THALIS: Human-Machine Analysis of Longitudinal Symptoms in Cancer Therapy

Carla Floricel, Nafiul Nipu, Mikayla Biggs, Andrew Wentzel, Guadalupe Canahuate,
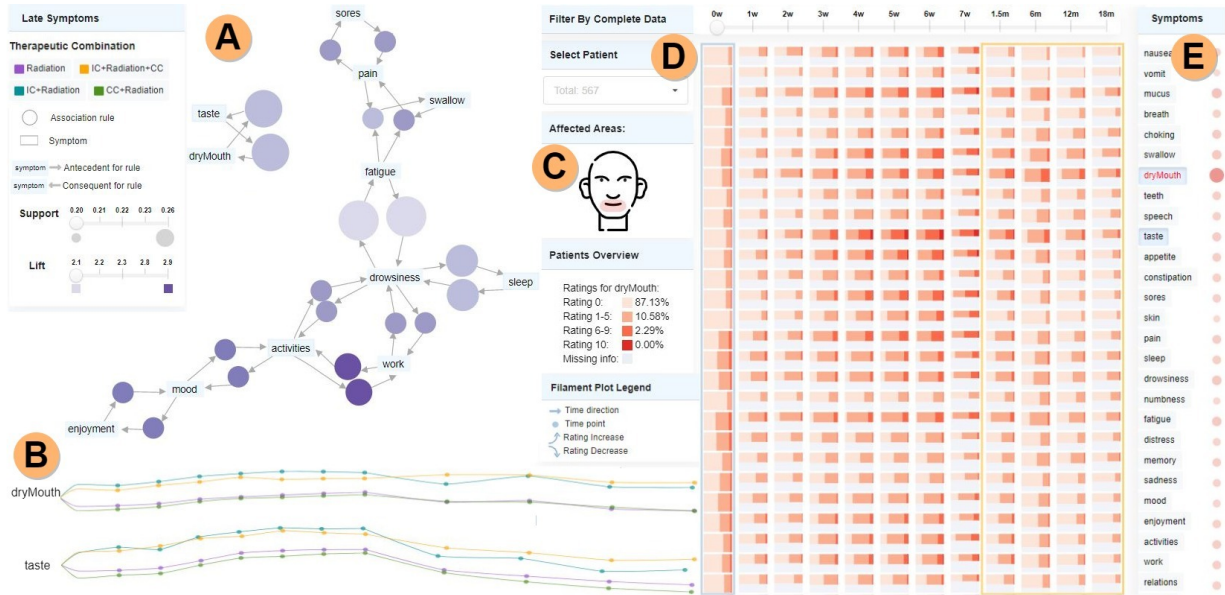Lisanne Van Dijk, Abdallah Mohamed, C.David Fuller, and G.Elisabeta Marai

Fig. 1. THALIS analysis of longitudinal symptom data. A) Association Rule Diagram panel, showing here association-rule-mining (ARM) relationships among the most frequent late-stage symptoms; rules are represented using bubbles, with size and color encoding the support and lift metrics. B) Symptom trajectory panel—filament plots encode here mean rating values per therapeutic combination, with more frequent observations in the acute stage (left-end) than in the late stage (right-end). C) Sketch of areas affected by the selected symptoms (dry mouth and taste). D) Cohort symptom panel showing via summarization with shade and height the percentile rating distribution. E) Correlation matrix showing associations with the selected symptom. Image cropped and edited for in-print legibility.

Abstract—Although cancer patients survive years after oncologic therapy, they are plagued with long-lasting or permanent residual symptoms, whose severity, rate of development, and resolution after treatment vary largely between survivors. The analysis and interpretation of symptoms is complicated by their partial co-occurrence, variability across populations and across time, and, in the case of cancers that use radiotherapy, by further symptom dependency on the tumor location and prescribed treatment. We describe THALIS, an environment for visual analysis and knowledge discovery from cancer therapy symptom data, developed in close collaboration with oncology experts. Our approach leverages unsupervised machine learning methodology over cohorts of patients, and, in conjunction with custom visual encodings and interactions, provides context for new patients based on patients with similar diagnostic features and symptom evolution. We evaluate this approach on data collected from a cohort of head and neck cancer patients. Feedback from our clinician collaborators indicates that THALIS supports knowledge discovery beyond the limits of machines or humans alone, and that it serves as a valuable tool in both the clinic and symptom research.

