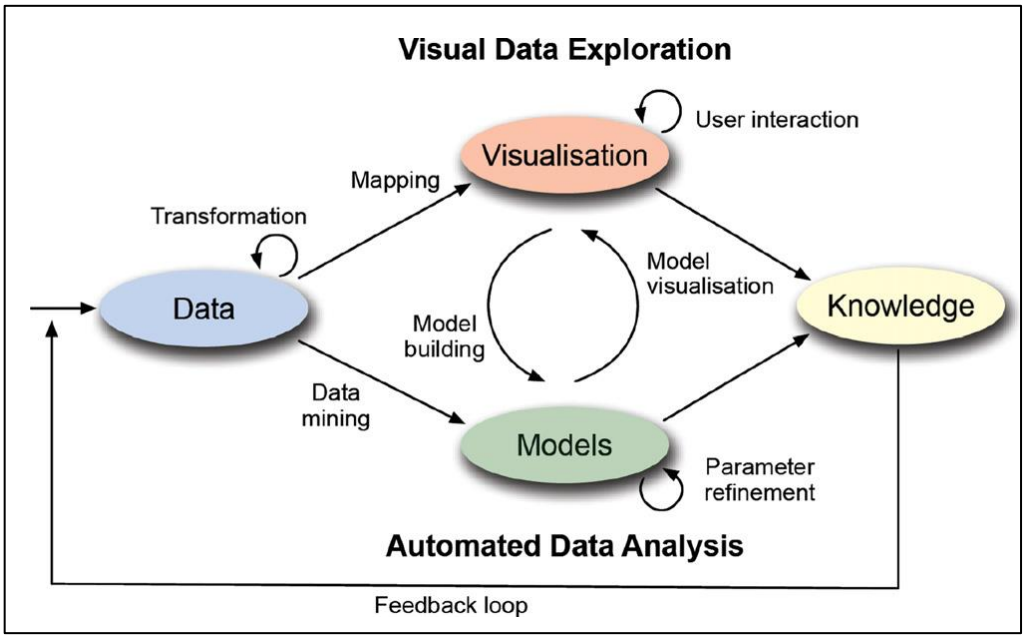


Figure 11. The visual analytics process*. The oval shapes represent the different phases and the arrows represent the transitions between these phases in the process [70].*

Several visual analytics systems were built to assist investigative analysis with different data types and visualization approaches (Table 4). For example, The prostate cancer cohort analysis[71] focused on identify and view multiple patients' histories at once and automated the process of calculating statistical measurement to identify different correlation scores for prostate cancer patients[72]. This tool used the data of prostate cancer that are stored into a database. The

HARVEST[73] NLP system extracted observation from multiple care settings including inpatient, ambulatory, and emergency department encounters and combined the output with different visual analytics approaches[16, 17, 73]. However, dealing with cancer data are more complex because they contain multiple types and instances of documents such as radiology, pathology, lab test and clinical notes and the aggregation process and summarization of different observations into high-level representation is not found in the HARVEST system (Figure 12).

The Cohort Comparison -CoCo- tool was built to find the similarities and differences between the sequence of events in two datasets[74]. The CoCo used well-structured data and didn't support the extraction of data from free text.

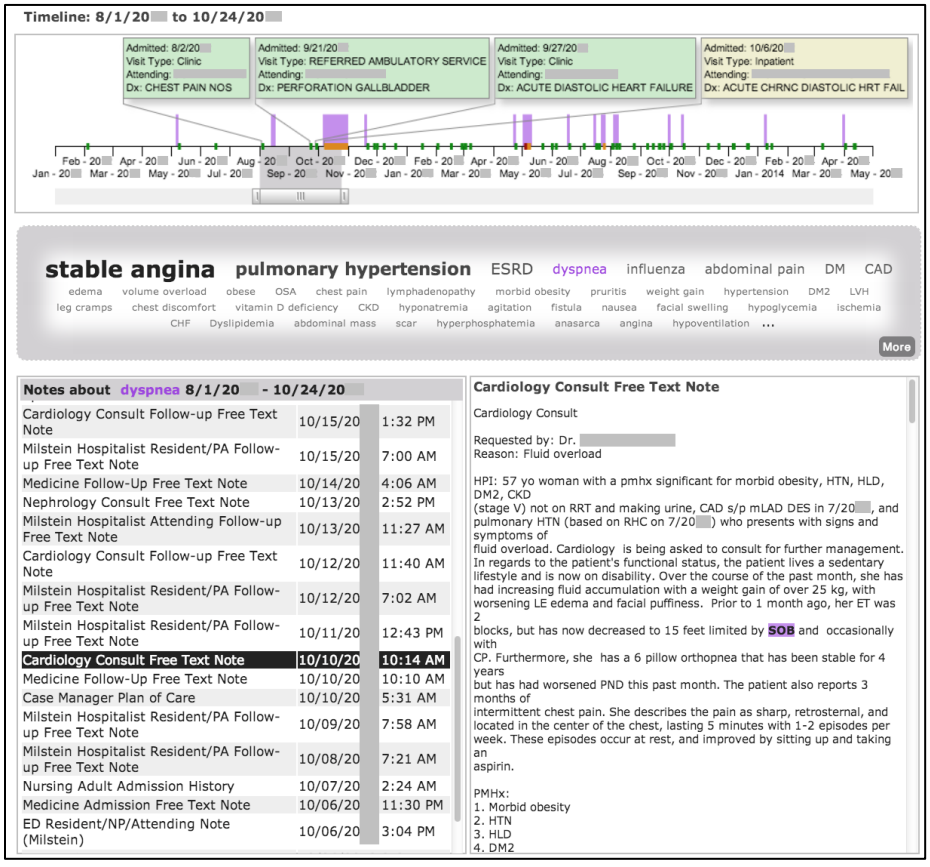


Figure 12. HAVREST system of deidentified patient*. Includes a timeline, document problem, notes panel list and the actual clinical note[73].*

Approaches such as EventFlow[75] (Figure 13-1) and Outflow[76] (Figure 13-2) visual analytics tools supported temporal data visualization and various sequences of events within the cohort. The limitation of the EventFlow and Outflow tools was the assumption that the data contained well-defined time points and was only available in abstracted data, which is not the case when working with cancer data in the clinical notes. EventThread[77] presented prior work on summarizing event sequences into similar sequences clusters called threads (Figure 13-3). EventThread used a tensor-based approach and grouped similar threads to visualize the evolution of patterns over time[77]. The main limitation with the above systems was the usage of a fixed-width time interval to identify the threads. This could be an issue in cancer medical data where the length of cancer care and the intervals between different events is not consistent between different patients due to the difference in the pace of the disease progression.

Table 4. Visual analytical tools that use various types of EMR data, visualization approaches and evaluations

Tool	Task	Data type			Visualization approach	Evaluation				
		Structured	Unstructured	Temporal/Event		Case study	User study	Usability study	Focus group	Observational study
CAVA[78]	The CAVA tool uses interactive visual analytics to facilitate the process of preforming retrospective cohort studies	✓			Interactive Histogram-based Interactive pie charts Flow diagram Tabular view	✓		✓		
COCO[74]	Cohort Comparison is an interactive visual analytics tools that combines human driven analysis and automated statistics to identify the similarities and differences between the sequence of events in two Cohorts	✓		✓	Interactive scale bar Color coded triangles for events sequences	✓				
Composer[79]	Visual analysis tool for identify a cohort of patients for an orthopedic surgeon to detect changes in patient physical functions after performing different procedures	✓		✓	Filter sidebar Histograms distributions Outcome Score Comparison graph Dynamic Score Scales / Normalization	✓				

Table 4 continued

COQUITO[80]	Cohort Query that uses Iterative overview to identify temporal conditions on the dataset. This interactive tool uses a novel visual query to empower domain experts in exploring and generating different queries based on the temporal aspect of the data.	✓	✓	Bar Charts Treemap Query view panel	✓
DecisionFlow[81]	A visual analytics tool that aggregates high-dimensional sequences of temporal event and provide multi-view visualization and statistical analytics. DecisionFlow provide a dynamic temporal sequence query using a scalable representation.	✓	✓	Temporal Flow diagram Time edge to highlight episodes overlapping. Interactive statistical graphs including color-coded bubble chart	✓
EventFlow[75]	A visualization tool to query and display interval and points events. EventFlow provides different controls to explore the records that contain interval events.	✓	✓	For interval events vertical-colored bars Color-coded triangles with timeline Subsequence Query Module Overlap Query Module Graphical query Module	✓

Table 4 continued

EventThread[77]	An integrated and scalable visual analysis system that combine large scale and high-dimensional event sequence data with different summarization, and visualization approach to support different event sequences such as clustering, pattern discovery and stage analysis.	✓		✓	Line map metaphor Event Flow diagram List view of entities, threads and events Tree map of each thread	✓		
EventThread2[82]	a visual analysis designed that build upon EventThread to incorporates progression analysis of event sequence data. This system designed to overcome the time scale issue in EventThread	✓		✓	Cluster view Sequence view Query view Event view Stage transition view Thread view Entity List view	✓	✓	
HARVEST[73]	A visualization tool that aggregates and summarize data from different multiple care setting using *NLP*	✓	✓	✓	Timeline panel Problem Cloud panel Note display panel	✓		
LifeFlow[83]	A visualization tool that summarizes all the possible spacing events within different sequences to support users' exploration, identify patterns and anomalies in the data	✓		✓	Using the Icicle Tree Visualization Color-coded event bar Zooming, Tooltip, Sort, Align Overlay distribution of gap between events	✓	✓	

Table 4 continued

Lifelines2[84]	An interactive visualization tool that searches and explore medical records for event sequences using a color-coded representation of different events to represent an overview of events and episodes of multiple patient record and provide a link to all the detailed in the recodes.	✓	✓	Color coded triangles to represent the events Timeline Align, Rank, Filter display	✓	✓		
Prostate Cancer Cohort Analysis[72]	An interactive tool to visualize multiple prostate cancer patients' histories with respect to temporal events in the data to identify early determinants for predication of prostate cancer	✓	✓	Color coded lines for different phases Static attributes for filtering Dynamic querying			✓	✓
OutFlow[76]	An interactive visualization tool that summarizes and aggregate different temporal events and view all the paths within the cohort of patients. The Outflow graph captures all event paths that help to recognize how different progression paths may lead to better/ worse results	✓	✓	Graph-based visual presentation Filtering, Brushing, Tooltips, Panning & Zooming Symptom Selection	✓			

Table 4 continued

Timespan[85]	An exploratory visualization tool that developed to identify temporal events in acute stroke patients records to gain a better understanding about the treatment process.	✓	✓	Stacked bar graph (Bertin-Style matrix)[86] Timeline Interactive Legend		✓	✓
VISCARETRAILS[87]	A visualization system that summarizes multiple sequences of timed event relative a given root event in pancreatic cancer patients. This tool provides a way to explore patients EMR and understand different patterns of care	✓	✓	Timed Word Tree Distribution using boxplot	✓		
STORYFLOW[88]	A visualization system that generates a layout to visualize the dynamic relationships between various entities in a text	✓		Horizontal timeline Multiple lines bundled together during a time period	✓	✓	

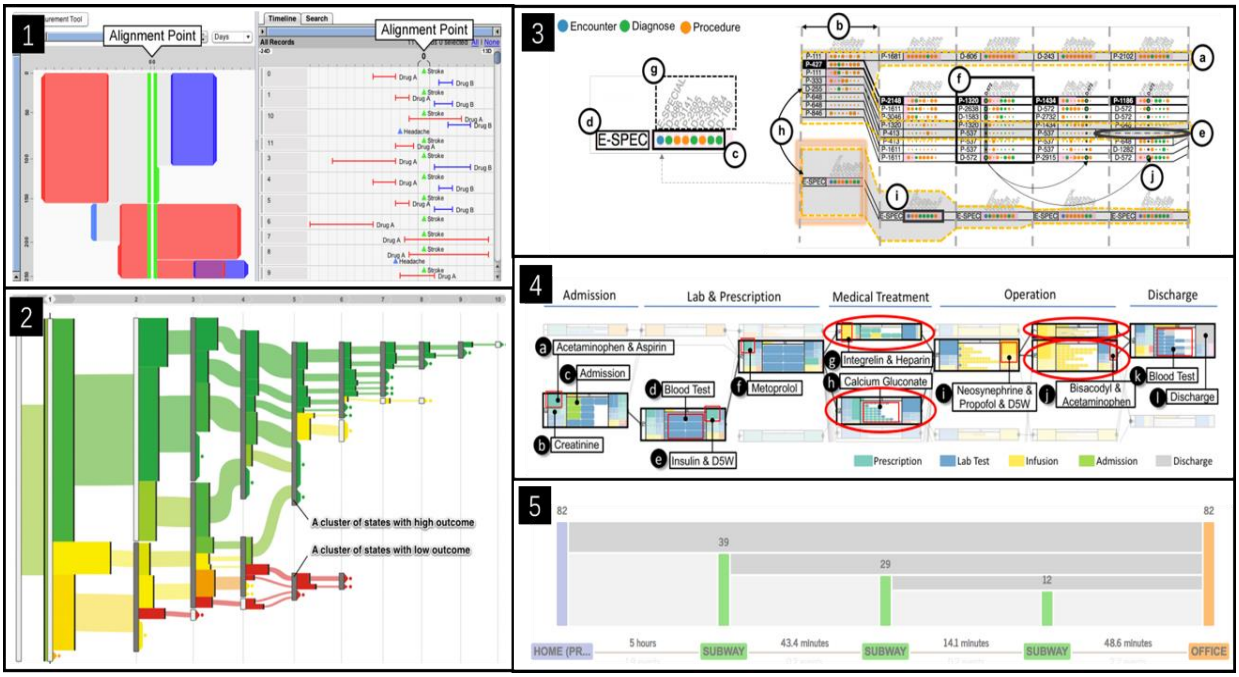


Figure 13. Examples of visual events summarization systems*. (1) EventFlow[75] aggregation of temporal event data from a cohort of patients and visualization of multiple clinical pathways using tree-like overview and detailed a timeline display. (2) OutFlow[76] explores point and interval events from a cohort of patients using color-coded edges that map to patient outcome to identify temporal patterns. (3) EventThread[77] uses tensor analysis to visualize multiples threads as segmented linear stripes, following a line map metaphor. (4) EventThread-2[82] provide a higher-level summary of temporal sequences patterns using a node-link visual design. (5) MAQUI [89] represents frequent patterns by applying hierarchy-based visualization and implements a timeline to reveal the temporal information.*

Other visual systems utilized flow-based approach such as CarePre[90] (Figure 14-A) and Sankey diagram-style timelines[91] to visualize various events and transitions over time. Sankey layouts can uncover useful knowledge about the specific events in the cohort and the trajectories of each event over time. Frequence's visualization[92] adopted a modified version of Sankey diagram[91] layout to represent different events patterns in the cohort of patient with lung disease (Figure 14-B). Huang et al. created an interactive visualization that showed disease-diseases association for chronic kidney disease overtime using a Sankey layout (Figure 14-C). Unlike the previous works which focused mainly on dealing with structure data with clear semantics,

DeepPhe tackled the challenges associated with the raw output of the NLP and used different approaches such as machine learning, and inference rules to build the temporal model for cancer cohort and provide a link to the original clinical documents.

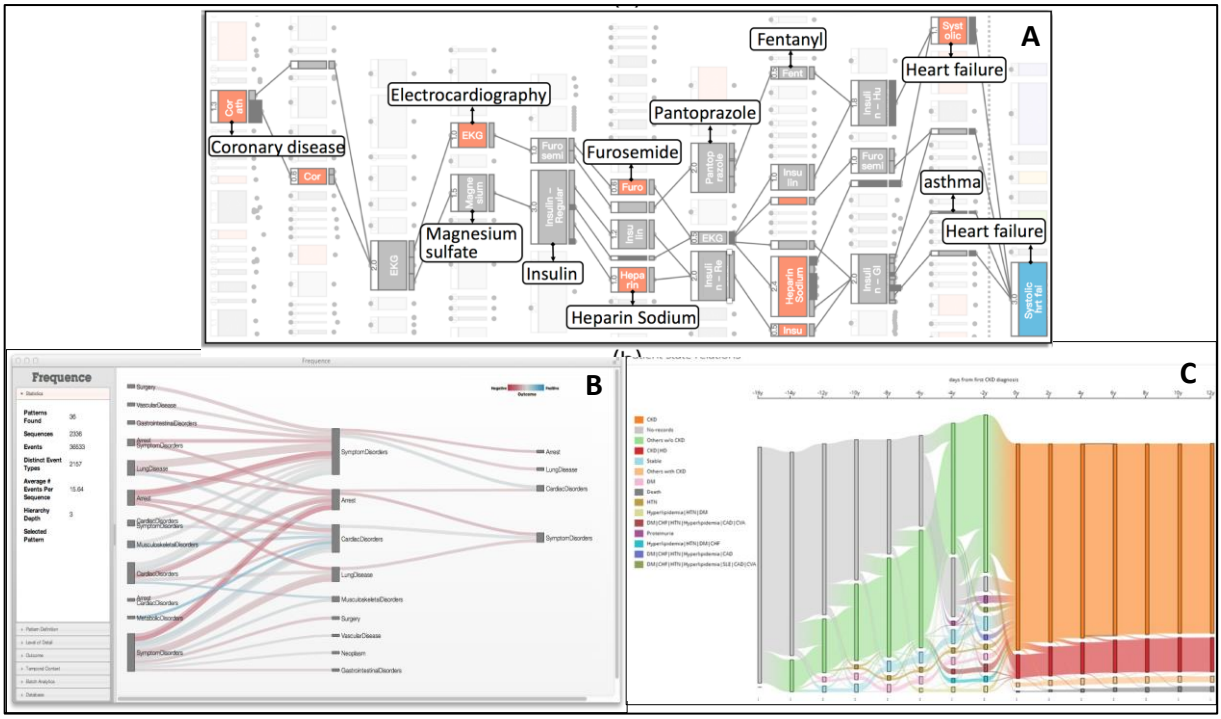


Figure 14. Examples of Sankey-based visualization systems*. A. CarePre[90] displays patients medical record using a Flow-based visualization. B. Overview of different events in the cohort of patient with lung diseases using Frequence[92]. C. Overview of related comorbidities for patient's cohort with chronic kidney disease through time[33].*

3.0 Hypotheses and Specific Aims

With all the valuable information that already was presented by the current version of DeepPhe-Viz to meet user requirements, there are still many extracted vital requirements (i.e., treatment, finding, and comorbidity) that already stored in neo4j and need to be viewed and presented to the user. Providing this new information using different views and interactive approaches helps answering more complex questions about the cohort (e.g., which patients in the current cohort with a specific treatment). Table 5 summarizes the limitations of the current DeepPhe-Viz system, the proposed solutions and the goals and purposes of these solutions

Table 5. Limitation in DeepPhe-Viz, the proposed solutions, plan, and the purpose.

Limitation in DeepPhe-Viz	Plan	Purpose
Utilize the content of the clinical text	• Aggregate different mentions of terms into high-level groups for each patient in the cohort using the hierarchy relationship between the terms. **For example: chemotherapy mentions are aggregated into drug group** • Aggregate high-level groups for all the patients in the cohort using the documents timestamped. • Using an interactive heatmap visualization to view the change in the frequencies of the high-level groups through time of diagnosis for all the patients in the cohort.	Help cancer researchers filter patients based on specific mentions of clinical terms at a specific timestamped and review trend in the terms' frequencies through time of diagnosis. **For example** **Identify patients with chemotherapy?** **Identify patient with comorbidity?**
Include the temporal transitions of the episodes of care for all the patients in the cohort	• Identify the sequences of episodes for each patient in the cohort • Resolve the unknown episodes • Remove duplication in the episode's sequence • Divided the episode sequence into an overlap pairs of episodes • Combine all the patients in the cohort who have the same pair of episode transition at the same order • Using an interactive Sankey visualization to view the change in the episodes transitions for all the patients in the cohort.	Help cancer researchers filter patients based on specific sequences of episodes and review different patterns of transitions in the cohort **For example:** **Identify number of patients who have treatment to medical episode?**

In this project, I utilized the output of the NLP pipeline to visualize a high-level representation of terms and events mentioned in the clinical documents using interactive and comprehensive visual analytics approaches for all the patients in the cohort. Viewing these terms

is required to answer key research questions about the patients' characteristics that were mentioned in the clinical documents. In addition, DeepPhe-Viz currently viewed different temporal events only at the patient level. Thus, aggregating of different temporal events in the cohort view and visualizing the changes in the sequences of events through time are still unmet. These additional steps are central to understand the different patterns and trend of cancer care and treatment among patients.

1) Aggregate the extracted terms from neo4j into different semantic groups to provide a user with a high-level summary of the contents of the clinical documents including symptoms, treatments, complications, and responses for all the patients in the cohort and to identify changes in the semantic group frequencies through time, which provide investigators with insights to compare patients' treatments progresses.
2) Display the broad distributions of terms through time in the semantic groups using visual interactive artifacts for all the patients in the cohort to enable the identification of patients with specific mentions at specific time point.
3) Views the overall transition of episodes of care for all the patients in the cohort and identify the patients with a specific episode sequence to determine different patterns of care.

3.1 Hypotheses

- I hypothesize that including the above functionalities into DeepPhe-Viz help cancer researchers identify different aspects of cancer cohort in timely and accurate manner.
- I hypothesize that the additional functionalities and views are useful and helpful to cancer investigators to identify a cohort of cancer patient with different data characteristics.

3.2 Specific Aims

This dissertation addressed the following specific aims

Aim 1: **visualize the change in the distribution of high-level summary of clinical documents contents through time of diagnosis using visual analytical artifact.**

To achieve this goal, first the extracted terms that are stored in neo4j were aggregated into high-level groups (semantic groups) to address the summarization challenges of the raw NLP output. Then I used an interactive heatmap to view the change in the prevalence of each sematic group through time. This provides the user with an overview of the different terms in the clinical documents and helps in answering different and challenging users' requirements and filters patients with specific term mentions. In addition to that, the change in the semantic groups prevalence through the time of diagnosis for all the patients in the cohort is also included to provide a visual representation of changes in the clinical term distributions.

Aim 2: visualize different patterns of episodes of care sequences for all the patients in the cohort view using visual analytical artifact.

To achieve this goal, first I Identified the sequence of episodes in the clinical documents for each patient on the cohort, then I inferred a new episode for the unclassified episodes. All the episodes then were aggregated into in the cohort view. I adopted Sankey diagram layout to represent the sequence of temporal events using the episodes of care within the cohort. This approach is commonly used to provide an overview of the transitions between different events and provides a complete bird's eye view of the different sequences within the cohort and at the same time allows the user to interact and drill down to more details about the relation between certain events in the cohort.