Index Terms—Temporal Data; Application Motivated Visualization; Life Sciences; Mixed Initiative Human-Machine Analysis

✦

1 INTRODUCTION

Thanks to advances in therapeutic care, nowadays cancer patients may survive for years after treatment. However, they are plagued with

---
- *C. Floricel, N. Nipu, A. Wentzel, and G.E. Marai are at the University of Illinois at Chicago. E-mail: cflori3@uic.edu.*
- *M. Biggs and G. Canahuate are at the University of Iowa.*
- *L.V. Dijk, A. Mohamed, C.D. Fuller are with the MD Anderson Cancer Center at the University of Texas.*

long-lasting or permanent residual sequelae, whose severity, rate of development, and resolution after treatment vary largely between survivors [16, 97, 98]. At the same time, patient questionnaires and electronic health records storing such patient responses are leading to larger than ever oncological symptom data collections. These symptom data collected from cohorts of patients [84] offer important information that can improve clinical decision-making and individual care delivery both during and after treatment [69, 77], and could be critical for the efficient detection and resolution of longitudinal symptoms. These factors have led to healthcare provider demands to better understand symptom development and prevention based on cohort data.

However, the meaningful interpretation at the individual patient level of symptom repositories is plagued by data and analysis issues that have prevented their practical use in clinical care. These issues include

the wide range of symptoms, their partial co-occurrence, their variability among patients and across time, and, in the case of head and neck cancers (HNC) and other cancers that employ radiation therapy, further symptom dependency on the anatomical location of the tumors and the course of therapy prescribed. To explore these issues, symptom cluster research aims to identify co-occurring symptoms and to understand the underlying mechanisms that drive these clusters, often using machine learning [74, 86]. At the same time, HNC analysis results based on factor analysis (e.g., PCA) do not always scale to larger patient datasets [10]. Furthermore, due to methodological limitations, symptom research analyzes either individual symptom evolution or symptom clusters at a single timepoint. Consequently, there is growing interest in alternative machine learning approaches for this type of longitudinal data. Last but not least, these approaches need to make sense in an applied healthcare setting and need to be actionable by clinicians. Therefore, there is also growing interest in mixed human-machine analysis, and a need to leverage and balance computational and human effort for symptom data analysis.

In this work, we present an interactive data mining environment to support the clustering, exploration, and analysis of longitudinal symptoms collected from cohorts of cancer patients. Our approach intertwines association-rule and factor analysis unsupervised models with custom visual statistical encodings and visual analysis, in order to estimate the longitudinal symptom evolution of an individual patient, in the context of cancer therapies and similar patients. This visual analysis methodology was successfully developed through an interdisciplinary, remote, geographically-distributed collaboration.

This work contributes: 1) a description of the application domain data and tasks, with an emphasis on the multidisciplinary development of clustering tools for symptom data in cancer therapy; 2) the design of a novel blend of data mining and visual encodings to predict and explain longitudinal symptom development, based on an existing cohort of patients; 3) the description of customized interactive encodings: interactive association-rule diagrams, filaments, and percentile heatmaps; 4) an implementation of this approach in a visual symptom explorer named THALIS: THerapy Analysis of LongItudinal Symptoms (Fig. 1); 5) a qualitative evaluation by domain experts using an existing head and neck symptom repository; 6) a start-to-end description of the design process and of the lessons learned from this successful, multi-site remote collaboration.

2 RELATED WORK

Electronic Medical Records and Cohort Visualization. Electronic Medical Records (EMR) store patient longitudinal information, often in the form of time series. In general, time-series visualization has utilized point graphs, circle graphs, line graphs [41], parallel coordinate plots [45], or stacked bar charts and their variations [3] to encode time-oriented nominal, ordinal or quantitative data, including in cancer [76,89,104]. For EMR data, Plaisant et al. have introduced personal patient summary visualization using timelines [80,93,94], or matrix-based representations [23]. Loorak et al. [60] proposed a stacked bar graph approach to explore patients' treatment processes, while Baumgartl et al. [8] explored storyline visualizations from EMR to detect pathogen outbreaks. Rogers et al. [83] showed outcome trajectories of different patient procedures using line charts. However, most of these approaches are not scalable for large EMR datasets. Wong et al. have employed summarization techniques to overcome issues of scale via tree-based encodings [102] and Sankey-based representations [101], while Karpefors's tendril plot [51] introduced a clustered timeline view of outliers and trends for dense clinical trial data. However, none of these approaches include details about individual patients. In contrast, we consider scalable encodings for patient cohort data, and indicate incomplete data and uneven time steps.

In healthcare, patient cohort visual analysis applications span disease evolution statistics extracted from EMRs [42, 101], cohort history comparison [9, 17, 105], inter-cohort medical image attribute comparison [56, 72, 88], survival risk analysis in cancer [69], and cohort heterogeneous medical data analysis [5, 91]. As often the case in clinician-driven visual analysis based on statistics, the visual encod-

ings in these works include conventional representations such as histograms [8], bar charts [60], time-series plots [33,47], matrices [23,66], radial charts [35], and scatterplots [58]. Similarly, our work builds on patient cohort data, however, our focus is on interpreting individual patient data in the context of similar patients, and contributing visual encodings and workflows which improve the human-machine analysis of symptom data.

Human-Machine Integrated Cohort Cluster Analysis. Cohort analysis uses unsupervised learning methods such as factor analysis (e.g. PCA), partitional (e.g. K-means), or hierarchical (e.g. agglomerative) clustering. Cluster analysis is traditionally visualized using methods such as scatterplots [73], matrices [82], radar charts [69], dendrograms [28], and heatmaps [1]. Temporal clustering is an open problem in symptom research due to the issue of missing data [6, 65, 90]. Additionally, cancer patient clustering takes into account clinical variables such as the disease stage, treatment plans, medication, treatment toxicity etc. [64, 103]. For HNC patients, Wentzel et al. have introduced spatially-informed distance measures and clustering approaches to group patients based on similarity [25, 97, 99], although they did not consider symptom data. Gunn et al. [34] and Rosenthal et al. [85] have studied specifically symptom burden for HNC patients by clustering patients based on reported symptom ratings and clinical covariates to find similarities between symptoms and HNC patients using heatmaps and cluster heatmaps, but do not consider temporal data, nor do they analyze patients that underwent specific treatments, respectively. In contrast, our approach explores groups of similar patients based on symptom load, while also capturing temporal changes in their symptoms. Moreover, we consider the impact of different treatment plans.

Cohort analysis often relies on domain expert interaction to help support human-machine integrated workflows. For more general clustering, several interfaces have afforded user interaction for iterative re-clustering and visualization of unstructured cluster data [13, 14], although these rely on generic abstract encodings such as scatterplots. Other tools support model building for biostaticians [20], although these do not consider spatial or temporal outcomes, and are targeted towards statisticians and not clinicians. Angelelli et al. [5] proposed an interactive system for hypothesis generation with retrospective cohort study data using a data-cube-based model that used linked views for spatial and nonspatial data. Other applications have integrated interactive interfaces with application-specific visual encodings with linked views [32, 96, 98] to support active collaboration between data analysts and domain experts. However, none of these approaches consider temporal changes in outcome data or nuanced quality of life outcomes, and do not account for missing data.

Rule Visualization. Association rules have been visualized via scatterplots, matrix views, node-link representations, mosaic plots, and parallel coordinates plots, as indicated by two surveys [12, 46], and also as grouped matrices [39]. More generally, rule-based modeling and visualization are common across domains that seek to understand causality. Colored shapes have been used to indicate information flow in interacting processes [27]. In biological modeling, interactive node-link visual representations have been used for rule-based intracellular biochemistry [31, 87]. Visual causal vectors have been used to indicate causality between data elements [95], and animated causal overlays have been used to highlight causal flows and to indicate the relative strength of the causal effect [7]. Whereas our work seeks to identify temporal relationships among data based on association rules, these relationships are not necessarily causal, and they have different features than biochemical pathways.

3 BACKGROUND

HNC Therapy and Symptom Collection. HNC treatment is a complex, longitudinal process that utilizes a variety of therapies, and whose cornerstone is radiotherapy. For example, patients may be prescribed chemotherapy first (induction therapy), and then radiotherapy, or they may be prescribed both chemotherapy and radiotherapy concomitantly. The type of treatment prescribed can result in both short-term (acute, or during treatment) symptoms and in long-term (late, or after treatment) or even permanent sequelae affecting the patient's quality of life.