Aim 3: **Conduct user studies with translational cancer researchers to assess the usability of the system**

To achieve this goal, I evaluated the design using a range of predefined tasks, calculated time completion and correction for each task, and collected user feedback using usability scale questionnaire and open-ended survey.

3.3 Significance and Innovation

Despite the success of applying informatics approaches to extract and visualize clinical data from EMRs, most prior research has focused on structured data. Some studies have explored methods to extract, integrate, and utilize heterogeneous clinical data including structured and narrative data from different types of notes in EMR. However, these studies lacked modeling individual mentions into high level summaries, classifying documents into episodes of care, including the temporality of clinical mentions, and linking extracted mentions and summaries to the original clinical documents. In this study, we extended the development of DeepPhe-Viz platform to address the need of cancer researchers to visualize the rich content of clinical notes in an abstract and meaningful representation using visual analytics

We applied interactive visualization features that complement the DeepPhe NLP and used visual analytics features in cancer domain to support cohort identification and hypothesis generation. To my knowledge this is the first proposed work to (1) aggregate individual mentions into higher-level meaningful semantic groups to facilitate the challenge of summarizing the extensive content of enormous amount of clinical text associated with cancer patients' histories, and (2) use an interactive time-based heatmap to view high level semantic groups of different

mentions in the clinical documents is necessary to easily answer key research questions about the cancer patients that are mentioned in the clinical notes for retrospective research. In addition of that, (3) visualize the transition of episodes of care using temporal Sankey diagram to facilitate the understanding of the complex longitudinal patient histories which is an important challenge for understanding cancer treatment and outcomes. These visual analytics provide generalizable, concise, and comprehensive views of a longitudinal summarization of cancer patients trajectories using structure, and unstructured EMR data.

4.0 Methods

The DeepPhe system started by extracting and storing the terms from the cancer clinical notes into a graph database and then provided preliminary cohort and patient views of the extracted cancer data using web-based visual analytics tool DeepPhe-Viz. These views included selected subgroups of extracted terms and relationships based on the requirements and goals informed by the users. However, many terms and temporal events that were extracted and identified using DeepPhe NLP stored in neo4j were not visualized in the DeepPhe-Viz interface. Including additional views to represent a sequence of large scale and diverse timestamped terms and events using the document date and episode type provided the first step to identify patients with specific characteristics using time-stamped semantic groups or episode sequence. For example, identifying patients with specific comorbidities in the beginning of cancer diagnosis timeline. Many terms in clinical notes are vital to answer key questions about the cancer patients' diagnosis and treatment. In addition, certain episode sequences can provide the investigators with additional insights about the patients. For example, patients who have episode sequence of (medical-decision to treatment to medical-decision) episodes could indicate that there were more discussion and updates about the treatment plans for those patients due to various reasons such as treatment responses and side-effects.

4.1 Construct the Clinical Terms Visualization into DeepPhe-Viz

Figure 15 illustrates the architecture of DeepPhe-viz. I extended the JAVA API and neo4j API modules in the backend of DeepPhe-Viz to query and retrieve all the terms stored in neo4j with their corresponding semantic group using the is-a relationship. Then the retrieved data was

handled by a data processing module to preprocess, prepare and send a JSON format of the data to the front-end of the DeepPhe-viz. The front-end of the DeepPhe-Viz (client side) is implemented using JavaScript[93] and D3 library[94] to generate the interface using interactive diagrams and filtering techniques in both the cohort and patient views.

I started by aggregating the NLP extracted terms into high-level groups called semantic groups for each patient in the cohort. This step is important to summarize and facilitate the visualization of the change in the clinical terms' distributions in the clinical documents throughout the cancer diagnosis timeline. Then I normalized patients' document dates to combine patients with same semantic group at the same time point in the diagnosis timeline. Finally, I adopted an interactive visualization to demonstrate the change in semantic groups frequencies though time for all the patients in the cohort.

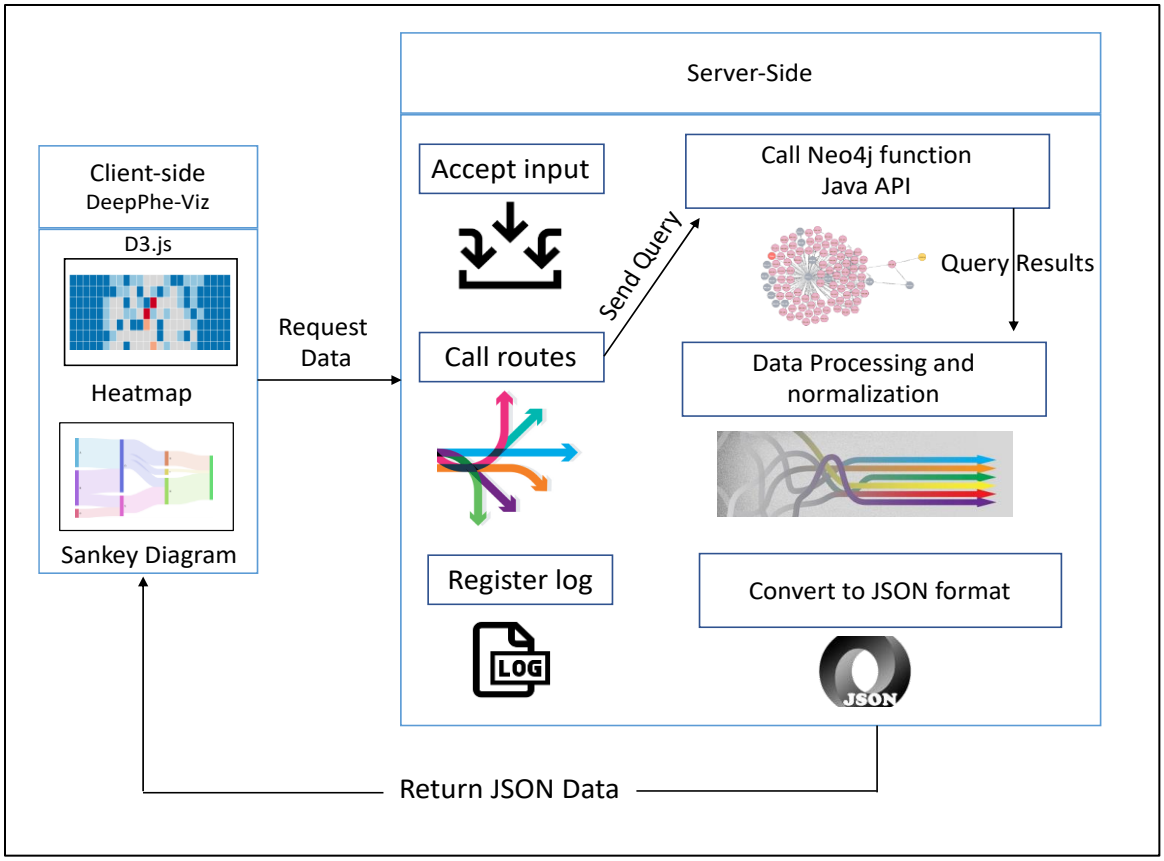


Figure 15. DeepPhe-viz server-client architecture*. The components of the server side including different routes calls, the data processing and normalization of the call results, and return results in JSON format to the client side that generate the different visualizations into the interface*

4.1.1 Aggregate Stored Terms in Neo4j into High-Level Groups

I extended the Java API plugins and node.js modules to retrieve all the terms and events in neo4j database. Terms are stored in neo4j as nodes and nodes related to one another using different relations. Each term stored in neo4j is associated with one or more documents, and each document is related to a patient in the cohort. I aggregated terms into higher-level concepts using the hierarchy relationships of the stored terms with the class node (root node). Figure 16 represents an example of how “Docetaxel” term (drug name) is stored in neo4j and the path to

the class node to determine that the semantic group (high-level representation) of this term which is drug.

Six semantic classes were identified including Finding, Drug, Procedure, Disorder, Lab, and Others. This step reformed the substantial amount of the NLP row output into a high-level representation to first provide a summarized and concise view and second to display the distribution of broad groups of terms in the clinical documents in the cohort view.

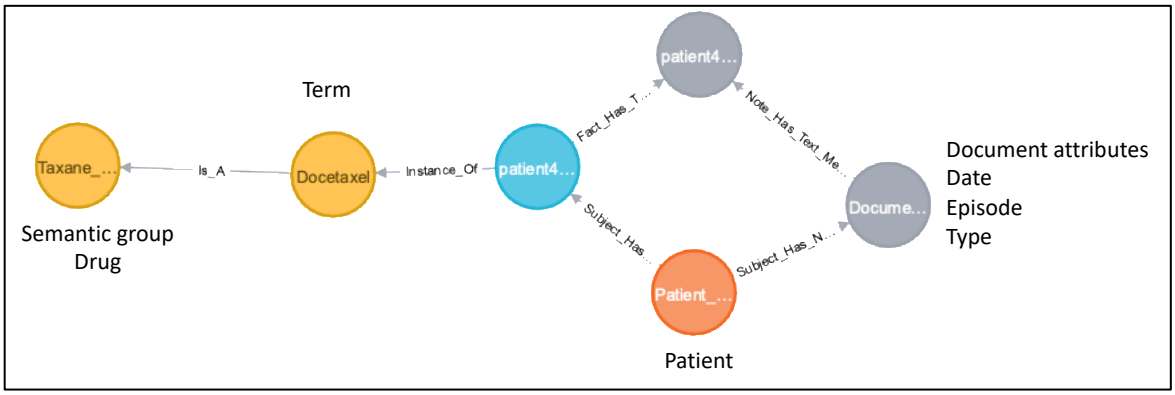


Figure 16. Example of terms stored in neo4j

To start the aggregation process, I used dictionary data structure for each patient in the cohort to store the results of neo4j cypher queries into a key-value pairs. The dictionary indexed the patient-Id, frequencies of terms for each semantic group and the document date where these terms were mentioned for each patient in the cohort (Figure 17).

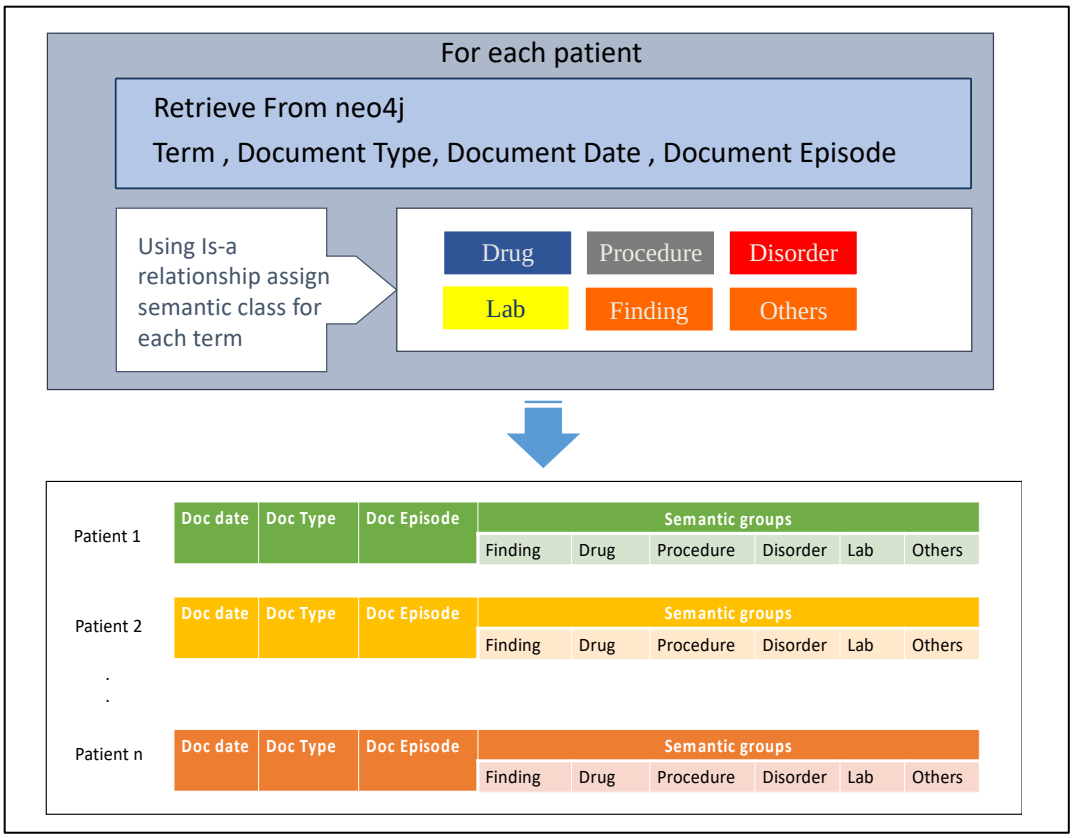


Figure 17. Aggregation of clinical terms into high-level concepts*. The terms from neo4j queries are stored into the dictionary and combined according to the corresponding semantic.*

4.1.2 Align and Normalize the Timeline for the Aggerated Semantic Groups

To demonstrate the change in the semantic groups frequencies through time for all the patients in the cohort. I combined patients with same semantic groups and the same timestamped using the document date where the term was mentioned. Figure 18-A shows an example of three patients in the cohort with different frequencies of semantic groups in the same month of the diagnosis timeline, and Figure 18-B shows the aggregation process of the same semantic groups for these patients.

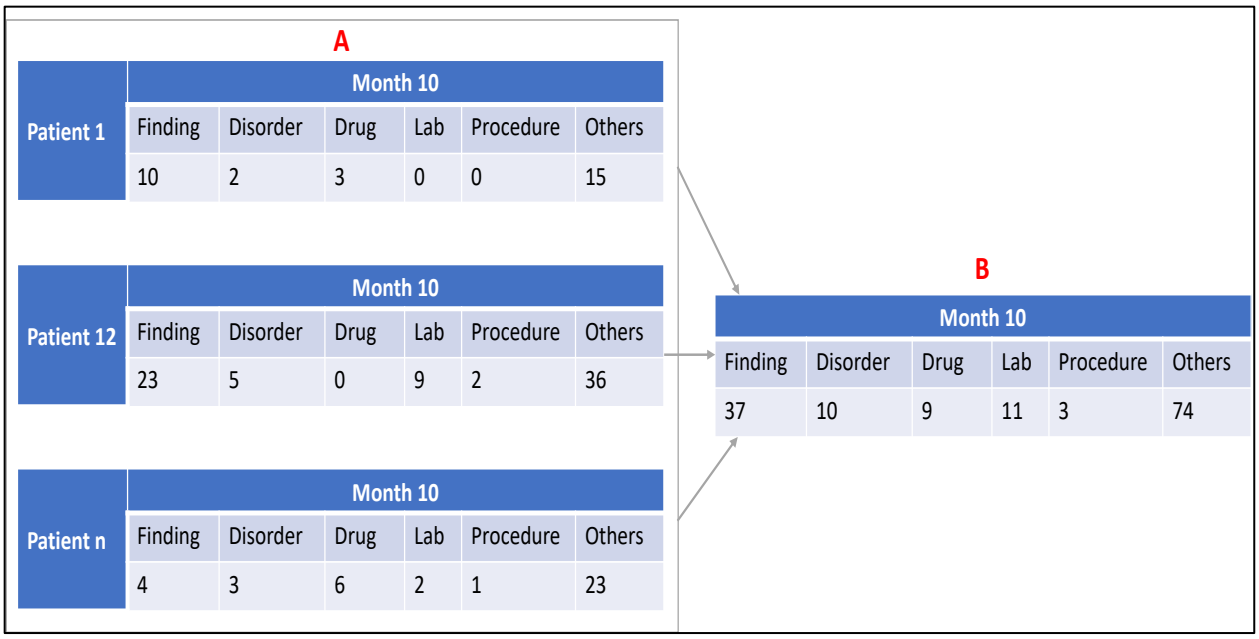


Figure 18. Example of the aggregation process of all the semantic groups frequencies for the patients with the same month of diagnosis.

However, the nature of cancer care where patients could have different distributions of documents through years of treatment with large gaps between these documents added a challenge in viewing the changes in the semantic groups in these documents through time. This issue is even harder when dealing with a cohort of patients with different documents distributions through time. Figure 19 illustrates how documents were scattered through time for 12 patients between 2000-2014. For example, patient 9 had documents from 2000-2001 and nothing until 2004. In addition to that, the timeline for all the patients' documents was not aligned which means that the starting document date for each patient in the cohort was different. This made the aggregation process of semantic groups with similar timepoint inefficient and uninformative.

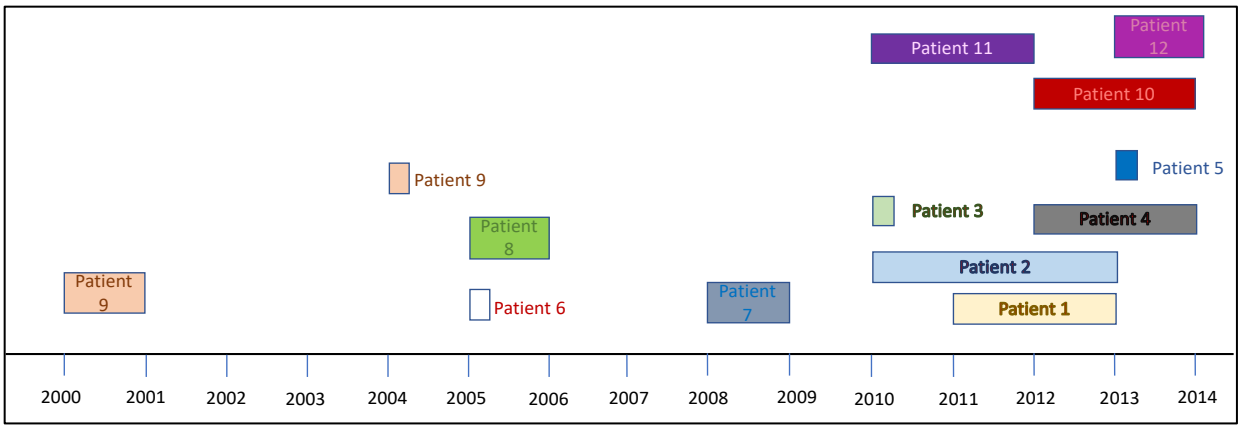


Figure 19. Illustrate the patients' documents timeline in the cohort

To solve these issues, I normalized all the documents dates in term of number of months starting from the first document date for each patient. Figure 20 shows how all the patients in the cohort had the same start point and they all had documents in the first 6 months of patient care. To normalize and align documents dates for all the patients in the cohort, first I sorted all the documents from earliest to newest using the documents date for each patient in the cohort. Then I assigned the document with the earliest date as the starting time point of care for each patient in the cohort. I labeled that document with the first month, and from that document date, I calculated the number of months between the first document and the subsequent documents (Figure 21). After assigning the number of months to each document in the cohort, I combined patients with the same semantic groups mentions with the same month of diagnosis for all the patients in the cohort.

Pseudo code

For each patient in the cohort:

Ascending order of the clinical documents according to the doc_date

Declare a month variable that assigned to each document:

First_doc. month= 1

Sub_doc. month:

Find diff = |date of first_doc – date of sub_doc|

If diff =0 then sub_doc. month= 1

else sub_doc. month= ceil (diff/30)

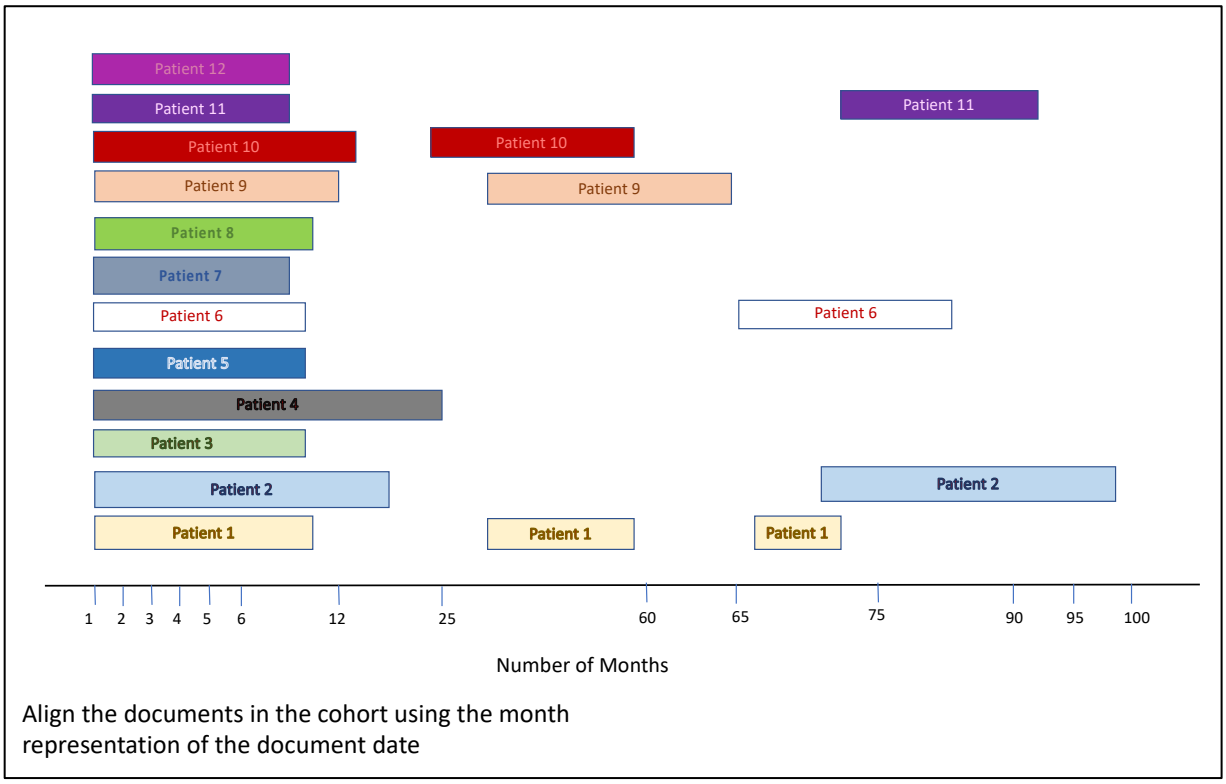


Figure 20. Normalize and align documents dates for all patients in the cohort

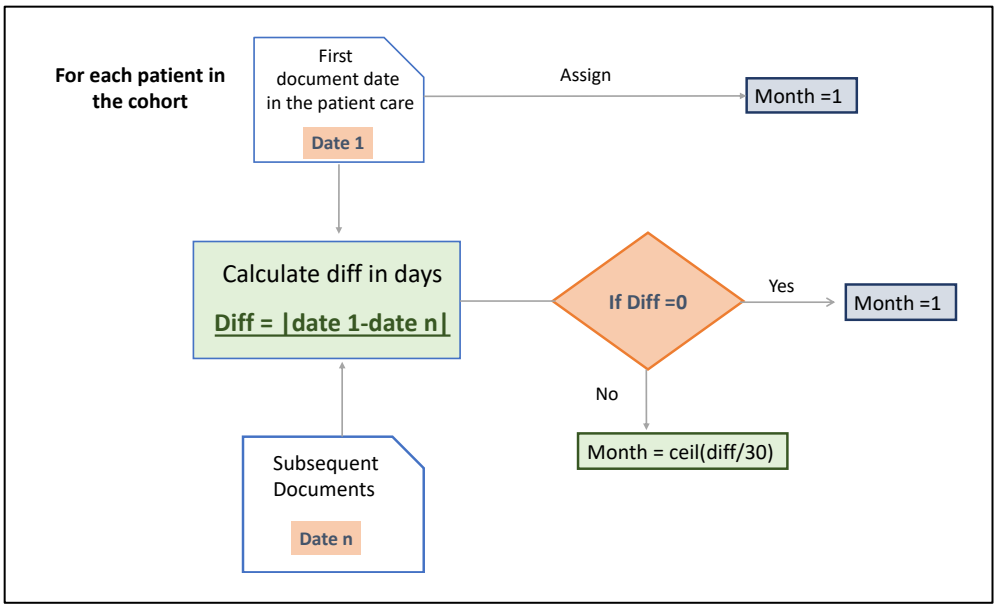


Figure 21. Calculate the month of patient care for each document in the cohort

4.1.3 Semantic Groups Prevalence Visualization

To support the visualization of the contents of the clinical documents for the patients in the cohort and to identify patients with specific terms mentions (i.e., treatment) I implemented an interactive visual artifact that visualize the change of each semantic groups through time of diagnosis. This change was represented by the prevalence of the semantic groups at each timepoint. The prevalence identifies the proportion change in the distribution of the semantic groups through time by considering the number of documents where the terms in the semantic groups were mentioned. Over cancer care period the number of documents is usually high in the early stages and causes high frequencies of terms in these timepoints. So, it is important to normalize the semantic groups distribution using prevalence quantity rather than terms frequencies due to the skewed nature of these clinical documents' distributions.

I utilized a heatmap chart view to represent the change in the clinical text contents using the assigned months for each document and the aggregated semantic groups. Heatmaps utilized a spatial matrix with cells colored after their values. These values showed the change in the prevalence to detect and filter patients with specific mentions at specific time point in their cancer care. Figure 22 shows the heatmap feature that was implemented into DeepPhe-Viz interface. The rows represented the different semantic groups, and the columns were the time periods of the documents date as months. The colored cell represented the different range of prevalence values and the legend in Figure 22 illustrates the various ranges of values for each color. I also implemented a tooltip feature in the heatmap map to provide the user with a summary of the selected cell values including the year, month, semantic group, and the prevalence value.

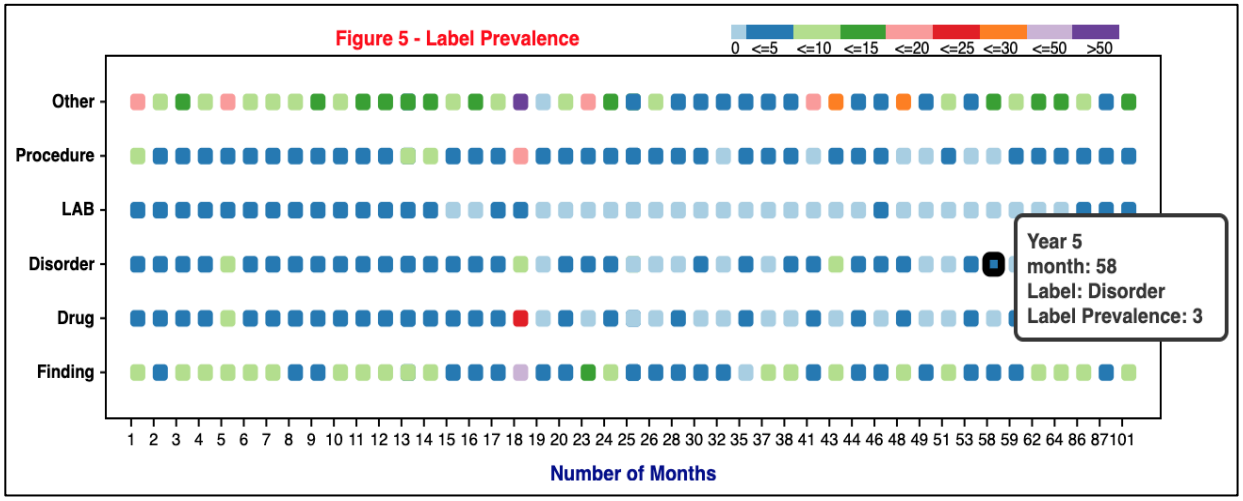


Figure 22. Heatmap visualization of semantic groups prevalence values in each month of patient care. *Color-coded cell to identify change in the values of the semantic group prevalence values, a tooltip feature to provide a brief information about the selected cell.*

A stacked-bar diagram has also been added into DeepPhe-Viz interface under the heatmap chart to illustrate the distribution of document types at each month of diagnosis for all the patients in the cohort. This provides an additional information to describe the various patterns of changes in semantic groups prevalence values through time in the heatmap (Figure 23) and provide a visualization of different types of clinical documents including radiology, pathology, progress note, discharge summary and clinical note and the change in frequencies for each type through the time of diagnosis for all the patient in the cohort.

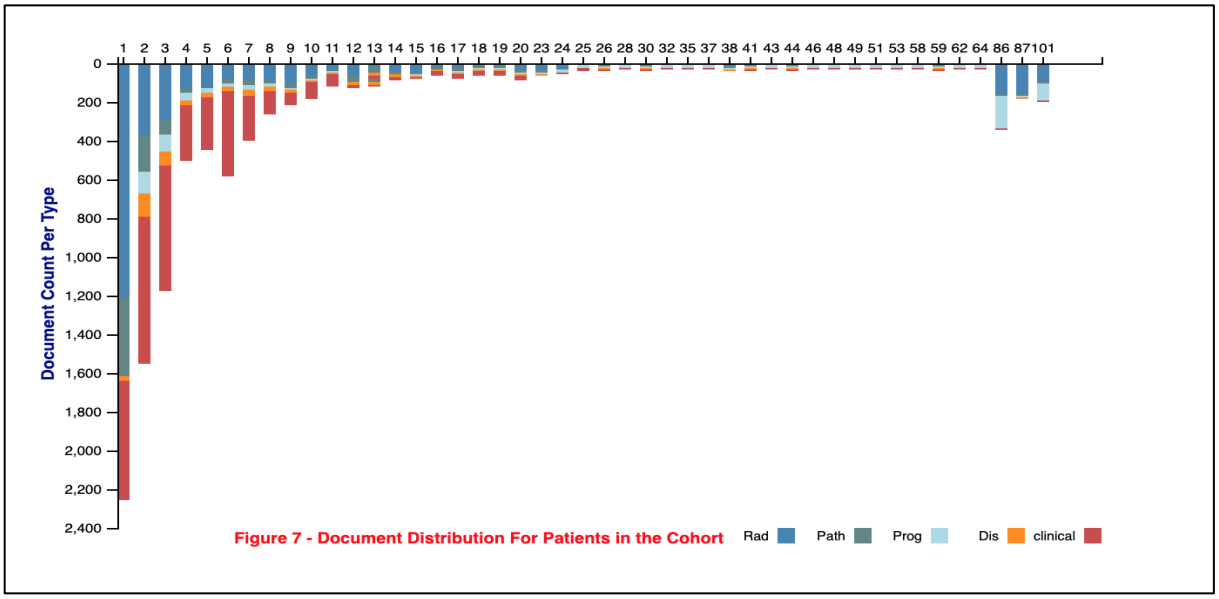


Figure 23.Stacked-bar visualization of different document type distribution for each month of patient care in the cohort.

4.2 Construct the Episode of Care Visualization into DeepPhe-Viz

4.2.1 Temporal Aggregation of Episode of Care in the Cohort View

Each document stored in neo4j was assigned with an episode of care, which represented key event intervals in the cancer care. These episodes were assigned using a supervised machine learning approach, where documents with specific mentions were classified to a predefined episode of care. For example, any document with tumor mention with no malignance diagnosis was classified by a domain expert as pre-diagnostic episode in the training set (Table 6). The machine learning algorithm classified clinical documents into the following episodes pre-diagnostic, diagnostic, medical decision making, treatment, follow-up, and unknown. The issue with the episode classifier was the number of documents that were assigned to unknown episode which didn't provide any information about the type of events in these clinical documents. Figure 24 shows the steps to prepare the episode data to be visualized into DeepPhe-Viz including (1) inferring the unknown episodes when possible and (2) identifying the episode sequences for each patient in the cohort by first removed any duplicated subsequent episodes in the episode sequence and divided the episode sequence into overlap pairs of each two subsequent episodes.

Table 6. The description of predefined episodes of care

Episode	Explanation
Pre diagnostic	Documents with tumor mention with no malignance diagnosis
Diagnostic	Documents with tumor mention with malignance or metastatic diagnosis
Decision making	Documents with possible treatments after malignance is confirmed
Treatment	Documents refer to different treatments
Follow-up	Documents appear after the treatment has finished
Unknown	Documents that are ambiguous and were not classified to any of the above episodes

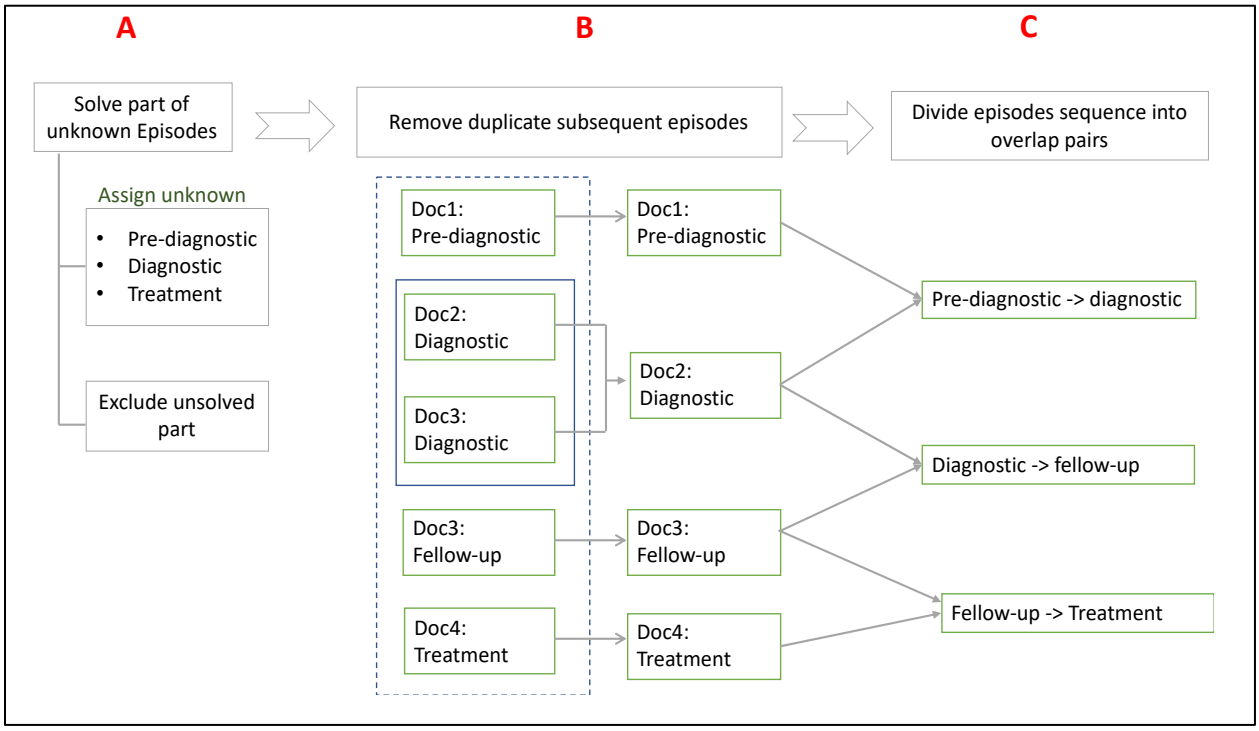


Figure 24. Pre-processing steps of documents with different episodes of care *that were stored in neo4j to identify the episodes sequences for all the patients in the cohort. A. The process started by solve the unknown episode. B. removed repeated episodes. C. spilt the episodes sequences into overlap pairs to show the transitions of subsequent episodes of care*

4.2.2 Infer Unknown Episode

Different types of documents were classified into different episodes of care using features about the documents type, date, and the progress of the tumor diagnosis. These episodes of care represented various temporal events about the treatment progression of cancer diagnosis. The classification of episodes of care informed an understanding about different trends and patterns in the cancer treatment to support the idea of precision medicine by identifying personalized treatment plan and comparing patients with same/different episodes of care.

Pre-diagnostic episode was assigned mostly to radiology reports where the tumor was identified but malignancy was not confirmed. Diagnostic episode was assigned to surgical pathology reports because pathology results confirmed the malignancy of the tumor. Treatment episode was assigned to document types of clinical note, progress note and discharge summary where treatment protocols were performed according to the tumor diagnosis. Follow-up episode was assigned to document types of radiology report and clinical note after treatment phase was completed (Table 7). The unknown episode was assigned to all types of documents where the machine learning algorithm could not assign any of the previous episodes to the document.

Table 7. The different type of clinical documents for each episode of care

	Document Type				
Episode	Radiology	Surgical pathology	Clinical note	Progress note	Discharge summary
Pre-Diagnostic	✓				
Diagnostic		✓			
Medical	✓		✓		✓
Treatment			✓	✓	✓
Follow-up	✓		✓		
Unknown	✓	✓	✓	✓	✓

To resolve the issue of the clinical documents with unknown episodes, I used the episodes assignments from the classifier and the contextual information about the document type to infer the episode for the documents assigned with unknown episode. From Table 7, all surgical pathology reports documents were only assigned to **diagnostic episode** so any document with **surgical pathology report type** and unknown episode is now assigned to diagnostic episode. Furthermore, certain type of documents including **progress notes and discharge summary** were only assigned to **treatment episode**. Thus, I assigned documents with any of these two types and with unknown episode to treatment episode. For the **radiology reports** if the date of the document is **within the first 3 months** of the patient care I assigned that document with a **pre-diagnostic episode** (Figure 25).

4.2.3 *Error* Analysis

An error analysis was performed to identify the potential reasons for classifying different type of documents into unknown episodes using the machine learning episode classifier. A random sampling of different type of documents with unknown episodes were manually reviewed to categorize the challenges in the clinical documents' contents.

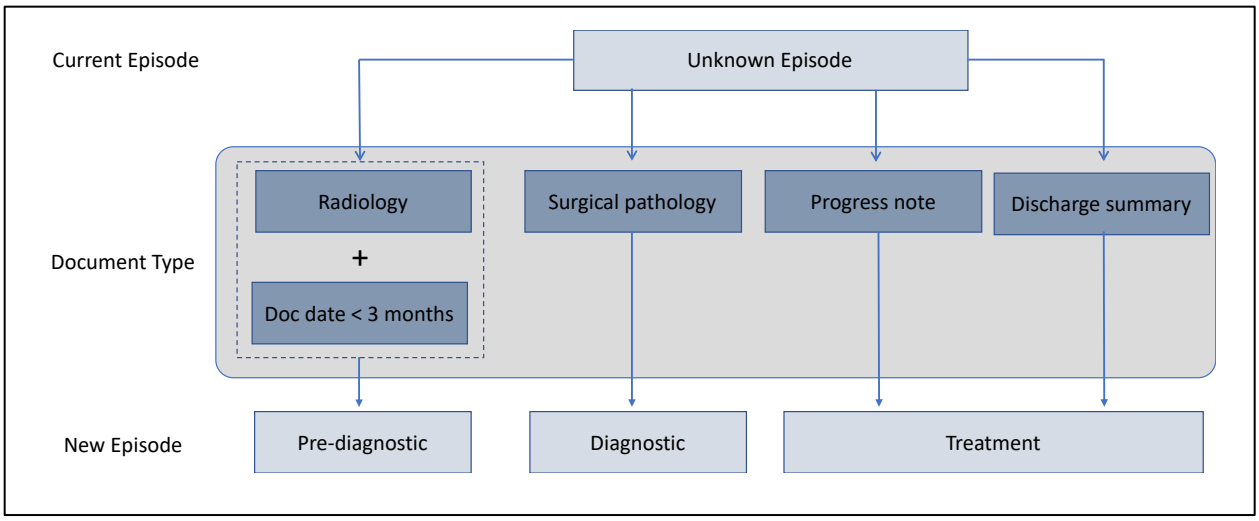


Figure 25. Infer part of the unknown episode for documents with specific types or within a date limit

4.3 Remove Duplicated Episodes in the Sequence

For each patient in the cohort, various sequences of episodes of care were occurred. For example, patient 1 could have the following sequence of episodes:

Pre-diagnostic → diagnostic →medical decision making →treatment →diagnostic →follow-up.

Some patients could have consecutive repeated episodes in the sequence like the following episode sequence:

Pre-diagnostic →diagnostic →medical decision making →treatment →treatment.

From the above example, this patient had two subsequent documents that were assigned to treatment episode. This transition between two documents with the same episode of care didn't provide any additional information about the change in episode sequence. Thus, I merged any subsequent documents assigned to same episode into one document with the same episode of care to reduce number of episodes sequences to provide an informative and compact

visualization (Figure 24-B). So, for the above example the episode sequence for the patient after merging similar episodes is:

Pre-diagnostic →diagnostic →medical decision making →treatment.

4.3.1 Divide Episode Sequences into Pairs

I divided the sequences of episodes for each patient into overlapping pairs to understand the different patterns of episode transitions in the cohort for all the patients and to identify patients with similar sequence of episodes at the same time point at their cancer care. Each pair represented two consecutive episodes using the document date. For example, if a patient has the following episode sequence: *Pre-diagnostic→ diagnostic→ medical decision making→ treatment*. Each two consecutive episodes represented a pair, so the first pair in the sequence was (*pre-diagnostic→ diagnostic*) and the second pair was overlapping with the previous pair (*diagnostic → medical decision making*) and finally (*medical decision making → treatment)* Then I aggregated all the patients who have the same pairs of episodes at the same time to be able to view all the episodes sequences in the cohort level for all the patients. Figure 26 shows an example of three patients with different pairs of episodes sequences, for patient-1 the first pair of episodes was *pre-diagnostic → diagnostic* episode. Both patient -2 and patient-n had the same pair of episodes at the same order (the 1st pair of sequence). Then I combined patients with the same episodes pair into one entry in the data structure with a variable that assigned to the number of patients.

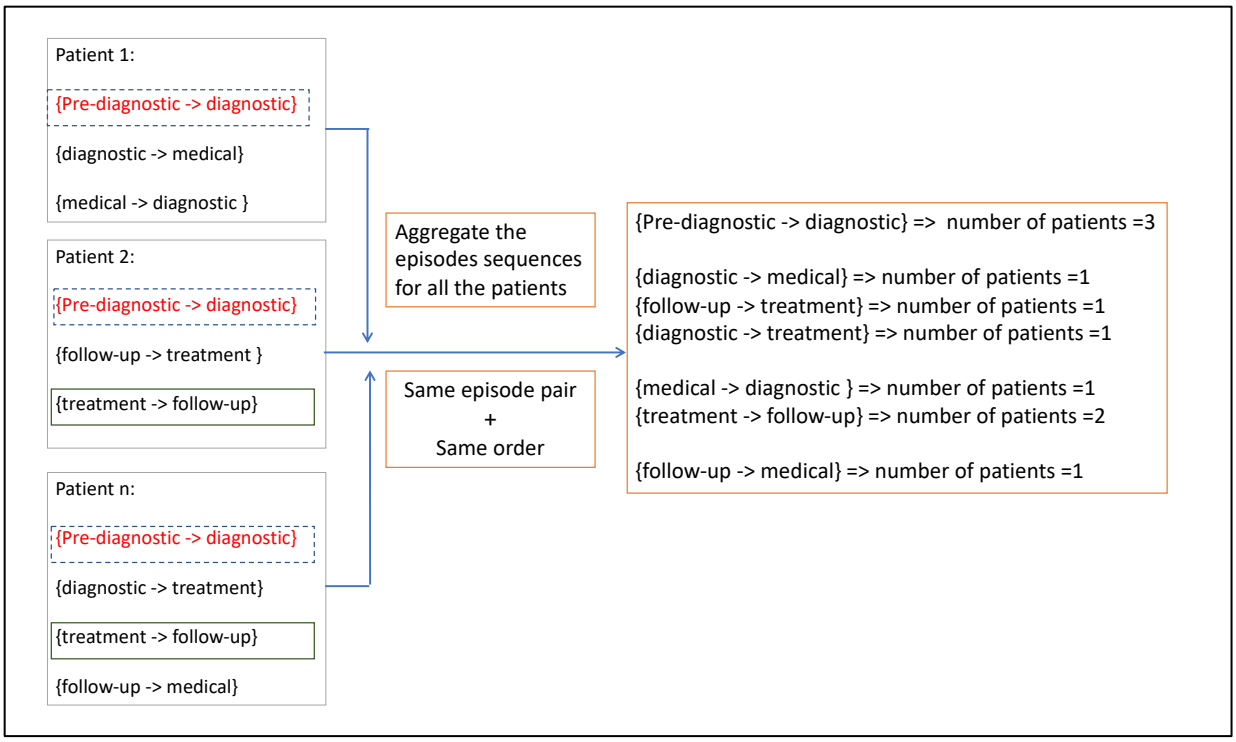


Figure 26. Example about how to aggregate patients with same episodes pair with the same order

4.3.2 Episode Sequence Visualization

I adopted Sankey diagram, which is a type of flow-based layout to obtain an overview of the various episode sequences in the cohort. The purpose of this visualization feature is to provide a visual representation of different pattern of the patients transitions between different episodes of care in the cohort. Historically, Sankey was used in various applications[95] to express different transitions between many to many states or objects.

Figure 27 shows the episode transition chart that was developed based on Sankey visualization in DeepPhe-Viz with multiple level of connection to show the episode transitions from the beginning to the end of cancer care for all the patient in the cohort. It consisted of nodes(rectangles) that represented different episodes of care with a width relative to the number

of patients with that episode and the links between these nodes are represented using lanes or curves with a width relative to the number of patients who have the episodes transition. Each node in each level represented an episode of care with a given color code. The legend in the episode transition chart showed the type of episode for each colored node in the chart. A tooltip feature also was implemented for the nodes and the links in the chart to provide an explanatory brief to the user about the type of episode and the number of patients who had the episode of interest.