In addition to clinical and imaging data [26], continuous efforts at MD Anderson have included over 1000 patients in a standardized symptom and quality of life monitoring program. The questionnaires are collected on paper at discrete time points, i.e., weekly at the time of the treatment appointment. The questionnaires are based on MDASI (MD Anderson Symptom Inventory) [18], a multisymptom patient-reported outcome measure for clinical and research use. MDASI's thirteen core items include symptoms found to have the highest frequency and/or severity in patients with various cancers and treatment types, whereas the additional MDASI-HN inventory [84] considers nine symptoms specific to HNC, such as swallowing difficulties, and six additional symptoms that interfere with major activities of daily life, such as enjoyment of life. The compliance rates within head and neck trials are between 60% and 90%. However, these patient-generated health data have not been utilized so far in direct patient care, due to a lack of computational hybrid analytics connecting therapy with the side effects and health state of the patient.

Symptom Clustering Research. Cancer patients experience multiple co-occurring symptoms often related to each other and to the therapy applied; however, much of symptom clustering research focuses on single symptoms. In contrast, the term "symptom cluster" (SC) denotes two or more interrelated symptoms that develop together and may or may not be caused by the same underlying mechanism. Several studies have identified symptom clusters in cancer patients [4, 21, 29], though symptom cluster research is still an emerging field. The two most common methods used to determine SCs are: factor analysis (e.g., principal component analysis, i.e., PCA) [53, 55, 86] and cluster analysis (e.g., hierarchical agglomerative clustering) [30, 37, 44, 75]. However, these approaches have not dealt with changes in symptoms over time, which remains an elusive goal.

Association Rule Mining (ARM), introduced by Agrawal and Srikant in 1994 [2], is an alternative unsupervised data mining method, used to identify interesting relationships within data. ARM has been applied to risk management and marketing [36, 52], and more recently, in clinical settings [57], although not in symptom clustering.

4 DESIGN

4.1 Collaboration Setting and Design Process

Our system was developed through a remote collaboration between three different research groups over the course of two years. During this collaboration, our visual computing research group worked closely with oncology and data mining experts. The core team includes 3 radiation oncology experts with clinical and research experience, a senior data mining expert, a data-mining graduate student, and a team of visual computing researchers with varying expertise. Our team met weekly to produce informative, mixed machine-human analyses of longitudinal symptom data collected from HNC patients who were undergoing treatment at the MD Anderson Cancer Center in Houston, Texas. This work is part of a longer, six-year-long collaboration between the lead investigators who had been working together on a series of related projects using oncology patient data.

Due to the long-term and remote nature of our collaboration, spanned on three sites, we employed team-science principles [69]. Our design process blended an agile design process based on regular team meetings along with an Activity-Centered-Design (ACD) approach to the design of the visualization system [68]. The ACD paradigm is an extension of human-centered-design, with emphasis on user activities and workflow. We note that in the ACD paradigm, the value of a tool depends on the value of the activity, not only on the number of people who use the tool (e.g., a tool serving the two researchers who will find a cure for Alzheimer's has no lesser value than a tool serving a larger population who are selecting pet names) [70]. Thus, the ACD paradigm is particularly well suited for tools in scientific research, particularly when we consider the scarcity of trained domain experts, as opposed to the large availability of untrained users, and the importance of slow thinking [49], including scaffolding.

Through a series of iterations, the research team met to define functional specifications, prototype the interface, evaluate prototypes, and decide on changes in the specifications. Moreover, because this approach was designed around developing interfaces that can be shared and designed remotely during the COVID-19 pandemic, our approach proved to be an effective alternative to approaches that rely on in-person group meetings. Additionally, because the ACD paradigm is focused on supporting the collaborators' activities, our collaborators stayed motivated to continue to attend meetings even during circumstances that required remote meetings and exceptional work conditions for clinical practitioners [78].

4.2 Activity and Task Analysis

THALIS serves oncologists who have experience in symptom research. Our collaborators also had extensive experience using basic unsupervised machine learning methods such as factor analysis via principal component analysis (PCA), which they had used to determine that symptom burden varies over time and over patient populations. However, PCA results obtained on smaller datasets did not generalize on larger datasets, so over the course of the project, the group's interests shifted from PCA to alternative approaches. Furthermore, predicting the symptom trajectory of an individual patient in the clinic based on the population data in the repository was not possible computationally because of data issues. Additionally, the oncologists expressed frustration due to repeated patient failures in following instructions aimed to reduce the symptom burden, such as following a prescribed regimen of swallowing exercises or taking the prescribed pain medication. The physicians felt that having the means to explain to patients a predicted symptom trajectory, in the context of other patients, could be beneficial in terms of therapy adherence.