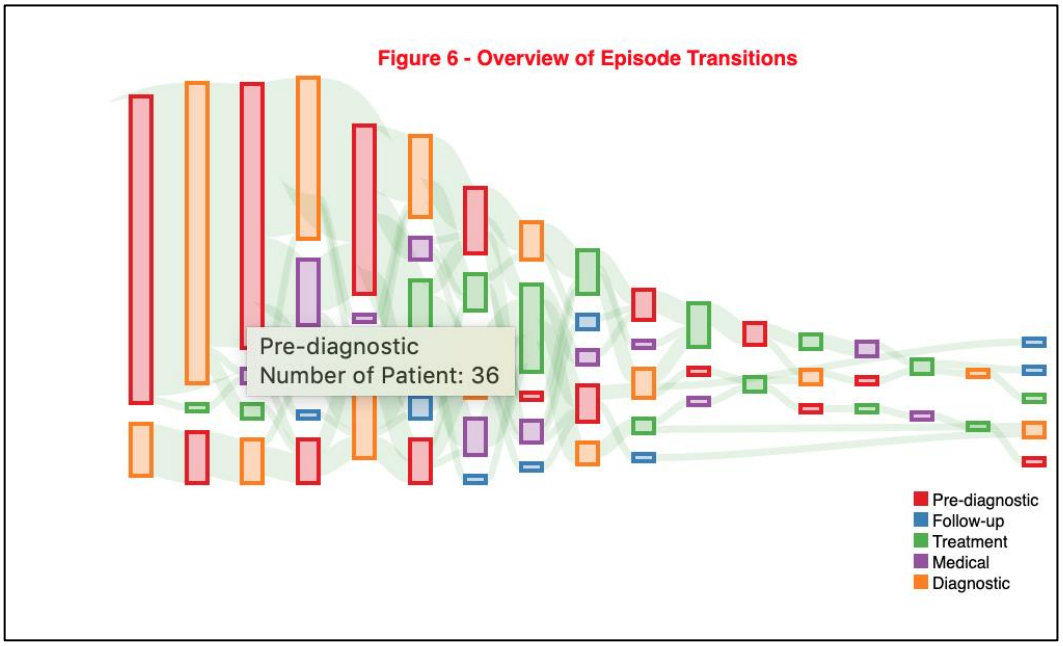


Figure 27. Episode transition chart based on Sankey visualization*, nodes [rectangles] with different colors represent episodes of care and the links between the nodes is shown using the green curve*

4.4 User Interaction with Visual Analytics

I used D3.js (Data-Driven Documents) JavaScript library to create the label prevalence (heatmap), the document distribution (stacked-bar diagram) and the episode transition (Sankey diagram) in the DeepPhe-Viz web browser. I used D3 dynamic user interaction with visualizations to facilitate the filtering feature between different charts. Figure 28 shows how the filtering feature between the different charts in the DeepPhe-Viz was utilized from the original cohort. Figure 28-A illustrates the filtering features using the label prevalence cell cohort of patients. The stack bar document distribution chart, the episode transition chart and the patient list showed the data for the patients who had “Drug” mentions in month 18 in their cancer care. Also Figure 28-B shows how the user can narrow down the label prevalence chart, stack bar document distribution chart and the patient list in a cohort level using a specific episode sequence (pre-diagnostic to diagnostic).

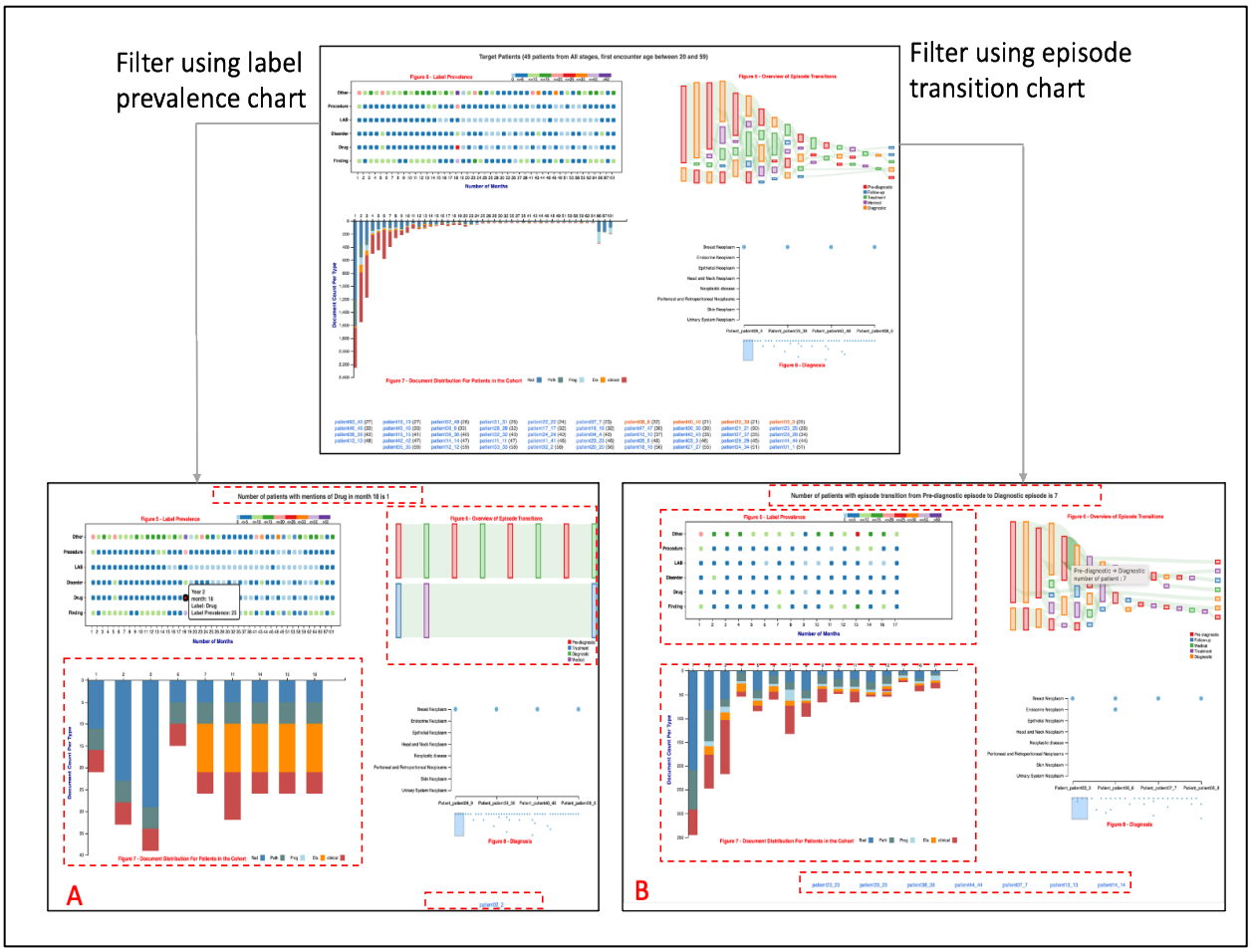


Figure 28. Interaction of heatmap and Sankey diagram visualization in DeepPhe-Viz*. The original views of all the patient in the cohort. A. The changes in different views [with red dash boxes] after filtering using the label prevalence chart. B. The changes in different views [with red dash boxes] after filtering using the episode transition chart.*

5.0 Evaluation of the Enhanced DeepPhe-Viz Interface

5.1 Dataset

A collection of 49 breast cancer patients' data were obtained from the University of Pittsburgh Medical Center (UPMC) electronic medical records between August 2000 through September 2014. This data was used in a previous study (Cancer Phenotype Extraction from Clinical Reports STUDY19020168) and was approved by the University of Pittsburgh Institutional Review Board (IRB Submission - STUDY20090033). This dataset contained several various types of de-identified clinical documents including clinical notes, radiology reports, surgical pathology reports, discharge reports, progress notes. Each document went through DeepPhe pipeline to extract all the required terms, attributes, and relationships. These terms and relationships were used to generate cohort and patients-level summaries. In addition, each clinical document was classified to an episode of care representing different phases in the cancer patient's history including pre-diagnostic, diagnostic, medical decision-Making, treatment, fellow-up, unknown episode using machine learning approach.

5.2 Participants

We recruited participants through professional connections at University of Pittsburgh, the Magee Women's Research Institute and the Hillman Cancer Institute using convenience sampling strategy[96]. We contacted faculty members, graduate students, post-docs, clinical residents, and clinical fellows who have experience in clinical chart reviewing and clinical cancer research. Participants were informed verbally that screenshots, audio and logfiles is recorded during the study. Participation in both studies was voluntary where no compensation was offered.

5.3 Usability Study

The purpose of this study was to acquire feedback about the usability of the DeepPhe-Viz tool and to uncover any usability problems with the system. Two participants were asked to perform a think aloud protocol with the goal to evaluate the usability of the DeepPhe-Viz visual analytics tools and to identify any major errors or difficulties with the system. Participants were asked to think aloud as they explore the tool to evaluate how the users use the tool, what the user understanding about the different features (different charts), and the various interaction functionalities in the tool like using filter the cohort using cancer stage or switching between cohort and patient level.

This study was conducted remotely using the Zoom web conference tool. Participants was informed about the study purpose; what data is collected and how it will be used. The participants were informed that their audio, screenshots and logfiles were recorded through the

study. After learning about the purpose and nature of the study, participants were asked verbally to consent to participate. Then the participants were given a brief 15-minute tutorial about the primary visualization and interaction techniques of the tool. The participants accessed the DeepPhe-Viz web-based tool a URL that is provided to them.

I reviewed the participants comments about the interface to identify any mismatch between user mental models and system. The results from the observations and notes provided a list of usability modifications that used to tweak some changes in the interface.

5.4 User Study

The purpose of this study was to evaluate the efficiency, error rate and user satisfaction of the DeepPhe-Viz tool. This study was conducted in-person in a roughly hour-long session. At the start of each session, participants were verbally informed at the beginning of the study that their activities will be recorded and retrieved using the server log files and screen/audio recording. The participants were provided with a description of the study, and they were asked to verbally consent to participate. Second, a short tutorial about the purpose, different functionalities, and various visualization features of the tool were provided to the participants**.** Third the participants were asked to use web page with a list of tasks that are linked to the DeepPhe-Viz interface a laptop that is provided to them (Figure 29)

Participants were asked to perform a set of tasks (18 tasks) using DeepPhe-Viz. These tasks were developed to assess the success of the DeepPhe-Viz tool in meeting user requirements developed through previous qualitative inquires. They were focused on the use and interpretation of the visualization and examine the user's ability to identify a cohort of interest with specific

criteria then drill down to more details about a specific patient. Each task involved a variety of different types of questions about the data, focused on different aspects of the data, and varied in complexity and difficulties. I grouped the 18 tasks into 5 high-level tasks using the evaluation goals of these tasks. Tasks 1-4 assessed the participants' ability to identify the distributions of the available data. Tasks 5-7 assessed if the participants could filter the cohort based on associated restrictions including categorical or focus exploration. Tasks 8-11 evaluated the interactions between the different features in the DeepPhe-Viz to find a patient(s) with two or more criteria. Tasks 12-15 assessed if the users could view the different temporal patterns in the cohort using the episode transition chart and the label prevalence timeline in DeepPhe-Viz. Finally, tasks16-18 assessed if the users could drill down to the patient level from the cohort level using different criteria including categorical or temporal aspects of the cohort (Table 8).

The participants submitted their answers for each task using the submit button in the web page where these answers were saved in the server log files (Figure 30). The time to complete each task and the screenshots of the participant activities were recorded.

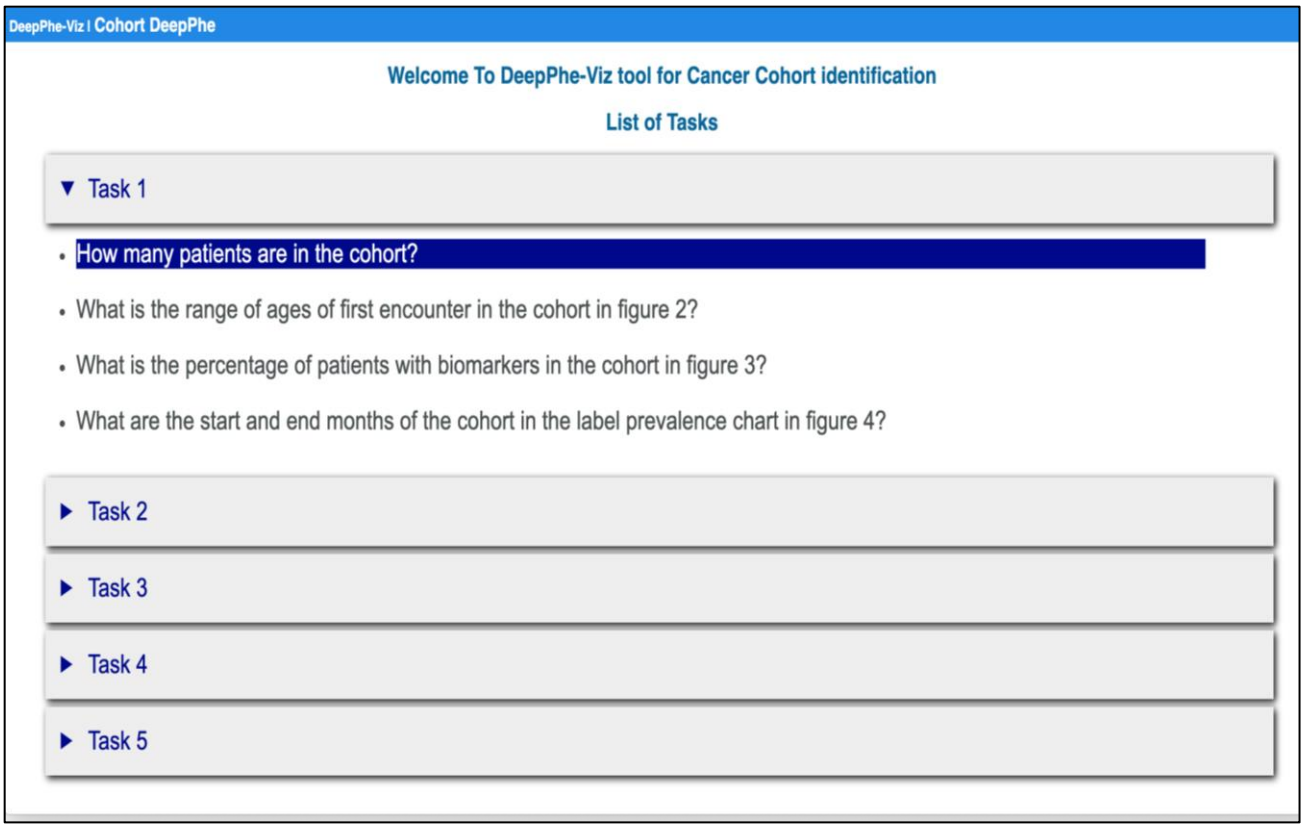


Figure 29. *The* ***first web page that contains the linkable task lists for the users to start the user study***

Table 8. List of tasks to evaluate user requirements using DeepPhe-Viz. Rows *with shading represent the goal of the subtasks and are not shown to the participants*

Task		No	Task	Requirement
A: Distributions	The following tasks are defined to assess if the DeepPhe-viz tool provides the users with an overview of different aspects about the cancer cohort and the ability to explore the distribution of the available data	T1	How many patients are in the cohort?	R1: Overview Identify the number of patients in the cohort
		T2	What is the range of ages of first encounter in the cohort in figure 2?	R1: Overview Identify the overview of the age distribution in the cohort
		T3	What is the percentage of patients with biomarkers in the cohort in figure 3?	R1: Overview Explore the distribution of biomarkers in the cohort
		T4	What are the start and end months of the cohort in the label prevalence chart in figure 4?	R1: Overview Identify the date range for the clinical documents in the cohort
B: Filtering	The following tasks are defined to assess if the users can filter the cohort based on associated restrictions including categorical or focus exploration.	T5	How many patients have stage 2 diagnosis? Name one of them?	R2: Filters R3: Stages Assess if the user can identify patients with specific cancer stage
		T6	How many patients in the cohort with age range between 30 and 40 years old inclusive at the time of their first encounter in figure 2? Name one of them?	R2: Filters R4: Temporal Assess if the user can identify patients with specific range of first age of encounter.
		T7	From the diagnosis chart, how many patients have a diagnosis of Epithelial Neoplasm in the Diagnosis chart in figure 8? Name one of them?	R2: Filters Assess if the user can identify patients with specific diagnosis

Table 8 continued

C: Interaction	The following tasks are defined to assess the interactions between the different features in the DeepPhe-Viz to find a patient(s) with two or more criteria	T8	For one of the stages in figure 1, no patients in the cohort have biomarkers. Which stage is this? Name one patient with this stage of cancer?	R2: Filters R3: Stages R5: Biomarkers and Genomics Assess the user ability to find patients with multiple clinical criteria using stages and biomarkers
		T9	Find one or more patient with **stage 0** in figure 1? What are the type and distribution of biomarkers found for those patients in figure 3?	
		T10	Using the episode transition chart in figure 6: 1. Find one patient who experienced a transition from a treatment episode to a medical episode 2. For that patient, what the month and label have the highest prevalence in the Label prevalence chart in figure 5?	R2: Filters R4: Temporal R6: Treatment R7: Search Assess the ability to conduct search based on temporal events and the interaction between different features in the DeepPhe-Viz
		T11	For Finding label in month 2 in the Label prevalence chart in figure 5: What is the highest number of patients that have the episode sequence of pre-diagnostic to diagnostic in the episode transition chart in figure 6?	

Table 8 continued

D: Temporal Pattern		The following tasks are defined to assess if the users can view the different temporal patterns in the cohort related to the episodes of care transitions or change in the label prevalence trends in the clinical documents through time.	T12	Find number of patient(s) with Drug mention in month 5 from the Label prevalence chart in figure 5? Name one of them?	R2: Filter R4: Temporal Assess if the user can identify a time point corresponding to change in clinical term prevalence
			T13	Find the highest number of patients who have medical to pre-diagnostic episode in the episode transition chart in figure 6? Name on of them?	R2: Filter R4: Temporal R6: Treatment Assess if the user can identify patients corresponding to a specific episode sequence
			T14	Find the highest number of patients who have treatment to pre-diagnostic episode in the episode transition chart in figure 6? Name one of them?	
			T15	Find the highest number of patients who have Pre-diagnostic -> diagnostic-> pre-diagnostic -> diagnostic episodes sequence in the episode transition chart in figure 6? Name on of them?	

Table 8 continued

E: Drill-Down		The following tasks are defined to assess if the users can drill down to the patient level from the cohort level using different criteria including categorical or temporal aspects of the cohort	T16	Find a patient with treatment to follow-up episode sequence in the episode transition chart in figure 6? What is the treatment for that patient in the Cancer and Tumor Summary in the patient level view?	R2: Filter R4: Temporal Assess the user ability to interact with the different views of the DeepPhe-Viz for patients with a specific the temporal events
			T17	Find a patient who has Finding mentions in month 18 using Label prevalence chart in figure 5. For that patient: 1. Find how many documents have treatment episode in the patient level view? 2. What type of documents they are?	R2: Filter R4: Temporal R7: Multi-records Assess the user ability to filter the cohort using different criteria in the cohort level to identify original text used to extract observations in the patient level
			T18	How many surgical pathology reports in the patient level view for the patient with stage 2A, first encounter age between 30 and 36, Finding mentions in month 2 and episode sequence from diagnostic to medical?	

5.5 Measurement

The evaluation of DeepPhe-Viz performance was based on the dependent variables in (Table 9) where time and accuracy were measured for each task. The task completion measured the time starting from clicking the task link in the first page of the DeepPhe-viz web page (Figure 29) to the time when the users submitted their answers (Figure 30), and the task correction measured whether the participants submitted the correct and complete answer for each task. Both the start time, end time and answer for each task were stored in the log file for each user. I used Python Jupyter Notebook[97] to extract the required information from the log files and R[98] to analyze the data. For each task the completion time was calculated for each participant using the difference between the start and the end time for each task. For the task correction, we considered the answer correct if the answer was completely right as for some tasks two parts of answers were required. If part of the answer was right and the other part was wrong, we considered that as half correct and half incorrect. A completely wrong answer was considered as incorrect answer.

Table 9. Dependent variables for the task evaluation

Task	Time to finish each task
	Accuracy of the found results
	Feedback about the visualization tool

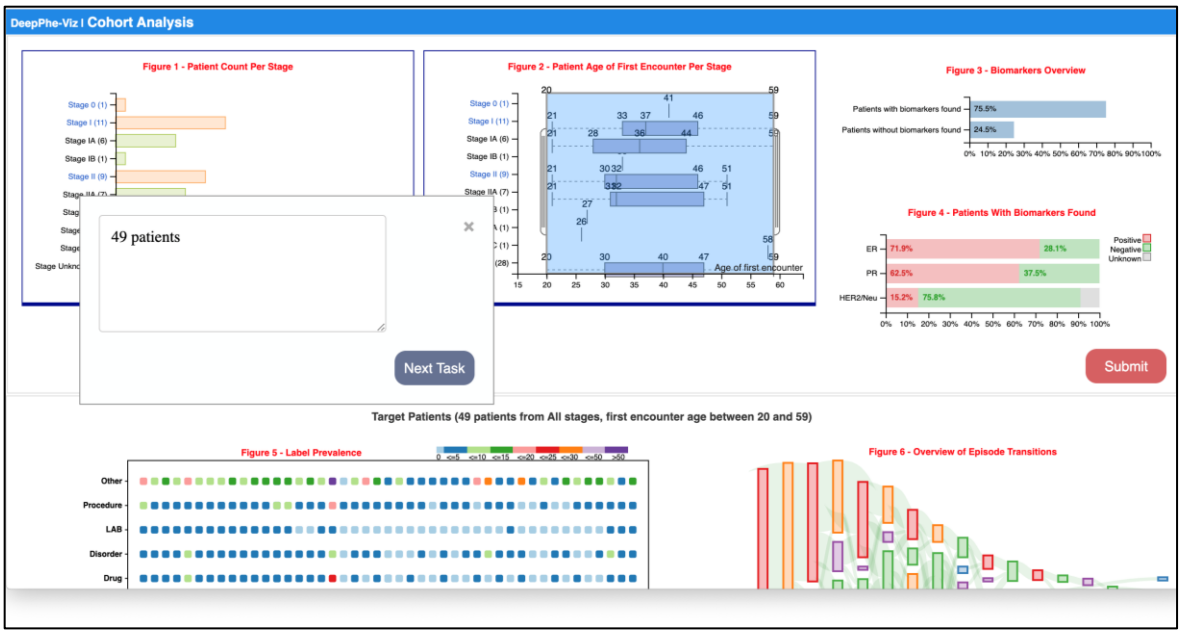


Figure 30.The user interface to submit the answer for each task

5.6 System Usability Scale Questionnaire (SUS)

After finishing all the tasks, the participants completed the system usability scale (SUS)[99] and open-ended questionnaires that was created using Qualtrics XM software (Appendix A). The purpose of system usability scale was to provide a qualitative assessment to gather more subjective feedback about the user experience and the usability of the tool.

The open-ended questionnaire below was designed to assess if DeepPhe-Viz helped tackle the user's information challenges. These challenges were identified in the contextual interviews at the early stages of the development of the DeepPhe project. Table 10 illustrate the identified user challenges.

- What worked well about the tool?
- What did not work well?

- Which tasks did you find particularly easy? Why?
- Which tasks were challenging? Why?
- What ways could the system be improved to better meet your needs?
- Which features of the system you think it was the most useful? Why?
- Which features of the system you think it was the least? Why?
- Is there information you would like, but cannot find in the system? If yes, what are they?

Participants' response for each open-ended question were manually transcribed and analyzed to identify and condense participants feedback into various topics or themes. We used the user challenges that were identified in the early stages of developing DeepPhe-Viz as a basis to inform the development of high-level themes in preliminary codebook (Table 10) [100]. This analysis was intended to identify themes and sub-themes regarding perceptions of the usability of the DeepPhe-Viz tool and information seeking regarding cancer investigators and capture the nuances of the participants needs and challenges with the extended version of DeepPhe-Viz.

Table 10. Sample question to evaluate if the system tackles the user challenges[101]

Challenges	**Sample question (open ended)**
Availability	Does the system provide you with the information that you need?
Accessibility	Is there information you would like, but cannot find in the system?
Quality	Do you feel the system is beneficial? What ways could the system be improved to better meet your needs?
Interpretation	What features of the system you think it will be useful the most? The least? Does the system help you to perform tasks easier?

6.0 Results

6.1 UPMC Dataset Description

Table 11 provide a description about the dataset that was processed through the DeepPhe pipeline and viewed using DeepPhe-Viz interface. The dataset included 49 female breast cancer patients were seen at the University of Pittsburgh Medical Center (UPMC) from August 2000 through September 2014. This dataset ((Table 11) contained 1512 of different types of de-identified clinical documents where clinical notes represent 41%, radiology reports represent 37%, surgical pathology reports are11%, discharge reports are 4%, and progress notes are 7% of the total number of clinical documents. These clinical documents were classified into 6 episodes of care using machine learning classifier. The unknown episode was assigned to 77% (1165 out of 1512) of the total number of the clinical documents where clinical notes represented the largest portion of documents assigned to unknown episode with 46.3% (539 out of 1165) of the total number of documents with unknown episodes, then radiology reports with 29.2% (340 out of 1165), and the remainder was distributed as following: pathology reports with 10.4% (121 out of 1165), progress notes with 8.7% (102 out 1165) and discharge summary with 5.4% (63 out of 1165).

Inferred rules using document type and date assigned a portion of documents with unknown episode into a new episode of care. The number of documents with unknown episode and assigned to a new episode was 379 (32.5 %) where 93 radiology reports with unknown episode were assigned to pre-diagnostic episode, 121 pathology reports were assigned to

diagnostic episode, 121 progress notes and 63 discharge summaries were assigned to treatment episode. The only document type that the unknow episode was not resolved was the clinical notes. We excluded 786 documents with unresolved unknown episode which represented 52 % of the total number of clinical documents in the dataset.

The results of the error analysis for the episode inference are shown in Table 12. Clinical notes represented the most type of clinical documents with unknown episode. One issue with clinical note was the mislabeled of these documents. These documents included information about radiology results, pathology results, progress note or discharge summary, but they were labeled as clinical note. The other issue was the existing of clinical documents that are irrelevant to cancer diagnosis. For example, there could be a radiology note related to a non-cancer disease such as pneumonia or a pathology note describing the excising of hernia.

Table 11. Description of the UPMC dataset *including the document types and the assigned episode of the care for each document type for 49 patients. The first column represented the type of clinical documents. The second and the third column showed the count of different episodes for each type of clinical document that the machine learning generated. The fourth column illustrated the number of inferred Unknown episodes according to the document type. The fifth column represented the number of documents with unknown episodes that could not be inferred and excluded from the study. Finally, the last column showed the total number of episodes from the machine learning and the inferred unknown episodes after excluding the un-inferred Unknown episodes.*

Document Type	**ML Classifier**		**Inferred unknown episode**		**Total number of Episodes from ML and inferred unknown**
	Episode type	**Count**	**Number of Inferred Unknown Episode**	**Number of un-inferred Unknown episode (Exclude)**	
Clinical Note	Medical Decision	47	-	-	47
	Treatment	34	-	-	34
	Follow-up	2	-	-	2
	Unknown	539	-	539	0
	Total	622	0	539	83
Radiology	Pre-diagnostic	178	93	-	271
	Medical Decision	1	0	-	1
	Follow-up	39	0	-	39
	Unknown	340	0	247	0
	Total	558	93	247	311
Surgical Pathology	Diagnostic	39	121	-	160
	Unknown	121	0	0	0
	Total	160	121	0	160
Discharge Summary	Treatment	3	63	-	66
	Unknown	63	0	0	0
	Total	66	63	0	66
Progress Note	Treatment	4	102	-	106
	Unknown	102	0	0	0
	Total	106	102	0	106
Overall Total	**All Episodes**	**1512**	**379**	**786**	**726**

Table 12. Error analysis of a random number of clinical documents *to identifies machine learning classifier's challenges within the different type of clinical documents*

Challenge	Explanation
Issue with clinical note document type	Clinical note documents contain disparate type of other type of documents such as discharge summary, pathology report, progress note and radiology reports
Document not related to cancer type	Some documents in the cancer treatment are not related to the cancer care journey including other conditions that cancer patients could have. For example, injury or pneumonia.
Duplication of the documents	Some of the documents exist twice in the dataset which produce redundant unknown episodes
Content of clinical documents	Some content of the clinical documents is more challenging to be classified to a specific episode of care because they include information about the patients that are not related to specific episode of care.

6.2 Usability Study

Two participants conducted a useability study for the DeepPhe-Viz interface focusing on the display of the interface and interpretation of the tasks list context. Both participants had experience in biomedical research (associative professor and assistant professor at school of medicine).

The participants suggested some modification in the interface appearance including color platelets, adding figure numbers, positions of the charts, legends, and tooltips in the DeepPhe-Viz. In addition to that the participants helped in the development and phrasing the tasks list in the first page to cover a wide range of aspects about the DeepPhe-Viz functionalities.

6.3 User Study

6.3.1 Participants

Twelve cancer investigators representing a variety of cancer translational expertise with an experience range from 3 years to 22 years participated in the user study. Six of the participants were clinicians including medical doctor (1), medical oncologist (3), medical resident-oncology (1) and physician assistant-oncology (1). The other six participants were translational cancer researchers including an assistant professor (1) with research interest in breast cancer, postdoctoral fellows (1), graduate students (2), clinical research coordinator who worked with cancer EMR data, and finally a medical student (1) who worked on project using cancer EMR data. Their demographic information is summarized in Table 13.

Table 13. Demographic summary about the user study participants

Number of participants (n)	12
Clinician	6
Researcher	6
Gender (N)	
Female	4
Male	8
Job Title(N)	12
Medical Oncologist	3
Doctor of Medicine	1
Physician Assistant-Oncology	1
Assistant Professor- Oncology Laboratory	1
Doctor of Medicine- Oncology Resident	1
Postdoctoral Scholar- Oncology Laboratory	1
Graduate Student- Cancer Research	2
Medical Student- Cancer Research	1
Clinical Research Coordinator- Oncology Laboratory	1
Experience in cancer research (year)	
Mean	8.75
Median	7
S. D	5.17
Max	22
Min	3

6.3.2 Task Completion Time

Twelve participants performed 18 tasks using DeepPhe-Viz. Figure 31 shows the total time completion for all the 18 tasks for each participant where user 10 (researcher) showed the least average time completion in 49.8 seconds (S.D. 27 seconds) and user 1 (clinician) showed the highest average time completion in 188.5 seconds (S.D. 87 seconds). The least average time completion was for Task 3 in 26.5 seconds (S.D 20.5 seconds), while Task 17 took the longest time with average time completion 148 seconds (S.D 82.5 seconds) for all the 12 participants (Figure 32).

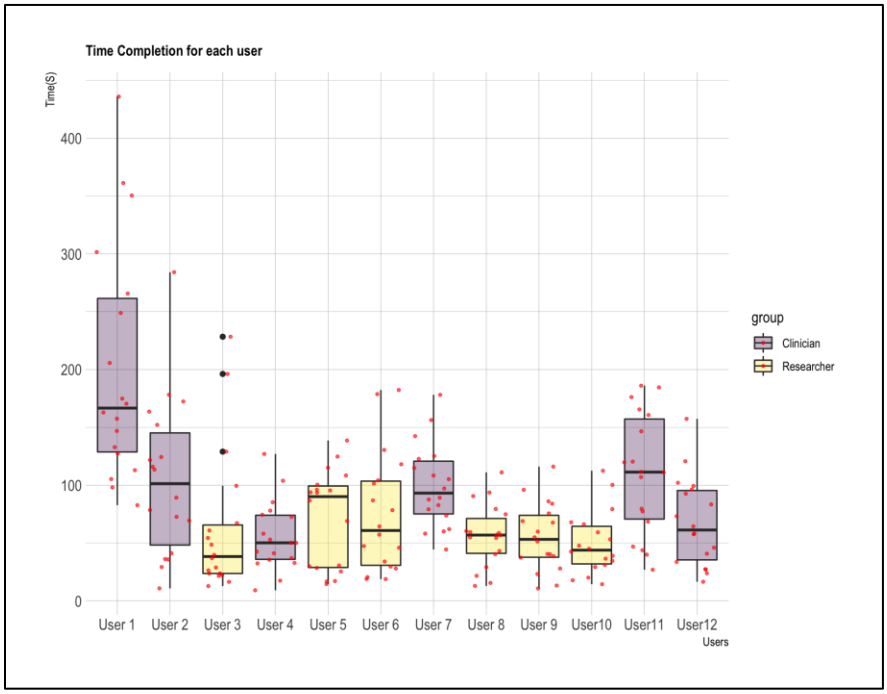


Figure 31. Time completion for each participant *(12 in total). The boxes and whiskers represent the median and interquartile ranges. The solid black circles represent the outliers.*

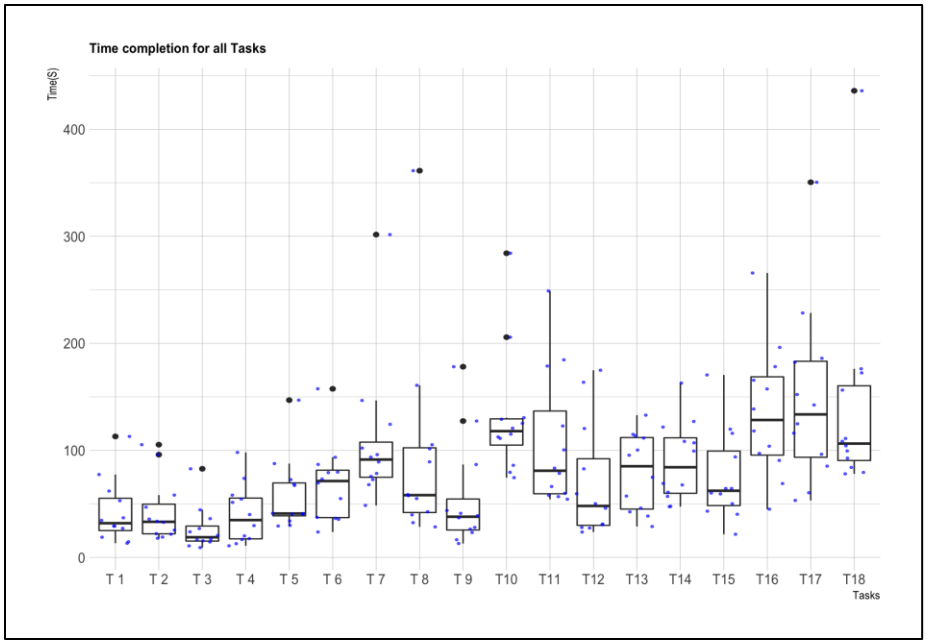


Figure 32. Time completion for each task *(18 in total). The boxes and whiskers represent the median and interquartile ranges. The solid black circles represent the outliers.*

Figure 33 shows the total time completion for all the high-level tasks for all the participants. Task A average completion time was completed in 75.9 seconds with the least amount of time, then task B in 156.9seconds. The average completion time for Task D was 157.2 and task C in 194.2 and finally task E in 283.4 seconds. Table 14 shows the average time completions for the 5 high-level tasks (A, B, C, D, E) for each participant type (clinician and researcher)

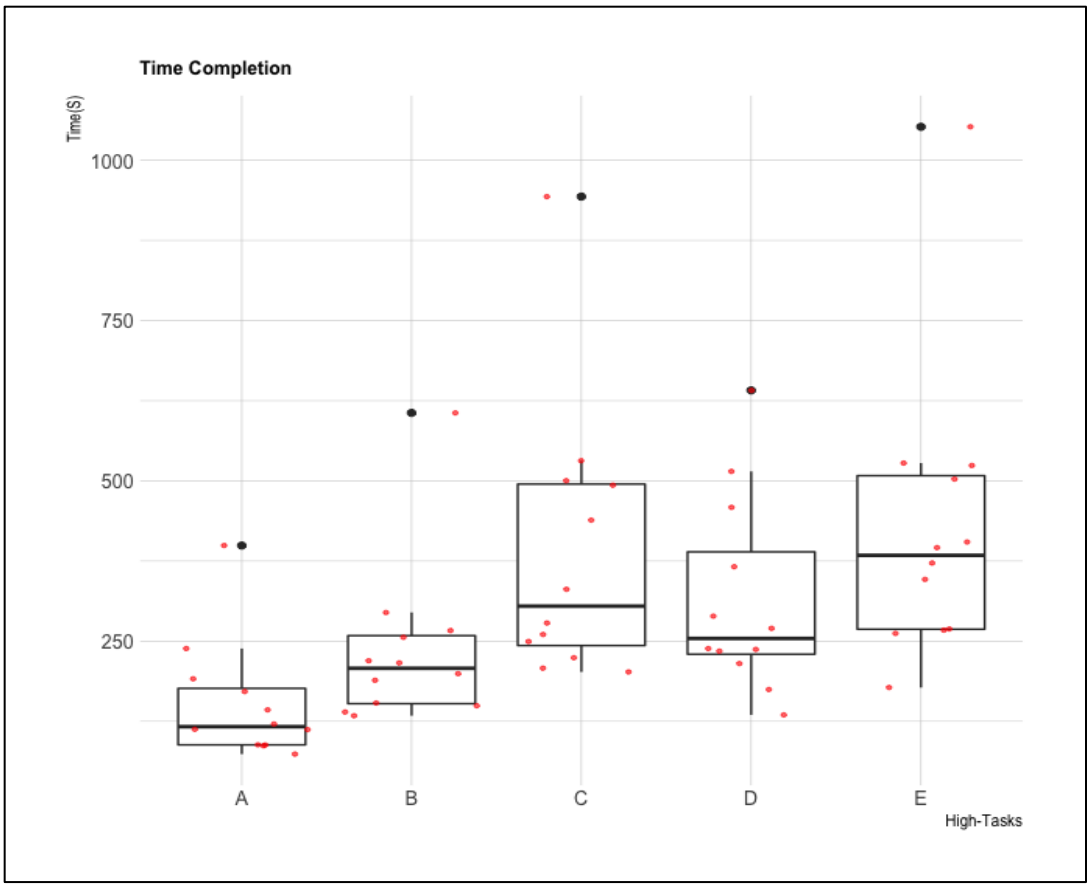


Figure 33. High-level tasks completion time for all participants*. The boxes and whiskers represent the median and interquartile ranges. The solid-colored circles represent the outliers.*

Table 14.Time completion of the high-level tasks for each participant type

High-Level Task	Sub-Tasks	Assessment	Clinicians		Researchers	
			Avg (s)	S.D (s)	Avg (s)	S.D (s)
A	T1, T2, T3, T4	Identify the distributions of the available data	49.9	29.1	26.2	18.9
B	T5, T6, T7	filter the cohort based on associated restrictions including categorical or focus exploration	99.4	62	57.5	25.1
C	T8, T9, T10, T11	the interactions between the different features	122.9	85.9	71.3	41.5
D	T12, T13, T14, T15	view the different temporal patterns in the cohort	103.7	42.7	53.6	22.4
E	T16, T17, T18	drill down to the patient level from the cohort level using different criteria	171.8	94.2	111.6	49.5
Total			547.6	313.9	320.3	157.1

A two-way repeated measures ANOVA was performed to evaluate the interaction between participant type, task type and time completion. There was a statistically significant interaction between participant type on the overall time completion (p-value = 0.000133) where the average completion time for clinicians was 1.7 slower than the average completion time for researchers (547.6/320) for all the tasks (Figure 34).

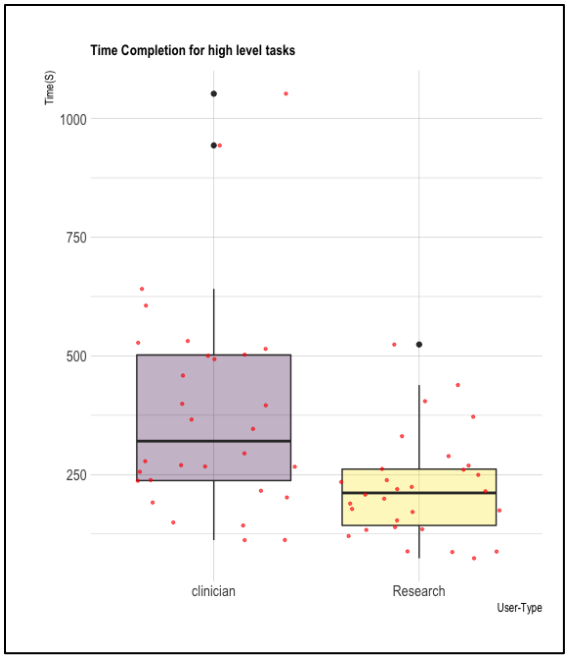


Figure 34. Overall time completion for each type of participants*. The boxes and whiskers represent the median and interquartile ranges. The solid-colored circles represent the outliers.*

Therefore, the effect of participant type variable was analyzed for each high-level task (Table 15). P-values were adjusted using the Bonferroni multiple testing correction method. The effect of participant type was significant at task D (p-value = 0.044) but not at the other tasks where task A (p-value = 0.093), task B (p-value = 0.145), task C (p-value = 0.168), and task E (p-value= 0.096) (Figure 35).

Table 15. P-values for the corresponding the high-level tasks for each participant type

	Clinicians					
Researchers		**A**	**B**	**C**	**D**	**E**
	A	0.093				
	B		0.145			
	C			0.168		
	D				0.044	
	E					0.096

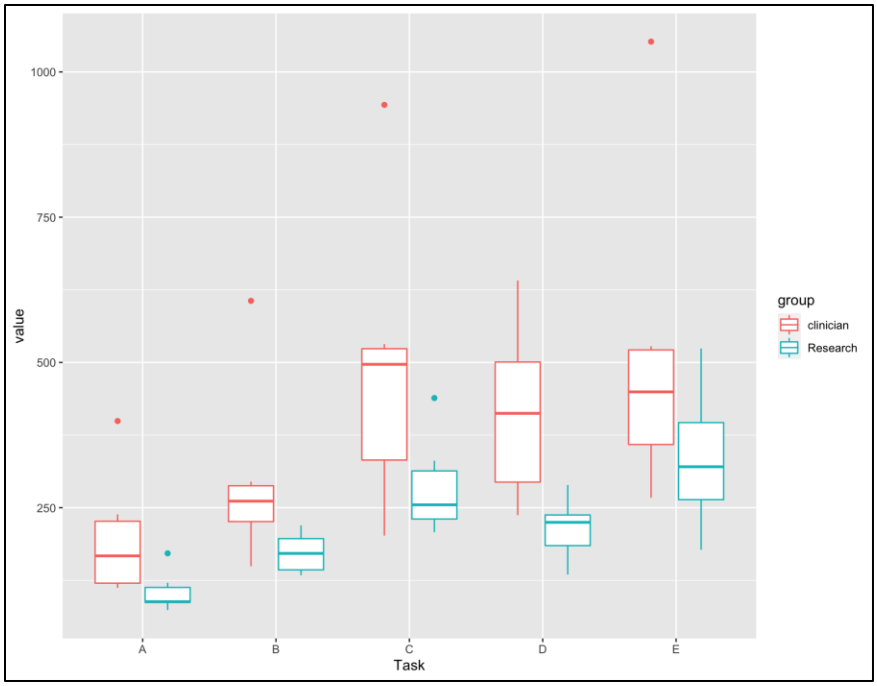


Figure 35. High-level tasks time completion for each type of participants*. The boxes and whiskers represent the median and interquartile ranges. The solid-colored circles represent the outliers.*

6.3.3 Task Correctness

The percentage of correct answers submitted by the participants was 95%. Of the two types of participants, 3 participants (2 clinicians and 1 researcher) found all the correct answers 100% (Table 16). The correctness score was 97.2% for clinicians and 93.5% for the researchers' participants. A Chi-square test showed no significant difference between clinicians and researchers in the rate of task correctness ($\chi 2(6.6667$, p-value = 0.6718).

Table 16. Correct and incorrect scores for the two types of participants

User	Group	Correct	Incorrect
User1	Clinician	18	0
User2	Clinician	17.5	0.5
User4	Clinician	16.5	1.5
User7	Clinician	17.5	0.5
User11	Clinician	17.5	0.5
User12	Clinician	18	0
Overall Percentage	Clinician	97.2%	2.8%
User3	Researcher	15	3
User5	Researcher	15.5	2.5
User6	Researcher	17.5	0.5
User8	Researcher	18	0
User9	Researcher	17.5	0.5
User10	Researcher	17.5	0.5
Overall Percentage	Researcher	93.5%	6.5%
Overall total	All	206	10
Percentage	All	95%	5%

Table 17 shows the correctness and incorrectness scores for the 18 tasks participants answers. The total number of incorrect answers was 10. Task 8 assessed the user ability to find patients with multiple clinical criteria using stages and biomarkers and Task 13 assessed the user ability to identify patients corresponding to a specific episode sequence. Those two tasks have the highest incorrect rate for all the participants (4.5 incorrect scores) where the participants found an answer, they believed to be correct, but it was not the correct answer**.** For the

remaining 5.5 incorrect answers, the participants only submitted part of the answer and not the complete answer.