Accounting for evolving requirements and specifications, we summarize the project activities and their corresponding visual analysis tasks as follows:

A1. Analyze alternative symptom clustering approaches, and apply them to an existing symptom dataset
- T1.1. For each approach, show similar patients, based on symptom severity at a specific time point
- T1.2. For each approach, detect correlations among symptoms, during and after treatment
- T1.3. For each approach, detect patient outliers and trends

A2. Analyze longitudinal symptom progression in the dataset, with particular emphasis on the acute versus late stage of symptoms, and different therapy options
- T2.1. Analyze the patient symptom trajectories as a whole, by therapy type, and by stage
- T2.2. Compare symptom trajectories by therapy type
- T2.3. Summarize symptom ratings for the entire cohort, by stage

A3. Map an individual patient to its relevant cohort, and explain their longitudinal symptom trajectory in the context of the cohort in an actionable manner
- T3.1. Show an individual patient in the context of the cohort
- T3.2. Display demographic and diagnostic patient data, and indicate patients with similar diagnostic attributes
- T3.3. Display the anatomical locations affected by a symptom
- T3.4. Filter a patient's symptoms by association rule

Our evaluation describes example workflows centered on these activities. Non-functional requirements included a request for the A3 data to be displayed in a manner amenable to audiences with low visual literacy, awareness of variability in symptom ratings across patients, and awareness of missing data.

4.3 Data Analysis

In accordance with the ACD paradigm for data visualization [68], the project requirements were based on a starter dataset, which was then expanded during the duration of the project. Patients who had completed fewer than two questionnaires were not included in the analysis. The final dataset included 699 HNC patients.

For each patient, two types of information were recorded: 1) Patient demographics and diagnostic data, which covered three attribute types:

quantitative data (e.g., age, weight, or the total radiation dose); ordinal data (disease stage), and nominal data (e.g., therapeutic combination); and 2) Longitudinal symptom data, as time-series attributes with quantitative values (ratings for 28 symptoms) over a maximum of 12 time points. The symptoms were further grouped in three categories: core symptoms common for all cancer types (fatigue, disturbed sleep, distress, pain, drowsiness, sadness, memory, numbness, dry mouth, lack of appetite, shortness of breath, nausea, and vomiting), HNC specific symptoms (difficulty swallowing, difficulty speaking, mucus in throat, difficulty tasting food, constipation, teeth/gum issues, mouth/throat sores, choking, and skin pain), and ratings of symptoms' interference with daily life (work, enjoyment, general activity, mood, walking, relationships). The symptoms were rated on a 0-to-10 scale ranging from "not present" (0) to "as bad as you can imagine" (10) for the core and HNC specific items, and from "did not interfere" (0) to "interfered completely" (10) for the interference items. Each patient rated all 28 symptoms during a questionnaire completion (time point).

The dataset included a total of 12 time points. Because of the desired longitudinal aspect of the analysis, we separated these points into three categories: baseline (week 0), acute stage (on-treatment period), and late stage ($>= 6$ weeks after treatment). For acute time points during treatment, data was collected every week (at most 7 weeks), while after treatment, time points data was collected at lower granularity, at 6-weeks, and 6-, 12-, or 18-months post-treatment. Previous timepoint values were substituted for missing values; missing baseline values (i.e., for the first timepoint) were marked with 0. Patients with no symptoms recorded during the acute or late phases were not included in the analysis for that time frame.

4.4 Environment Design

The design followed a parallel prototyping approach [22], a method proven to lead to better design results by opening up the visual encoding and interaction space, which in turn elicits more detailed and constructive feedback than in serial prototyping. THALIS was implemented in Python and JavaScript with the D3.js library [11]. The top design is based on coordinated multiple views of the data, in order to support both layering and separation of information and workflow components, and the ability to integrate visually heterogeneous data. A main clustering panel allows the analysis of patient groupings based on similarity (Fig. 2), respectively the analysis of symptom groups via association rule mining (Fig. 1.A). A second main panel supports the longitudinal analysis of patient symptoms (Fig. 1.B), in coordination with the other panels. The remaining panels supports explicitly the context-analysis of cohort symptom data. The panels are connected through explicit filtering operations, brushing and linking.

4.4.1 Clustering Panel

Because of the experts' interest in activities A1 and A3, the clustering panel shows a therapy cluster view of patients (Fig. 2). Alternatively, the panel shows an association graph view of symptoms (Fig. 1.A), illustrating the two main clustering approaches of this project (A1). These views are coupled with computational modules for clustering.

Therapy Cluster View. In prior research, the clinicians had analyzed a subset of the patient data using factor analysis and had identified distinct groups of patients with high, medium, and low symptom burden, depending on the therapeutic combination, which they had illustrated via heatmaps and dendrograms. However, they were also aware that the heatmap representation did not illustrate well outliers in the patient dataset, nor did it support well individual patient analysis, and they were also not confident about the therapeutic distinction between these groups. We agreed that a scatterplot view, color-mapped to the different therapies, would serve activities A1 and A3 better, by capturing more clearly individual patients and cohort patterns in the data.

We first organized the symptom ratings into a patient-symptom matrix for the selected time point, where each element (i, j) corresponds to the score given to symptom j by patient i at that time point. Prior research in symptom cluster for HNC [34] had applied hierarchical clustering using Ward's method [48] with Euclidean distance on the
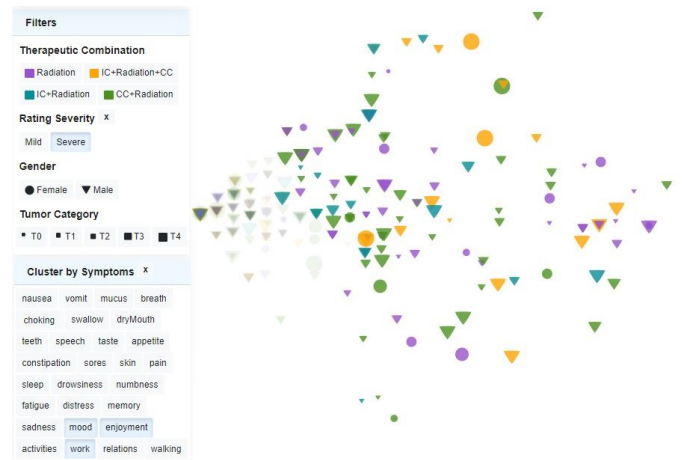

Fig. 2. Custom scatterplot of patients at a specific time point, for a selected rating severity. Left position is associated with a lower symptom burden, calculated based on the symptoms selected in the list. Shape, size, and color encode demographic, diagnostic, and therapy features (see legend). In this example, highlighted patients correspond to the high rating severity group, indicating that the three symptoms selected (mood, enjoyment, and walk) severely affect the vast majority of patients across all therapies, genders, and tumor sizes. Outliers are easily noted.

patient-symptom matrix to group patients based on their raw symptom ratings. After alternative clustering with complete and average linkages, we found that Ward's method generated larger, more informative groups of high symptom patients, which made sense to the clinicians. We identified two patient groups with high and low symptom burden (T1.1). This two-group clustering was preferred by clinicians, who found it easier to compare two groups instead of more. The axes of the scatterplot correspond to the first two components obtained by applying PCA to the patient-symptom matrix. Clusters for a specific time point are extracted and displayed, while clusters for different timepoints can be investigated via the time slider, which will update the scatterplot.

The scatterplot was customized to separately capture acute and late symptom burden distribution as identified by the symptom clusters, and to reflect via marker color, shape, and size the therapeutic combination administered to each patient, their gender, and their disease stage (T3.2) (Fig. 2). The data can be filtered by attributes, and filtering operations update the other views. A filtering control panel serves double duty, by also providing the plot legend. This customized scatterplot encoding effectively captured the symptom distribution across the patient population, patient outliers, therapeutic distribution across the data (T1.1, T1.3, T2.3).

To assess the symptoms' impact on clustering, we also provide an option for dynamically recalculating the clusters based on user-selected subsets of symptoms (Fig. 2) and update the scatterplot accordingly.

Association Rule Diagram View. Driven by the factor analysis limitations discussed earlier, this project pursued Association Rule Mining (ARM) as an alternative, novel approach to symptom cluster analysis (A1). ARM is an unsupervised data mining technique for identifying relationships within the data [2]. In marketing applications, an association rule in the form $X \rightarrow Y$ indicates the pattern that if a customer purchases X, they will also purchase Y, where the patterns are extracted from relational data expressed as transactions. Similar to the strong positive correlations found between items in a supermarket basket, relationships within clinical data can help identify disease comorbidities [43, 54, 57].