Table 17. Correct and incorrect scores of each task

Task	Correct	Incorrect
Task 1	12	0
Task 2	12	0
Task 3	12	0
Task 4	12	0
Task 5	11.5	0.5
Task 6	12	0
Task 7	11	1
Task 8	9.5	2.5
Task 9	10.5	1.5
Task 10	12	0
Task 11	12	0
Task 12	10.5	1.5
Task 13	10	2
Task 14	12	0
Task 15	11.5	0.5
Task 16	11.5	0.5
Task 17	12	0
Task 18	12	0
Total	206	10

6.3.4 System Usability Scale Questionnaire Results

The SUS provided a 100-point scale with certain cutoffs for different usability levels. The average SUS score for all the participants was 65.2 which is considered a good score[102]. The participant's distributions were normally distributed but didn't show equal variance so Welch- t-test was used for the analysis. The average SUS scores were slightly higher for clinicians (65.42)

than the researchers (65) but were not statistically significantly different (p-value = 0.79) (Figure 36).

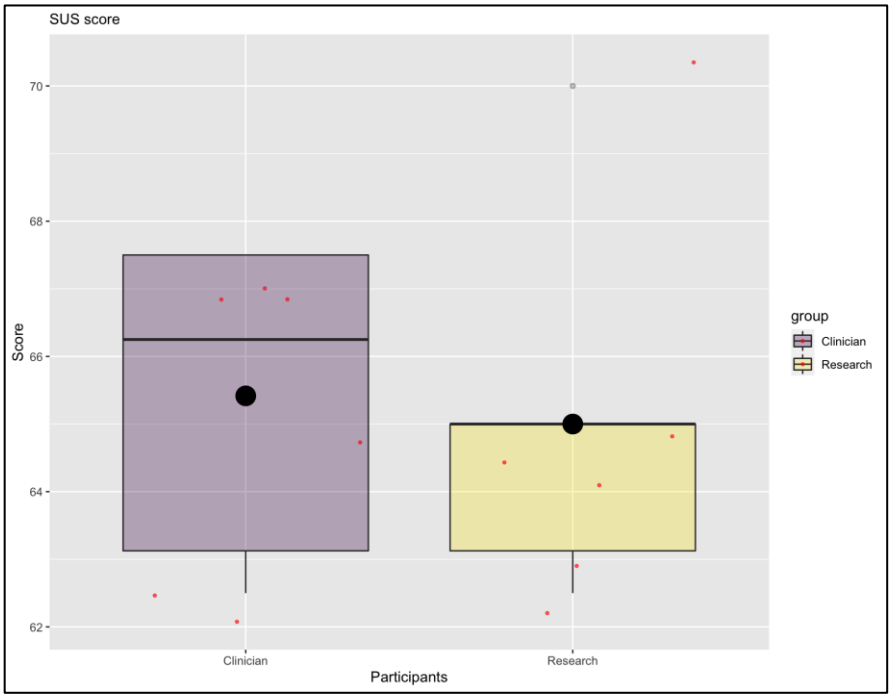


Figure 36. **System usability scales by participants**. *The boxes and whiskers represent the median and interquartile ranges. The solid circles represent the mean SUS scores.*

6.3.5 Open-End Questions

An open-ended questionnaire was filled by the participants after performing the tasks using DeepPhe-Viz interface. We developed a codebook using the participants' responds for basic questions that address the strengths and weakness of the DeepPhe-viz tool to obtain qualitative feedback about (1) the participants perceptions about the tool, (2) usability and the layout of the tool (3) structure and the interactive between the tool features (4) challenges and difficulties using the tool and (5) missing data and functionalities that the participants were needed. Table 18 provides a summary of the high-level themes used to code the participants responses, a breakdown of each of the sub-steps involved in each theme, and an example response for each theme.

Table 18. Main categories and themes associated with DeepPhe-Viz tool conveyed by participants.

Perception of the tool	N=12	Example Quote
Interface usability		
Easy to operate	7	"Simple, easy-to-use interface"
Interactive	4	"Easy to read and point-and-click format is very user friendly"
Well visualized	1	"Well visualized to understand at a glance"
Simple to use	3	"The interface is simple to use and the quick sorting of cohorts as parameters"
Very functional	1	"It is very functional and well thought out."
Nice interface	2	"Nice interface, easy to read and point-and-click format is very user friendly"
Color scheme	1	"Color scheme was bright, and contrast of colors groups was clear very interactive"
Clinical information	2	"A lot of high-level clinical information is available and easy to see the pattern of the timing of care."
Visual summary	2	"This tool provides a wealth of information that allows us to obtain clinically desirable data on large number of patients in a fraction of the time"
Useful	5	"The interface is simple to use and the quick sorting of cohorts as parameters were narrowed allowed me to rapidly find useful information related to the patients I was searching for "
Structure of the tool		
Diagrams and figures	10	"I liked how all of the diagrams were interconnected"
Drill down	5	"You could use this functionality to really drill down to your cohort of interest"

Table 18 continued

Click and scroll	7	"Very easy to click and scroll”
Filter information/ patients	13	“Filtering patients of interest by cancer diagnosis and stage and age at time of diagnosis”
parse patients	1	“Finding the number of patients for a certain category was easy”
Functionality of the tool		
Age feature	7	Selecting a cohort based on stage and age - and biomarkers, if possible"
Stage feature	5	Selecting a cohort based on stage and age - and biomarkers, if possible"
biomarker feature	4	Selecting a cohort based on stage and age - and biomarkers, if possible"
diagnosis feature	6	“Counting the number of diagnoses was difficult in the example”
Episode feature	7	"The episode transition diagram was useful, but it was sometimes hard to be able to focus your mouse on a specific path you were looking for.”
Label feature	6	“Identifying charts which include certain terms. This would be helpful for retrospective chart reviews”
Number of patients	4	"Graphical display and showing the number of patients in parenthesis were helpful. "
Patient level	5	“Looking at patient level view and understanding it was not easy."
Type of information in the tool		
Wealth of information	3	“This tool provides a wealth of information that allows us to obtain clinically desirable data on a large number of patients in a fraction of the time it would have taken to perform this manually other positive points: color scheme was bright, and contrast of colors groups was clear very interactive "
High level clinical information	1	“A lot of high-level clinical information is available and easy to see the pattern of the timing of care”

Table 18 continued

Granular information	1	“Notes were all assembled in the patient view so more granular information about treatment can be easily seen."
Information need	3	“Knowing the prevalence of information available may not significantly change the question I would be asking "
Available information	2	“Inclusion of more diagnostic details such as molecular testing or next generation sequencing”
Guidance requirement		
Explanation needed	4	"I think it will need some instruction and brief explanation for each chart and how to use for the first-time users."
Additional data to the tool		
Clinical data	6	“This tool needs to include more types of clinical data”
Missing information	1	"Mainly the missing information that I still have to extract.”
Treatment data	5	"The ability to pick specific drug treatments. ex.) palbo, AIs, Tamoxifen, bone targeting agents"
Biomarker data	3	“This tool needs to include more types of clinical data such as recurrence free time, overall survival, ER/PR/HER2 expression level, menopausal status, PAM50 types (if possible), and others.”
Survival data	3	“This tool needs to include more types of clinical data such as recurrence free time, overall survival”
Therapy response	2	“Inclusion of all chemotherapies received in the patient information section, inclusion of responses to each line of therapy "
Additional Functionalities to the tool		
Save patients of interest	1	"I would also appreciate a way to save your patient”

Table 18 continued

Export feature	2	“Also, an excel export function would be useful”
Term search filter	6	"I would LOVE to be able to use a search by keyword as shows in the breakdown"
Advance features	4	"It would be great to link the patient case files to actual samples we have in the tissue bank. "

7.0 Discussion

In this dissertation, I aimed to extend the functionalities of a cancer visualization tool to meet some of the user requirements that were not implemented into DeepPhe-viz preliminary cohort view. These requirements were identified through contextual inquiries with cancer researchers. The goal of adding these features was to improve the cancer cohort identification process by uncover aspects about the cancer cohort that are hidden in the free text part of the EMR data in timely and accurate manner. While there are well-known visualization approaches to convey the structure format of clinical data, visualizing information within the free text of the clinical documents in a useful fashion remains an open research question. The interactive heatmap and Sankey visualization, along with the other features of DeepPhe-Viz demonstrate a contribution to this discipline for clinical and translational research.

Using a task-based evaluation setup including the task time completion and task correction we demonstrated the user ability to utilize the different functionalities of the DeepPhe-Viz and drilled down to patients of interest with more complex characteristics. In addition to that, the results showed that different types of users who are involved in cancer clinical or cancer translational research could benefit from the DeepPhe-Viz tool.

Furthermore, the results of the System Usability Scale (SUS) and the questionnaire indicate that the additional functionalities and views were useful and helpful to cancer investigators to identify a cohort of cancer patient with different data characteristics. The first functionality embodies the utilization of the visual analytic heat map artifact to encapsulate the content of the clinical documents and demonstrate the change in the semantic groups.

These semantic groups were identified using the is-a relationship incorporated into neo4j from NCIt. Grouping terms into high-level sematic groups facilitated the visualization of the clinical terms into the DeepPhe-Viz and will enable multi levels of granularity to be explored and viewed within DeepPhe-Viz. These semantic groups are indictive of aspects of care and provide a way to summarize the data into high level groups. However, one of the semantic groups contained many terms that were aggregated under "Other" group because they were not related to any of the defined semantic groups (finding, disorder, drug, lab and procedure semantic groups). The "Other" semantic group didn't provide any information about these terms. Subgroups the terms in the "Other" semantic group into less granular groups could provide more information and facilitate the exploration of the clinical documents contents.

The heatmap view provided an overview of the clinical documents' contents for all the patients in the cohort which facilitated exploratory search[103]. Also comparing the change in prevalence between the semantic groups through the timeline of diagnosis for all the patients could highlight insights about cancer trajectories and patterns.

Aligning and normalizing documents dates supported the visualization of the prevalence changes for each sematic group for all the patients in the cohort. However, the assumption that the document with the earliest date for each patient is marked as the starting timepoint (month 1) was not an accurate and a reliable assumption. Incorporating additional features about the document such as various mentions of specific terms, events or the episode of care assigned to the document may provide a better way to identify if this document should be anchored to the first month. The incompleteness of the data set used in this project also contributed to the challenge of identifying the starting time point of the cancer diagnosis where for some patients only a portion of clinical documents was available.

Many participants expressed their desire to be able to identify additional mentions of clinical data in the clinical documents including survival data, recurrence, treatment responses, menopause status. The heatmap visualization provided the participants with the first step to filter out patient with specific semantic group, which narrowed down the number of patients that the investigators need to manually go through their clinical documents to identify patients with specific clinical data. For example, if the cancer investigator is looking for cancer patients with specific treatment such chemotherapy, the interactive heatmap could facilitate that process by allowing the investigator to view only patients with treatment mentions in their clinical documents at a certain time of cancer diagnosis.

Providing cancer researchers with high-level details about the different treatments using the heatmap representation was necessary to explore the trends in the contents of the clinical documents but not sufficient to identify specific mentions of treatment. Utilizing multiple-level visualizations moving from the semantic groups to specific mentions of treatment may support the user requirement to identify patients with specific treatments. One participant addressed this issue: *"I would have liked the ability to find out more specific treatment information under drug. Right now, the options under label preference are not that specific or helpful for my questions."* The second functionality represented the adoption of the Sankey diagram to demonstrate the temporal transitions between different episodes of care in the cohort and view patterns and trends in cancer treatment. One participant mentioned that the Sankey diagram was a useful feature in the DeepPhe-Viz: *"the Sankey graph of the treatment history. It makes it much easier to follow the treatment of a patient".*

However, excluding many clinical documents with unresolved unknown episodes from the Sankey diagram is still an issue. The classification of different documents into different

episodes of care were informed by previous work that modeled cancer care trajectories[104]. The machine learning classifier was unable to classify many documents into corresponding episode of care and assigned them to unknown episode. A preliminary error analysis was performed by manually reviewed number of clinical documents with unknown episodes. The type and date of clinical documents were used as features for machine learning algorithm to classify documents to predefined episodes. Clinical note document type was the most document type assigned to unknown episode, and this was because of the mislabeling of the document type and the content of the document. Number of these clinical documents headers was marked as clinical notes while the actual content of these documents was either discharge summary, progress note, or pathology report documents. Identifying the actual type of clinical document is essential for the SVM classifier as the type of document was one of the features to determine the episode of care for that document. Second, the existing of duplication in the clinical documents with unknown episode also increased the number of unknown episodes. Prior work quantified the effect of text duplication where deducting duplication in large number of documents is an important problem in machine learning and data mining[105, 106] and would improve the performance of the machine learning algorithm.

The other issue with assigning unknown episodes to different clinical document types were the indistinct nature of the clinical documents' contents where some of these documents contained events about the patients that were not related to cancer care. The machine learning algorithm identified specific terms in the clinical documents such as tumor, malignancy, treatment options as features to train the classifier and some of these clinical documents didn't have any of these indications. For example, an injury or chest pain notes. Furthermore, some clinical documents contained events that were related to cancer care but not to any of the

predefined episodes of care. For example, some events like procedures or imaging studies that happened after the diagnosis of cancer and before the treatment started cannot be classified to either diagnostic or treatment episodes. This issue could be improved by expanding the training data of the machine learning and including manually annotated documents with the proper episode of care for more ambiguous clinical notes.

Classifying the clinical documents into the episodes of care represent a high-level and granulated demonstration of the temporal transitions in the clinical documents and there is still a demand to automatically identify and track clinical events within these clinical documents using event extraction techniques, ordering[107], and reasoning[108]. For example "Clinical threading"[109] a semantic model that extracted and arranged temporal events using rule-based approach to assemble events on a timeline. Exploring less coarse demonstration of temporal events in the clinical documents and identify the occurring subsequences within a collection of longer event sequences would facilitate combining related events within a certain time frame such as combining multiple procedures that are related to cancer surgery event. Incorporating domain-specific knowledge, temporal reasoning and mining frequent patterns of temporal events could identify the transitivity inherent and temporal hierarchical relations between timestamped events. This would generate a detail representation of temporal disease progression and facilitate the comparison of different diagnosis and treatment regimens between patients. Finally, creating an interactive visualization that could highlight and identify connections between events and within different clinical documents could help solving the inherent ambiguity in linking related events within the cancer diagnosis timeline[4].

A few other participants indicated their concerns about different aspects of the tool. First the "*need of some instruction and brief explanation for each chart and how to use for the*

first-time users". Another participant specified the same issue by expressing the following: "*I struggled some with the advanced level questions as I hadn't used the system before".* This concern could be addressed by providing the participants with some usage scenario examples and well-structured documentation.

Other participants expressed the challenges faced them while using the Sankey diagram to identify patients with specific episodes transition. The challenge with the Sankey diagram was understanding the purpose and the elements of the episode transition chart. Classifying clinical documents into episodes of care and demonstrating the transitions between different episodes of care was a new concept for some of the cancer researchers and the Sankey diagram had never been used in this setting. Thus, providing the users with some usage scenario demo and simplify the layout of the Sankey diagram could help them understand the idea of the Sankey diagram and how to use it: *"The only difficult task was utilizing the map to transition between Encounter types. Only because I have never used one similar before and it was a new process for me. I think once I got more experience with the task it too would become intuitive."*

7.1 Participants' Feedback

Overall, the participants stated their need and appreciation of implementing such a tool for identifying cancer cohort from EMR data for retrospective study: *"This tool provides a wealth of information that allows us to obtain clinically desirable data on a large number of patients in a fraction of the time it would have taken to perform this manually"*. A participant specified his perspective about the tool saying that there "*lot of high-level clinical information is available and easy to see the pattern of the timing of care. Notes were all assembled in the patient view so more granular information about treatment can be easily seen*". Another

participant expressed the benefit of including the heatmap feature into the DeepPhe-Viz: *"identifying charts which include certain terms. This would be helpful for retrospective chart reviews".*

Most participants found DeepPhe-viz to be ''*easy to operate*", "*well visualized to understand at a glance*", "*simple, interactive*", "*user friendly*" and ''*very useful*.'' They mentioned that "the *diagrams were interconnected"* which provide them with the *"functionality to really drill down to cohort of interest"* details about the patient using "*point-and-click format*" and "*selecting the age brackets for cohort of interest"* where*" the sliding bar facilitated this".* They also stated that it provided "*honing in on the patient charts with desired characteristics without manually reviewing them".*

The participants mentioned number of suggested improvements to the interface. The first area of improvement was the episode transition diagram where the transition arcs between the episodes were overlapped which require more time to find a specific transition. One participant said that *"the episode transition diagram was useful, but it was sometimes hard to be able to focus your mouse on a specific path you were looking for".* Other participant also mentioned that "*to identify patient status change (pre-diagnosis, diagnosis) you have to scan the whole figure to find the information you need "*. This issue could be addressed by expanding the layout of the Sankey diagram, representing the different transitions arcs using color-coded pathways and providing the users with the ability to view only the episode transitions of interest using an interactive legend that shows or hides different episodes of care pathways

One participant pointed-out that the diagnosis chart needed some improvement in determining the number of patients with specific diagnosis where *"counting the number of diagnoses was difficult. I would prefer there be a way you just click on the diagnosis of interest*

and then it would give you a number". This issue could be solved by viewing the number of patients when selecting a specific diagnosis using a tooltip feature and updated the DeepPhe-Viz patients list component with the patients with the selected diagnosis. Another required functionality that the participants needed in the diagnosis chart was the ability to filter patients using the diagnosis and identifying patient with subtypes of cancers*: "Utilizing a means of searching by diagnosis (neoplasm) or by tumor marker would be nice."* Filtering patients using the diagnosis is feasible and can be incorporated into the tool however identifying patients with cancer subtypes requires additional steps to detect the hierarchy relationships between different cancer types and subtypes.

Participants provided feedback to extend the DeepPhe-Viz functionalities including the ability to filter patents using the biomarkers status: *"would be helpful to separate and filter patients based on biomarker status - this will be key in several different kinds of cancer (for example mutation status in melanoma but more importantly in being able to distinguish ER positive breast cancers or HER2 positive breast cancers or Triple negative patients).* And identifying patients with details about breast cancer subtypes: *"being able to separate breast cancer patients by histology (ductal, lobular, metaplastic, etc.)*. Moreover, incorporate additional clinical data including survival outcome, treatment, menopause status, recurrence was repeatedly mentioned by the participants: *"the ability to pick specific drug treatments. ex.) palbo, AIs, Tamoxifen, bone targeting agents"*.

Some of these recommendations are amenable and feasible to achieve including filtering by biomarker status or diagnosis. Including additional clinical data into the DeepPhe visualization will require linking individual mentions throughout the course of cancer treatment

to identify recurring concepts and include additional visualization that will embody the individual terms in each specific semantic group.

7.2 Limitation and Future work

This study is limited by the small size of the UPMC dataset for cancer patients. Only 49 patients' EMR were viewed in the DeepPhe-Viz and this dataset is mainly about breast cancer patients. We expect to be able to use DeepPhe-viz to view the data of a larger number of patients' EMR with breast, ovarian and melanoma cancer. Because DeepPhe-Viz uses only UPMC dataset thus far, our evaluation focused on patients from a single institution. However, DeepPhe NLP was designed using interoperability framework (FHIR, NCIt) which support using other institutions EMR into the DeepPhe NLP to assess the reproducibility of DeepPhe-Viz interface.

Another limitation with this study was the issue of excluding large number of documents from the episode transition chart with unsolved unknown episode. Increasing the size of the training dataset for the machine learning and including manually annotated episodes for documents with inexplicit features into the training set will help solving the issue of the unknown episode.

The evaluation study had several limitations. First, the study used user study to evaluate the interface. Alternative approaches such as conducting focus group study or developing a case study could have been used to address long-term and in-depth insights about the tool and may have found different results. Although we recruited participants with different roles in retrospective cancer research including clinical and translational, the small sample size with 12 participants in the user study could not be generalizable to all the cancer researchers. It is also

possible that the inexperience of participants with this type of interface is a limitation, however it provided us with insights about the current state of the participants potentials and capabilities. Future works on advancing the capabilities of the DeepPhe NLP to identify more ambiguous and challenges terms including treatment outcome, treatment repones, and survival rate. Identify these sentinel events within the clinical documents will facilitate the comparison between different treatment regimens and help in answering some temporal queries. Also integrate a full text indexing into DeepPhe-Viz will facilities text searching using keywords or events to identify similar patients.

The long-term goal of this project is to develop a generalizable computational infrastructure that will be used in other cancer domains especially cancers with rare incidences (e.g., Sarcoma).

In addition to that extend this model to extract all the information in other chronic diseases could be possible. additional efforts will be required to reuse the DeepPhe pipeline for other cancers or diseases. Including interviews with domain experts to inform the model entities and attributes and the rules to aggregate different mentions into a high-level phenotype. Extending different ontologies will be also needed to represent the required vocabularies, concepts, and relationship for each domain.

7.3 Conclusion

This study presented an extended version of DeepPhe-Viz, a platform for visualizing deep phenotype to identify cancer cohort visualization using visual analytics. DeepPhe-Viz is designed to help translation cancer researchers perform retrospective cohort studies and identify

a cohort of cancer patients using EMR. In this dissertation work, we aggregated the mentions of different clinical terms in high-level representation to summarize the substantial contents of different clinical documents throughout cancer treatment for all the patients in the cohort. We utilized the heatmap layout to view the high-level groups and the change in distribution throughout cancer care. We also presented temporal aggregation of all the episodes of care or all the patients in the cohort in the form of a sequence of states ordered in time using the adopted Sankey diagram.

We evaluated the DeepPhe-Viz with the new additional functionalities in term of task completion time, correctness, usability, and qualitative user feedback of the system. All the results from the evaluation study confirmed the necessity and the usability of this system to facilitate the process of using EMR data to conduct cancer retrospective studies and that the DeepPhe-Viz system offers a significant step forward to automate the process of manually reviews EMR data to answer researcher questions and investigate several hypotheses about cancer diagnosis, treatment, and prognosis. We were able to achieve the three aims we set out to achieve in this project.

Appendix A

Informational Script

The purpose of this research is to determine the usability of visual analytics tool being developed to identify cohort of cancer patients using electronic medical records. The goal of this study is to acquire feedback about the current usability and determine the usefulness of the system. You will be asked to perform a set of tasks that will evaluate the performance of the visual analytics tool by measuring time completion and correction of the each performed task. After you finish the tasks, you will be asked to answer a survey questionnaire about your experience using the tool. Your help will benefit us to improve our system and better serve cancer research needs.

Participation will occur either in person or via web-conference, and should take approximately one hour during the session, you will work through a set of tasks, during which you will be asked to think-aloud for some of the tasks. We will record the contents of your screen and audio via your microphone.

- The possible risks associated with this project are frustration due to using a system that is not fully developed, possibly fatigue and the infrequent risk of breach of confidentiality as well. Although every reasonable effort has been taken, confidentiality during internet communication activities cannot be guaranteed and it is possible that additional information beyond that collected for research purposes may be captured and used by others not associated with this study.

- Your participation is voluntary so you may withdraw from the study at any time should you experience undue frustration. If you choose to withdraw from this study, all data collected prior to the date of withdrawal will be continued to be used. All responses are confidential and will be stored securely. All data including recordings will be maintained so that your identity is anonymous, and you will not be identifiable in any way. Authorized representatives of the University of Pittsburgh Office of Research Protections may review your identifiable research information for the purpose of monitoring the appropriate conduct of this research study. Your current and future status with the University or any other benefits to which you are entitled will be the same whether you participate in this study or not.
- This study is funded by the National Cancer Institute. Due to funding being provided by the NCI/NIH, this research is covered by a Certificate of Confidentiality from the National Institutes of Health. have agreed. This protection includes federal, state, or local civil, criminal, administrative, legislative, or other proceedings. An example would be a court subpoena. There are some important things that you need to know. The Certificate does not stop reporting that federal, state, or local laws require. Some examples are laws that require reporting of child or elder abuse, some communicable diseases, and threats to harm yourself or others. The Certificate cannot be used to stop a sponsoring United States federal or state government agency from checking records or evaluating programs. The Certificate also does not prevent your information from being used for other research if allowed by federal regulations. This means that the researchers cannot release or use information, documents, or samples that may identify you in any action or suit unless you say it is okay. They also cannot provide them as evidence unless you.

- I understand that I may always request that my questions, concerns, or complaints be addressed by a listed investigator. This study is being conducted by Dr. Harry Hochheiser, who can be reached by email harryh@pitt.edu or by phone at 412 648 9300 and Saja Al-alawneh email saa144@pitt.edu or phone number 412-944-8646 if you have any questions. You may contact the Human Subjects Protection Advocate of the Human Research Protections Office, University of Pittsburgh (1-866-212-2668) to discuss problems, concerns, and questions; obtain information; offer input; or discuss situations if the research team is unavailable.

User Recruitment Script

The Department of Biomedical Informatics at University of Pittsburgh is conducting a user research study about visual analytics tool. This program allows users to visualize data regarding cancer patient records, both in aggregate, through displays of data summarizing the distribution of data in a cohort of patients, and in individual patient records. Our goal in this study will be to assess the usability of the system.

We are looking for 12 volunteers to spend approximately 60 minutes working with a web-based tool. The participants will be asked first to explore the tool and to "think aloud" as they are working. We will capture information from the computer as they work. Then the participants will be asked to work through a set of tasks to identify a cohort of cancer patients with specific characteristics. At the end of the study, the participants need to fill out a survey. We will capture information from the computer as they work. All sessions will be conducted via web conference or in-person, which will be recorded. All information will be kept confidential

and will only be used to assist us in improving the program. There is no benefit to you for participation in this program. participation is voluntary

If you are interested in participating or have any questions, please contact Dr. Harry Hochheiser at 412 648 9300, email harryh@pitt.edu or Saja Al-alawneh at 412-944-8646, email saa147@pitt.edu.

Sincerely,

Saja Al-alawneh

Department of Biomedical Informatics

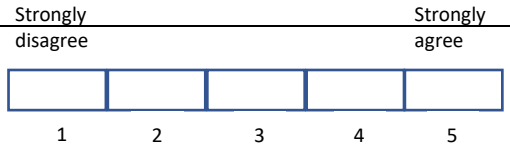
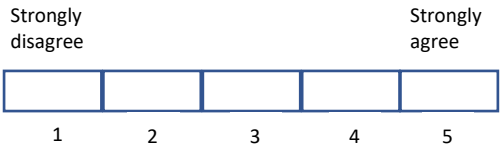
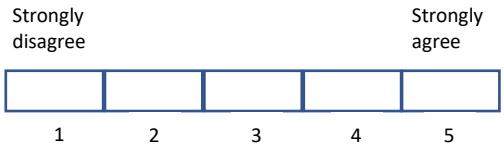
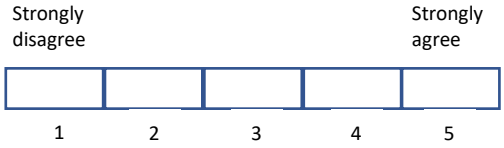
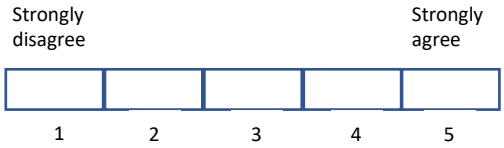
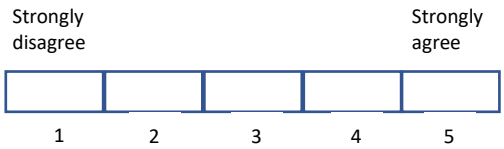
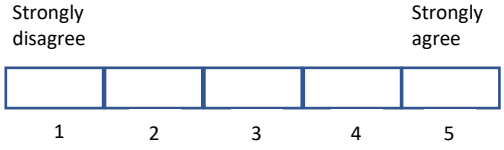
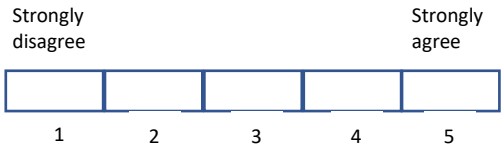
Task Answers

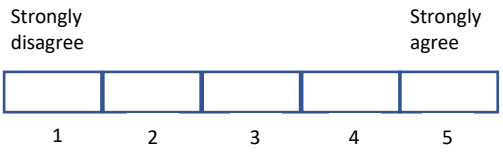
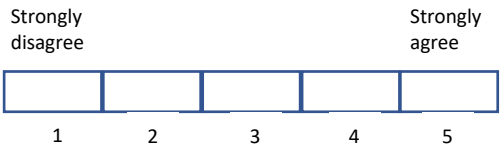
Task	Answers				
Task 1	49 patients				
Task 2	20-59				
Task 3	Biomarker distributions75.5%				
Task 4	Months: 1-101				
Task 5	9 patients patient93, patient47, patient 42, patient 41, patient40, patient17, patient16, patient04, patient01				
Task 6	17 patients patient26, patient 46, patient 45, patient06, patient28, patient17, patient16, patient47, patient30, patient21, patient36, patient32, patient24, patient04, patient10, patient43, patient37				
Task 7	5 patients patient92, patient21, patient37, patient24, patient44				
Task 8	Stage: 2B patient93				
Task 9	100% ER +				
Task 10	month 18 Label: other patinet02	month 1 Label: other patient41, patient17 patient35, patient36	month 5 Label: other patient45, patient04 patient42, patient27	month 10 or 1 Label: other patient29	month 3 Label: other patient46
Task 11	38 patients				
Task 12	22 patients patient92, patient46, patient36, patient40, patient93, patient17, patient38, patienr47, patient34, patient10, patient07, patient27, patient14, patient42, patient20, patient31, patient21, patient04, patient41, patient24, patient35, patient33				
Task 13	7 patients patient41, patient40, patient05, patient27, patient45, patient43, patient01				
Task 14	3 patients				

	patient05, patient29, patient14	patient27, patient04, patient44	patient10, patient42, patient34
Task 15	21 patients patient44, patient10, patient07, patient39, patient13, patient04, patient14, patient02, patient20, patient23, patient21, patient48, patient25, patient29, patient30, patient37, patient18, patient36, patient32, patient38, patient15		
Task 16	patient32 Treatment: Segmental Mastectomy	patient44 Treatment: Lumpectomy and Segmental Mastectomy	
Task 17	patient02 one document clinical note		
Task 18	patient16 3 documents		

SUS questionnaire

System usability scale questionnaire

I think that I would like to use this system frequently	
I found the system unnecessarily complex	
I thought the system was easy to use.	
There is information that I need but cannot find in the system.	
I found the various functions in this system were well integrated.	
I felt very confident using the system	
I think that the system provides the information that I need.	
I found the system very cumbersome to use	

I think that the system is beneficial.	
I could use the system without having to learn anything new.	

Open ending questions

- What worked well about the tool?
- What did not work well?
- Which tasks did you find particularly easy?
    - Why?
- Which tasks were challenging?
    - Why?
- What ways could the system be improved to better meet your needs?
- Which features of the system you think it was the most useful?
    - Why?
- Which features of the system you think it was the least?
    - Why?
- Is there information you would like, but cannot find in the system?
    - If yes, what are they?

Appendix B

Time completion (Row data)

T	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12
T1	113.002	29.295	36.99	52.971	14.592	18.874	62.013	29.162	13.266	34.601	77.418	27.122
T2	105.301	35.794	22.172	32.833	25.44	18.959	58.206	21.688	96.009	17.835	46.938	33.672
T3	82.768	36.211	16.446	9.171	17.132	20.53	44.447	15.54	10.776	14.427	26.965	23.892
T4	98.086	10.846	12.811	17.546	16.744	29.698	73.849	54.48	51.408	20.09	40.028	58.171
T5	146.911	72.664	67.208	41.161	29.947	34.026	87.655	40.434	40.552	29.319	68.516	40.822
T6	157.512	69.523	23.744	35.565	93.526	86.875	79.184	54.941	37.438	36.487	79.679	73.18
T7	301.595	124.371	48.581	72.654	96.018	78.454	89.171	93.688	75.621	67.94	146.613	102.131
T8	361.27	89.289	39.719	32.429	28.526	101.467	105.261	58.672	54.973	42.665	160.74	57.655
T9	127.402	41.181	26.417	37.022	86.791	27.991	178.153	12.965	23.223	38.963	43.819	16.54
T10	205.72	284.166	129.046	74.502	115.163	130.495	125.244	79.493	86.068	112.539	111.191	120.742
T11	248.915	78.544	54.376	58.079	100.423	178.775	122.752	56.788	59.91	66.218	184.543	83.402
T12	174.844	163.628	23.714	50.225	30.607	45.955	82.736	59.461	27.97	31.089	120.403	27.405
T13	132.883	113.47	28.875	42.577	95.468	57.312	114.92	74.855	38.64	100.315	111.49	46.007
T14	162.901	121.754	60.865	127.044	69.103	47.415	108.383	57.021	67.7	47.745	107.022	99.356
T15	170.464	115.917	21.62	50.176	93.862	64.406	60.126	43.176	40.313	59.315	119.781	64.427
T16	265.678	178.217	196.195	103.897	138.644	118.115	97.166	90.618	68.968	45.071	165.545	157.372
T17	350.474	152.166	228.379	85.281	124.791	182.324	142.417	60.409	116.076	53.25	186.024	96.341
T18	435.955	172.401	99.481	78.065	108.442	104.241	156.313	111.098	84.046	79.406	176.195	92.697

Time completion summary for all the participants data (in seconds):

Min	1st Quarter	Median	Mean	3rd Quarter	Max
9.171	38.340	69.035	84.197	111.266	435.955

- Shapiro-Wilk normality test:

Using Shapiro test, if p-value <0.05 then the hypothesis is rejected, and the distribution is not normally distributed

W = 0.83292, p-value = 1.687e-14

Time completion data for the clinician participants

- Summary (in seconds):

Min	1st Quarter	Median	Mean	3rd Quarter	Max
9.171	52.285	90.993	106.613	135.267	435.955

- Shapiro-Wilk normality test

W = 0.85615, p-value = 7.744e-09

Time completion data for the researcher participants

- Summary (in seconds):

Min	1st Quarter	Median	Mean	3rd Quarter	Max
10.78	29.09	54.43	61.78	86.81	228.38

- Shapiro-Wilk normality test

W = 0.8879, p-value = 1.692e-07

High Level Tasks

Task A (T1, T2, T3, T4),

Task B (T5, T6, T7),

Task C (T8, T9, T10, T11),

Task D (T12, T13, T15),

Task E (T16 T17, T18)

Time completion summary for high-level tasks data (in seconds):

- Summary:

Min	1st Quarter	Median	Mean	3rd Quarter	Max
73.91	176.95	252.78	303.11	396.71	1052.11

- Shapiro-Wilk normality test

W = 0.84444, p-value = 2.096e-06

High-level time completion data for the clinician participants

- Summary

Min	1st Quarter	Median	Mean	3rd Quarter	Max
112.1	237.5	320.6	383.8	502.2	1052.1

- Shapiro-Wilk normality test

W = 0.88089, p-value = 0.002956

High-level time completion data for the researcher participants

- Summary

Min	1st Quarter	Median	Mean	3rd Quarter	Max
73.91	143.05	211.50	222.42	261.69	524.05

- Shapiro-Wilk normality test

W = 0.92859, p-value = 0.04504

Ad-Hoc analysis using repeated measure ANOVA for the high-level tasks and participants type

- Check for extreme outliers

Task	group	id	value	Outlier	Extreme outlier
A	Clinician	1	399	TRUE	FALSE
A	Research	11	171	TRUE	FALSE
B	Clinician	1	606	TRUE	TRUE
C	Clinician	1	943	TRUE	FALSE
C	Research	9	439	TRUE	FALSE
E	Clinician	1	1052	TRUE	TRUE

- Overall participant type effect

Task	Group 1	Group 2	N1	N2	statistic	p-value adj.
A	Clinician	Research	6	6	2.07	0.093
B	Clinician	Research	6	6	1.73	0.145
C	Clinician	Research	6	6	1.61	0.168
D	Clinician	Research	6	6	2.68	0.044
E	Clinician	Research	6	6	2.05	0.096

- Overall Task Type Effect

Group 1	Group 2	N1	N2	statistic	p-value adj.
A	B	12	12	-4.51	0.009
A	C	12	12	-5.6	0.002
A	D	12	12	-5.25	0.003
A	E	12	12	-5.68	0.001
B	C	12	12	-5.26	0.003
B	D	12	12	-3.72	0.034
B	E	12	12	-5.44	0.002
C	D	12	12	2.35	0.386
C	E	12	12	-1.25	1
D	E	12	12	-2.55	0.268

Group	Group 1	Group 2	N1	N2	statistic	p-value adj.
Clinician	A	B	6	6	-3.36	0.02
Clinician	A	C	6	6	-4.32	0.076
Clinician	A	D	6	6	-4.65	0.056
Clinician	A	E	6	6	-4.04	0.099
Clinician	B	C	6	6	-4.13	0.091
Clinician	B	D	6	6	-3.39	0.194
Clinician	B	E	6	6	-4.30	0.077
Clinician	C	D	6	6	1.39	1
Clinician	C	E	6	6	-0.686	1
Clinician	D	E	6	6	-1.56	1
Research	A	B	6	6	-2.92	0.329
Research	A	C	6	6	-4.02	0.101
Research	A	D	6	6	-3.55	0.163
Research	A	E	6	6	-4.00	0.103
Research	B	C	6	6	-3.77	0.13

Research	B	D	6	6	-2.57	0.498
Research	B	E	6	6	-3.24	0.23
Research	C	D	6	6	1.98	1
Research	C	E	6	6	-0.992	1
Research	D	E	6	6	-1.90	1

Task correctness data

T	U1	U2	U3	U4	U5	U6	U7	U8	U9	U10	U11	U12
T1	49	49	49	49	49	49	49	49	49	49	49	49
T2	20-59	20-59	20-59	20-59	20-59	20-59	20-59	20-59	20-59	20-59	20-59	20-59
T3	75.5%	75.5%	75.5%	75.5%	75.5%	75.5%	75.5%	75.5%	75.5%	75.5%	75.5%	75.5%
T4	1-101	1-101	1-101	1-101	1-101	1-101	1-101	1-101	1-101	1-101	1-101	1-101
T5	9 P93	9 P47	9 P93	9 P93	9	9 P93	9 P93	9 P93	9 P93	9 P93	9 P93	9 P93
T6	17 P26	17 P47	17 P26	17 P26	17 P26	17 P26	17 P26	17 P26	17 P26	17 P26	17 P26	17 P26
T7	5 P37	5 P44	5 P92	5 P21	16	5 P92	5 P48	5 P21	5 P92	5 P92	5 P44	5 P92
T8	2B P93	2B P93	3B P93	2B P93	0 P15	0 P15	2B P93	2B P93	2B P93	2B P93	2B P93	2B P93
T9	ER 100%	ER 100%	ER	ER	ER	ER 100%	ER 100%	ER 100%	ER 100%	ER 100%	ER 100%	ER 100%
T10	3 other	3 other	3 other	3 other	1 other	3 other	3 other	1 other	1 Other	3 other	3 other	3 other
T11	38	38	38	38	38	38	38	38	38	38	38	38
T12	22 P92	P38	22 P92	6 P92	22 P92	22 P92	22 P92	22 P92	P92	22 P92	22 P92	22 P92
T13	7 P41	7 P41	7 P41	7 P41	7 P41	7 P41	3 P10	7 P41	7 P41	3 P10	7 P41	7 P41
T14	3 P05	3 P27	3 P27	3 P27	3 P27	3 P27	3 P10	3 P05	3 P10	3 P10	3 P05	3 P05
T15	21 P44	21 P44	21 P44	21	21 P44	21 P44	21 P25	21 P44	21 P44	21 P44	21 P44	21 P44
T16	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	Segmental mastectomy	lumpectomy	Segmental mastectomy
T17	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note	1 Clinical note
T18	3	3	3	3	3	3	3	3	3	3	3	3

SUS score for each question

Question	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	User 11	User 12
Q1	4	4	4	3	4	4	4	4	4	3	3	3
Q2	1	0	1	0	0	0	1	0	0	1	1	1
Q3	2	0	1	1	0	0	0	1	2	1	1	1
Q4	2	0	1	0	0	0	0	0	1	1	1	1
Q5	3	3	3	2	1	1	1	2	1	4	3	4
Q6	3	4	4	4	4	4	4	4	4	3	3	4
Q7	3	4	3	4	4	4	3	4	4	3	3	3
Q8	2	3	3	4	4	4	4	4	3	3	3	3
Q9	3	4	4	4	4	4	4	4	4	4	4	4
Q10	3	4	4	4	4	4	4	4	4	3	3	3
Total	26	26	28	26	25	25	25	27	27	26	25	27
SUS score	65	65	70	65	62.5	62.5	62.5	67.5	67.5	65	62.5	67.5

Strongly agree 5	Somewhat agree 4	Neither agree/disagree 3	somewhat disagree 2	strongly disagree 1
To calculate the SUS score, from each user, each question's score contribution will range from 0 to 4. For questions 1,3,5,7, and 9 the score contribution is the scale position minus 1. For questions 2,4,6,8 and 10, the contribution is 5 minus the scale position. Multiply the sum of the scores by 2.5 to obtain the overall value of SUS.				

- Summary of SUS scores for all participants

Min	1st Quarter	Median	Mean	3rd Quarter	Max
62.50	62.50	65	65.21	67.5	70

- Shapiro-Wilk normality test

W = 0.87705, p-value = 0.08034

- Summary of SUS scores for clinician participants

Min	1st Quarter	Median	Mean	3rd Quarter	Max
62.50	63.12	66.25	65.42	67.5	67.5

- Shapiro-Wilk normality test

W = 0.77516, p-value = 0.03473

- Summary of SUS scores for researcher participants

Min	1st Quarter	Median	Mean	3rd Quarter	Max
62.50	63.12	65	65	65	70

- Shapiro-Wilk normality test

W = 0.81374, p-value = 0.07784

- Welch Two Sample t-test

Data: value by group

t = 0.27735, df = 9.8854, p-value = 0.7872

alternative hypothesis: true difference in means is not equal to 0, 95 percent confidence interval:

[-2.9, 3.7]

sample estimates:

Clinician (mean)	Research (mean)
65.41667	65.00000

Open ended questions answer for all the participants

Question	Participant type	Answer
What worked well about the tool?	Clinician	I liked how all of the diagrams were interconnected and you could use this functionality to really drill down to your cohort of interest.
	Researcher	easy to operate, and the figures are clear
		This tool provides a wealth of information that allows us to obtain clinically desirable data on a large number of patients in a fraction of the time it would have taken to perform this manually other positive points: color scheme was bright, and contrast of colors groups was clear very interactive
	Clinician	Very interactive and easy to use, even as a first timer.
	Researcher	Very easy to click and scroll. Well visualized to understand at a glance
	Researcher	The interface is simple to use and the quick sorting of cohorts as parameters were narrowed allowed me to rapidly find useful information related to the patients I was searching for. It is very functional and well thought out.
	Clinician	easy to filter for patient data you are interested in
	Researcher	user friendly
	Researcher	honing in on the patient charts with desired characteristics without manually reviewing them
	Researcher	Simple, interactive
	Clinician	The ability to parse patients based on diagnosis, encounter type is very useful.
	Clinician	Nice interface, easy to read and point-and-click format is very user friendly. A lot of high-level clinical information is available and easy to see the pattern of the timing of care. Notes were all assembled in the patient view so more granular information about treatment can be easily seen.
What did not work well?	Clinician	Counting the number of diagnoses was difficult in the example. I would prefer there be a way you just click on the diagnosis of interest and then it would give you a number.
	Researcher	annotation of some of the terms

		would be helpful to separate and filter patients based on biomarker status - this will be key in several different kinds of cancer (for example mutation status in melanoma but more importantly in being able to distinguish ER positive breast cancers or HER2 positive breast cancers or Triple negative patients) the later figures are very sophisticated and will benefit from labels and or an explanation that may appear by hovering over the graph
	Clinician	None
	Researcher	I think it will need some instruction and brief explanation for each chart and how to use for the first-time users.
	Researcher	Little - I think this is a very good system at the current time.
	Clinician	Figure 5. - I would be curious to know what the "other" terms are
	Researcher	None
	Researcher	None
	Researcher	mainly the missing information that I still have to extract. Also, a way to perform basic plots (histograms) and statistics. Also, an excel export function would be useful
	Clinician	I struggled some with the advanced level questions as I hadn't used the system before. I thought the 'diagnosis type' as in the different types of cancer scrolling model was a little clunky.
	Clinician	Would have liked to see more details about other biomarkers for patients, particularly next generation sequencing or other diagnostic tests. Also, may be helpful to see if they were in clinical trials and when the clinical trial encounters were.
Which tasks did you find particularly easy? Why?	Clinician	Selecting the age brackets for your cohort of interest. The sliding bar facilitated this.
	Researcher	patients filter
		biomarker status was very clear generally very user-friendly once instructions explained age brackets were fun to move
	Clinician	Task 5 was particularly easy. I think the filtering was done very well and was intuitive to use.
	Researcher	Finding the number of patients for a certain category was easy. It is because the finding number changes when I click some categories.
	Researcher	Almost all of the tasks were intuitive and easy to sort through. Finding the patients by sorted Stage Cohorts really helped narrow a larger field to find what I was looking for.
	Clinician	filtering stage and age of diagnosis
	Researcher	having the visual summary and keys

	Researcher	identifying patients by stage and other basic characteristics (such as age)
	Researcher	all of them pretty much
	Clinician	Basic demographics was easy. Simple clicks to obvious places.
	Clinician	Finding the number of patients in a given cohort, because the information was displayed quickly and did not require many clicks to access.
Which tasks were challenging? Why?	Clinician	The episode transition diagram was useful, but it was sometimes hard to be able to focus your mouse on a specific path you were looking for.
	Researcher	None
		the tasks related to figure 6, maybe because there are many overlapping lines but generally doable
	Clinician	Task 1 was challenging because for Figure 2, the chart was a little confusing (cramped, too much information on one chart).
	Researcher	Looking at patient level view and understanding it was not easy.
	Researcher	The only difficult task was utilizing the map to transition between Encounter types. Only because I have never used one similar before and it was a new process for me. I think once I got more experience with the task it too would become intuitive.
	Clinician	nothing was particularly challenging
	Researcher	None
	Researcher	identifying patient status change (pre-diagnosis, diagnosis). you have to scan the whole figure to find the information you need
	Researcher	the sequence tasks take a bit more concentration but are not challenging per se
	Clinician	Utilizing the heat mat took a little while because I've never used it before
	Clinician	Multistep procedures to find patients with several specific characteristics, mainly because it required multiple checks to make sure I had selected the correct parameters.
What ways could the system be improved to better meet your needs?	Clinician	It would be great to link the patient case files to actual samples we have in the tissue bank. I would also appreciate a way to save your patient files/searches you perform in this app. This way, you don't need to recreate this every time. Also, include an export feature next to each diagram in case you want to use these figures.
	Researcher	1. think need better annotations of the terms used in the system
	Clinician	being able to filter by receptor status (ER, PR, HER2) as mentioned above being able to separate breast cancer patients by histology (ductal, lobular,

		metaplastic etc.) label last figure to indicate this shows patients with concomitant other cancers incorporate menopause status as a filterable option
	Clinician	None
	Researcher	Like the questions I solved, the example questions and answers (how to find it) will be helpful. This tool needs to include more types of clinical data, such as recurrence free time, overall survival, ER/PR/HER2 expression level, menopausal status, PAM50 types (if possible), and others.
	Researcher	It would fit my needs at the current time, but I think as the patient pool becomes more robust it will continue to grow in allowing sorting and saving overall time.
	Clinician	having treatment filters for routine dosing of chemo or radiation treatment so timing of dosing/completion can be found easily for each patient
	Researcher	None
	Researcher	I'm not sure--it probably is simplified as much as possible while still being able to reach the information needed
	Researcher	see Q12.
	Clinician	Utilizing a means of searching by diagnosis (neoplasm) or by tumor marker would be nice. I would LOVE to be able to use a search by keyword as shows in the breakdown.
	Clinician	Inclusion of more diagnostic details such as molecular testing or next generation sequencing, inclusion of all chemotherapies received in the patient information section, inclusion of responses to each line of therapy (progression, stable disease, partial response, complete response)
Which features of the system you think it was the most useful? Why?	Clinician	1.) I like the biomarker feature as we often use this to select patients. 2.) The case level view of each patient file was nice to be able to see the abridged clinical notes, which is something you don't typically get to investigate if you are not a clinician.
	Researcher	The Sankey graph of the treatment history. It makes it much easier to follow the treatment of a patient
	Clinician	filtering patients of interest by cancer diagnosis and stage and age at time of diagnosis. It was immensely helpful to then have a list of individual patients where i could then go and dig out more information - all of this in one place at my fingertips
	Clinician	I think the ability to filter patients and then go down to the patient-level data is really helpful to get relevant notes/labs/scans.
	Researcher	Graphical display and showing the number of patients in parenthesis were helpful.

	Researcher	Simple, easy-to-use interface. Even with the data narrowed down, having a simple interface that I could pick up on my first attempt made the system less cumbersome.
	Clinician	selecting a cohort based on stage and age - and biomarkers, if possible
	Researcher	having everything summarized and visual in one page
	Researcher	identifying charts which include certain terms. This would be helpful for retrospective chart reviews
	Researcher	Interactive, visualization
	Clinician	The breakdown of individual patient information was very useful because I could learn about each patient based on the filters that were applied.
	Clinician	The interface is very nice, it is very easy to navigate, and many aspects are intuitive.
Which features of the system you think it was the least? Why?	Clinician	I would have liked the ability to find out more specific treatment information under drug. Right now, the options under label preference are not that specific or helpful for my questions.
	Researcher	None
	Clinician	graph 5 where it indicated prevalence of specific data points. this was helpful to filter information but knowing the prevalence of information available may not significantly change the question i would be asking
	Clinician	None
	Researcher	NA
	Researcher	None
	Clinician	trends, in Figure 6 (from pre-diagnostic to diagnostic, etc.)
	Researcher	None
	Researcher	N/a
	Researcher	everything was useful
	Clinician	n/a
	Clinician	The total number of each type of note does not seem particularly useful to me, usually it is the information within the notes that are important to me but not the number of notes.
Is there information you would like, but cannot find in the system?	Clinician	Yes the ability to pick specific drug treatments. Ex.) palbo, Ais, Tamoxifen, bone targeting agents
	Researcher	Yes more specific treatment of that patient, and the switching between treatment
		Yes

If yes, what are they?		menopause status - which is available but not filterable subtypes of breast cancer - again this information was available for individual patients, but this would be very helpful if we can filter for this information
	Clinician	No None
	Researcher	Yes I like to see more different types of clinical data, such as survival, menopausal status, availability of tissue sample, and others
	Researcher	No None
	Clinician	Yes Itemize treatment history -rounds of chemo or rounds of radiation and dosing for each
	Researcher	Yes More detailed about specific cancer type and some of specific treatment
	Researcher	No None
	Researcher	Yes mentioned previously: labs, other biomarkers (next generation sequencing i.e.), survival, progression-free survival
	Clinician	Yes Ability to search by keywords as uploaded into the system
	Clinician	Yes When patients have next generation sequencing, what the mutations or abnormalities in the NGS are, and what the responses to each treatment that they received was

Bibliography

1. Raghavan P, Fosler-Lussier E, Embi PJ, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *Journal of the American Medical Informatics Association* 2014; 21: 221-230.

2. Xie F, et al. Application of text information extraction system for real-time cancer case identification in an integrated healthcare organization. *Journal of pathology informatics* 2017; 8.

3. Yuan Z, et al. Longitudinal Visual Analytics for Unpacking the Cancer Journey. *bioRxiv* 2018: 444356.

4. Stasko J, Carsten Görg, and Zhicheng Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization* 2008; 7: 118-132.

5. Kreimeyer K, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *Journal of biomedical informatics* 2017; 73: 14-29.

6. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010; 17: 507-513. DOI: 10.1136/jamia.2009.001560.

7. Crowley RS, et al. caTIES: a grid based system for coding and retrieval of surgical pathology reports and tissue specimens in support of translational research. *Journal of the American Medical Informatics Association* 2010; 17: 253-264.

8. Sohn S, Kocher JP, Chute CG, et al. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. *J Am Med Inform Assoc* 2011; 18 Suppl 1: i144-149. DOI: 10.1136/amiajnl-2011-000351.

9. Bejan CA, et al. Pneumonia identification using statistical feature selection. *Journal of the American Medical Informatics Association* 2012; 19: 817-823.

10. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proc AMIA Symp* 2001: 17-21.

11. Friedlin J, Overhage M, Al-Haddad MA, et al. Comparing methods for identifying pancreatic cancer patients using electronic data sources. *AMIA Annu Symp Proc* 2010; 2010: 237-241.

12. Hassanpour S, and Curtis P. Langlotz. Information extraction from multi-institutional radiology reports. *Artificial intelligence in medicine* 2016; 66: 29-39.

13. Newton KM, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *Journal of the American Medical Informatics Association* 2013; 20: e147-e154.

14. Uzuner Ö, et al. Identifying patient smoking status from medical discharge records. *Journal of the American Medical Informatics Association* 2008; 15: 14-24.

15. Rea S, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. *Journal of biomedical informatics* 2012; 45: 763-771.

16. Hochheiser H, Castine M, Harris D, et al. An information model for computable cancer phenotypes. *BMC Med Inform Decis Mak* 2016; 16: 121. DOI: 10.1186/s12911-016-0358-4.

17. Savova GK, et al. DeepPhe: a natural language processing system for extracting cancer phenotypes from clinical records. *Cancer research* 2017; 77: e115-e118.

18. Yuan Z, et al. Interactive Exploration of Longitudinal Cancer Patient Histories Extracted From Clinical Text. *JCO Clinical Cancer Informatics* 2020; 4: 412-420.

19. Doan S, et al. Natural language processing in biomedicine: a unified system architecture overview. *Clinical Bioinformatics* 2014; Humana Press: 275-294.

20. Friedman C, and Stephen B. Johnson. Natural language and text processing in biomedicine. *Biomedical Informatics* 2006.: 312-343.

21. Sager N, Carol Friedman, and Margaret S. Lyman. *Medical language processing: computer management of narrative data.*: Addison-Wesley Longman Publishing Co, 1987.

22. Sager N, et al. Natural language processing and the representation of clinical data. *Journal of the American Medical Informatics Association* 1994; 1: 142-160.

23. Zeng QT GS, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak* 2006; 6.

24. organization ib. HITEx Manual v2.0, https://www.i2b2.org/software/projects/hitex/hitex_manual.html.

25. Yin S, et al. Detecting hotspots in insulin-like growth factors 1 research through metamap and data mining technologies. In: *International Conference on Web Information Systems Engineering* (ed Springer), Berlin, Heidelberg, 2013.

26. Chapman WW, et al. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics* 2001; 34: 301-310.

27. Kreuzthaler M, Stefan Schulz, and Andrea Berghold. Secondary use of electronic health records for building cohort studies through top-down information extraction. *Journal of biomedical informatics* 2015; 53: 188-195.

28. Heintzelman NH, et al. Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text. *Journal of the American Medical Informatics Association* 2013; 20: 898-905.

29. Skeppstedt M, et al. . Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: An annotation and machine learning study. *Journal of biomedical informatics* 2014; 49: 148-158.

30. Botsis T, et al. Secondary use of EHR: data quality issues and informatics opportunities. In: *Summit on Translational Bioinformatics* 2010.

31. Xu H, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. In: *AMIA Annual Symposium Proceedings* (ed Association AMI), 2011.

32. Bill R, et al. Automated extraction of family history information from clinical notes. In: *AMIA Annual Symposium Proceedings* (ed Association AMI), 2014.

33. Lowe HJ, Yang Huang, and Donald P. Regula. Using a statistical natural language parser augmented with the umls specialist lexicon to assign snomed ct codes to anatomic sites and pathologic diagnoses in full text pathology reports. In: *AMIA Annual Symposium Proceedings* (ed American Medical Informatics Association), 2009.

34. Buckley JM, et al. The feasibility of using natural language processing to extract clinical information from breast pathology reports. *Journal of pathology informatics* 2012; 3.

35. Childs LC, et al. Description of a rule-based system for the i2b2 challenge in natural language processing for clinical data. *Journal of the American Medical Informatics Association* 2009; 16: 571-575.

36. D'Avolio LW, and Alex AT Bui. The Clinical Outcomes Assessment Toolkit: A Framework to Support Automated Clinical Records–based Outcomes Assessment and Performance Measurement Research. *Journal of the American Medical Informatics Association* 2008; 15: 333-340.

37. Cui L, et al. EpiDEA: extracting structured epilepsy and seizure information from patient discharge summaries for cohort identification. In: *AMIA Annual Symposium Proceedings* (ed Association AMI), 2012.

38. Sahoo SS, et al. Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. *Journal of the American Medical Informatics Association* 2014; 21: 82-89.

39. Zhang G-Q, et al. MEDCIS: multi-modality epilepsy data capture and integration system. In: *AMIA Annual Symposium Proceedings* (ed Association AMI), 2014.

40. Cunningham H, et al. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS Comput Biol* 2013; 9.

41. Cunningham H, et al. GATE, a general architecture for text engineering. *Computers and the Humanities* 2002; 36: 223-254.

42. Cai T, et al. Natural language processing technologies in radiology research and clinical applications. *Radiographics* 2016; 36: 176-191.

43. Hassanpour S, and Curtis P. Langlotz. Information extraction from multi-institutional radiology reports. *Artificial intelligence in medicine* 2016; 66: 29-39.

44. Xu H, et al. MedEx: a medication information extraction system for clinical narratives. *Journal of the American Medical Informatics Association* 2010; 17: 19-24.

45. L. Christensen PH, M. Fiszman. MPLUS: a probabilistic medical language understanding system. In: *Proceedings of the ACL-02 workshop on Natural language processing in the biomedical domain* (ed Linguistics AfC), Phildadelphia, Pennsylvania, 2002, pp.29–36.

46. Hirsch JS, Tanenbaum JS, Lipsky Gorman S, et al. HARVEST, a longitudinal patient record summarizer. *J Am Med Inform Assoc* 2015; 22: 263-274. DOI: 10.1136/amiajnl-2014-002945.

47. University H. Health NLP Cancer, https://healthnlp.hms.harvard.edu/cancer/wiki/index.php?title=Main_Page (2015).

48. Jiang TT, W. D. Xiao, and C. Zhang. A Standards-based Semantic Metadata Repository to Support EHR-driven Phenotype Authoring and Execution. *Medinfo 86* 2015: 1098.

49. Institute NC. NCI Metathesaurus, https://ncimeta.nci.nih.gov/ncimbrowser/ (2019.).

50. Liu H, et al. An information extraction framework for cohort identification using electronic health records. *AMIA Summits on Translational Science Proceedings* 2013: 149.

51. Horrocks I, et al. SWRL: A semantic web rule language combining OWL and RuleML. *W3C Member submission* 2004; 21: 1-31.

52. Drools D. Business Rules Management System (Java™, Open Source). http://www.drools.org.

53. Nadkarni PM, Lucila Ohno-Machado, and Wendy W. Chapman. Natural language processing: an introduction. *Journal of the American Medical Informatics Association* 2011; 18: 544-551.

54. Kersloot MG, et al. Automated SNOMED CT concept and attribute relationship detection through a web-based implementation of cTAKES. *Journal of Biomedical Semantics* 2019; 10: 14.

55. Wright-Bettner K, et al. Cross-document coreference: An approach to capturing coreference without context. In: *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis* LOUHI, 2019, pp.1-10.

56. Miller T, et al. Towards generalizable entity-centric clinical coreference resolution. *Journal of biomedical informatics* 2017; 69: 251-258.

57. Datta S, Elmer V. Bernstam, and Kirk Roberts. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *Journal of biomedical informatics* 2019; 100.

58. Buckley JM, et al. The feasibility of using natural language processing to extract clinical information from breast pathology reports. *Journal of pathology informatics* 2012; 3.

59. Coden A, et al. . Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model. *ournal of biomedical informatics* 2009; 42: 937-949.

60. Wieneke AE, et al. Validation of natural language processing to extract breast cancer pathology procedures and results. *Journal of pathology informatics* 2015; 6.

61. D'Avolio LW, et al. Facilitating clinical outcomes assessment through the automated identification of quality measures for prostate cancer surgery. *Journal of the American Medical Informatics Association* 2008; 15: 341-348.

62. Caban JJ, and David Gotz. Visual analytics in healthcare–opportunities and research challenges. *J Am Med Inform Assoc* 2015; 22: 260–262.

63. West VL, David Borland, and W. Ed Hammond. Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association* 2015; 22: 330-339.

64. Manning JD, Beatriz E. Marciano, and James J. Cimino. Visualizing the Data–Using Lifelines2 to Gain Insights from Data Drawn from a Clinical Data Repository. In: *AMIA Summits on Translational Science Proceedings* 2013.

65. Cook KA, and James J. Thomas. Illuminating the path: The research and development agenda for visual analytics.: IEEE Computer Society, 2005.

66. Kohlhammer J, et al. Solving Problems with Visual Analytics. *Procedia Computer Science* 2011; 7: 117-120.

67. Lee J, Maslove DM and Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *PLoS One* 2015; 10: e0127428. DOI: 10.1371/journal.pone.0127428.

68. Lee J, David M. Maslove, and Joel A. Dubin. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *PloS one* 2015; 10.

69. Caban JJ, and David Gotz. Visual analytics in healthcare–opportunities and research challenges. In: *AMIA* 2015, pp.260-262.

70. Keim D, et al. *Mastering the Information Age. Solving Problems with Visual Analytics,*. 2010.

71. Bernard J, et al. A visual-interactive system for prostate cancer cohort analysis. In: *IEEE computer graphics and applications* 2015, pp.44-55.

72. Bernard J, Sessler D, May T, et al. A visual-interactive system for prostate cancer cohort analysis. *IEEE Comput Graph Appl* 2015; 35: 44-55. DOI: 10.1109/MCG.2015.49.

73. Tanenbaum JS, Lipsky Gorman S, Liu C, et al. HARVEST, a longitudinal patient record summarizer. *Journal of the American Medical Informatics Association* 2015; 22: 263-274.

74. Malik S, et al. An evaluation of visual analytics approaches to comparing cohorts of event sequences. *EHRVis Workshop on Visualizing Electronic Health Record Data at VIS*. 2014.

75. Monroe M, et al. EventFlow: Exploring Point and Interval Event Temporal Patterns. *Human Computer Interaction Lab* 2013; 23.

76. Wongsuphasawat K, and David Gotz. Outflow: Visualizing patient flow by symptoms and outcome. In: *IEEE VisWeek Workshop on Visual Analytics in Healthcare* (ed Association AMI), Providence, Rhode Island, 2011.

77. Guo S, Xu K, Zhao R, et al. EventThread: Visual Summarization and Stage Analysis of Event Sequence Data. *IEEE Trans Vis Comput Graph* 2018; 24: 56-65. DOI: 10.1109/TVCG.2017.2745320.

78. Zhang Z, David Gotz, and Adam Perer. Iterative cohort analysis and exploration. *Information Visualization* 2015; 14: 289-307.

79. Rogers J, Spina N, Neese A, et al. Composer-Visual Cohort Analysis of Patient Outcomes. *Appl Clin Inform* 2019; 10: 278-285. DOI: 10.1055/s-0039-1687862.

80. Krause J, Perer A and Stavropoulos H. Supporting Iterative Cohort Construction with Visual Temporal Queries. *IEEE Trans Vis Comput Graph* 2016; 22: 91-100. DOI: 10.1109/TVCG.2015.2467622.

81. Gotz D and Stavropoulos H. DecisionFlow: Visual Analytics for High-Dimensional Temporal Event Sequence Data. *IEEE Trans Vis Comput Graph* 2014; 20: 1783-1792. DOI: 10.1109/TVCG.2014.2346682.

82. Guo S, et al. Visual progression analysis of event sequence data. *IEEE transactions on visualization and computer graphics* 2018; 25: 417-426.

83. Wongsuphasawat K, et al. LifeFlow: Visualizing an Overview of Event Sequences. In: *Proceedings of the SIGCHI conference on human factors in computing systems* 2011.

84. Wang TD, et a. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In: *Proceedings of the SIGCHI conference on Human factors in computing systems* 2008.

85. Loorak MH, et al. Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients. *IEEE transactions on visualization and computer graphics* 2015; 22: 409-418.

86. J.Bertin. *Readings in information visualization*. San Francisco, CA, USA: Morgan Kaufmann Inc., 1999.

87. Lins L, et al. Viscaretrails: Visualizing trails in the electronic health record with timed word trees, a pancreas cancer use case. *Workshop on Visual Analytics in Healthcare*. 2011.

88. Liu S, et al. StoryFlow: Tracking the Evolution of Stories. *IEEE Transactions on Visualization and Computer Graphics* 2013; 19: 2436-2445.

89. Law P-M, et al. Maqui: Interweaving queries and pattern mining for recursive event sequence exploration. *IEEE transactions on visualization and computer graphics* 2018; 25: 396-406.

90. Jin Z, et al. Carepre: An intelligent clinical decision assistance system. *ACM Transactions on Computing for Healthcare* 2020; 1: 1-20.

91. Froehlich P. Interactive Sankey Diagrams. In: *IEEE Symp on Information Visualization* 2005.

92. Perer A, and Fei Wang. Frequence: Interactive Mining and Visualization of Temporal Frequent Event Sequences. In: *Proceedings of the 19th international conference on Intelligent User Interfaces* 2014, pp.153-162.

93. team P. JavaScript, https://www.javascript.com/ (2021).

94. Bostock. M. Data-Driven Documents, https://d3js.org/ (2020).

95. Schmidt M. Der Einsatz von sankey-diagrammen im stoffstrommanagement. *Beiträge der Hochschule Pforzheim* 2006; 124.

96. Corbin J, and Anselm Strauss. *Basics of qualitative research: Techniques and procedures for developing grounded theory.*: Sage publications, 2014.

97. Jupyter P. Python Jupiter notebook (2021).

98. Team RC. R: A language and environment for statistical computing. Version 4.0. 2, (2020).

99. Brooke J. SUS: a "quick and dirty'usability. *Usability evaluation in industry* 1996; 189.

100. Thayer A, et al. Content analysis as a best practice in technical communication research. *Journal of Technical Writing and Communication* 2007; 37: 267-279.

101. Brooke J. *SUS: a "quick and dirty'usability*. Usability evaluation in industry, 1996.

102. A. Bangor PTK, and J.T. Miller. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 2009; 4: 114-123.

103. Jiang T. Exploratory search: a critical analysis of the theoretical foundations, system features, and research trends. *Library and Information Sciences* 2014; Springer, Berlin, Heidelberg: 79-103.

104. Defossez G, Rollet A, Dameron O, et al. Temporal representation of care trajectories of cancer patients using data from a regional information system: an application in breast cancer. *BMC Med Inform Decis Mak* 2014; 14: 24. DOI: 10.1186/1472-6947-14-24.

105. Chowdhury A, and Joshua Alspector. Data duplication: an imbalance problem?. In: *ICML'2003 Workshop on Learning from Imbalanced Data Sets (II)* Washington, DC, 2003.

106. Schofield A, Laure Thompson, and David Mimno. Quantifying the effects of text duplication on semantic models. In: *Empirical Methods in Natural Language Processing* 2017, pp.2737-2747.

107. Raghavan P, et a. Cross-narrative temporal ordering of medical events. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* 2014, pp.998-1008.

108. Cherry C, et al. A la Recherche du Temps Perdu: extracting temporal relations from medical text in the 2012 i2b2 NLP challenge. *Journal of the American Medical Informatics Association* 2013; 20: 843-848.

109. Pivovarov R, and Noémie Elhadad. Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association* 2015; 22: 938-947.