In this project, we extended the potential of ARM to symptom clustering applications. To this end, we adapted the most common ARM method to our problem: the Apriori algorithm [2], for frequent item-set mining and association rule learning. In our approach, the symptoms experienced at each time point by each patient are treated as a transaction. The algorithm first identifies frequent symptoms to determine sets of
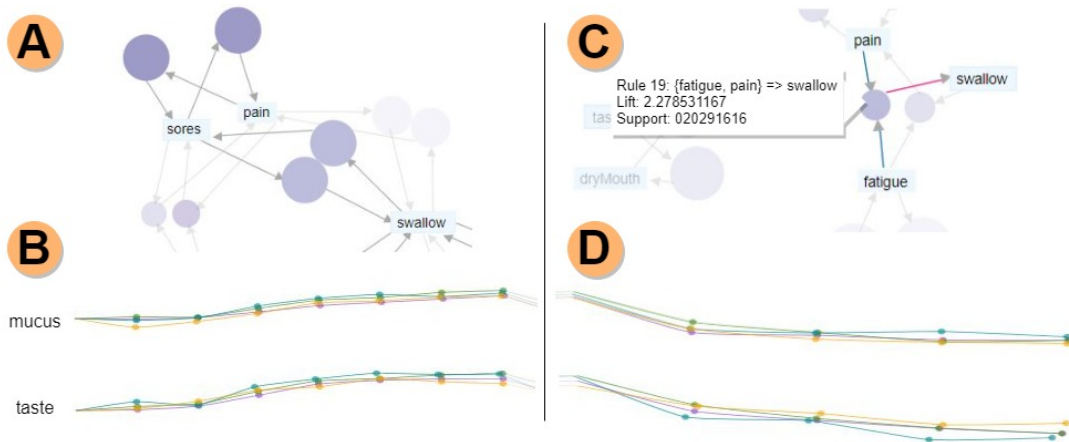
Fig. 3. Acute vs. late phase analysis. A) Association rule diagram for the acute phase. Rules are filtered based on support (frequency) and lift (dependency between symptoms); other rules are faded in the background. B) Mean rating value filament plots for all therapies, with the acute phase highlighted. All therapies follow similar trajectories for both mucus and taste, and towards the end of the acute phase, taste has a considerable increase in ratings for all therapies. C) Association rule diagram for the late phase, showing the antecedents (fatigue, pain) and consequent (swallow) for rule 14. D) Mean rating value filament plots, showing a slightly different trajectory for IC+Radiation.

Table 1. Example of 3 transactions containing 4 symptoms: fatigue, drowsiness, pain, and swallow.

tid	items
001	fatigue, drowsiness
002	pain, drowsiness
003	fatigue, pain, swallow

symptoms that co-occur with high certainty and then extends to larger symptom sets. Table 1 contains an example of three "transactions" from our data. Transactions were extracted from existing questionnaires. Missing ratings for a symptom within a questionnaire implied that the symptom was not included in the transaction. If a patient was missing an entire questionnaire, no transaction was generated for that patient. The ARM was performed using all the available data and no data imputation was performed.

We followed Agrawal and Srikant's proposed association rule [2] in the form:

$$X \rightarrow Y$$

which indicates that if a patient suffers from symptom X (the antecedent), they will also be affected by symptom Y (the consequent). Based on the first transaction in Table 1, such a rule can be:

$$\{\text{fatigue}\} \rightarrow \{\text{drowsiness}\}$$

where {fatigue} is the rule antecedent and {drowsiness} is the consequent. For itemsets larger than this pairwise example (e.g., last transaction in Table 1), either the antecedent or the consequent could contain multiple items.

Two standard measures, *support* and *lift*, are tuned to filter the association rules by a minimum value. *Support* is the measure of how often the transactions contain both X and Y, in our case, how frequently sets of symptoms X and Y occur together. The *support* of a subset of symptoms S is defined by:

$$\sigma(S) = \frac{|S|}{|T|}$$

where $|S|$ is the number of transactions that contain all the symptoms in set S and $|T|$ is the total number of transactions in the dataset. In Table 1, $\sigma(\{\text{fatigue,drowsiness}\}) = \frac{1}{3}$ as both symptoms appear together in 1 out of 3 transactions.

Lift is the measure of the importance, or strength of the rule, and it shows how more frequently than we'd expect by random chance do X and Y appear together. *Lift* is defined as:

$$\lambda(X,Y) = \frac{\sigma(X \cup Y)}{\sigma(X) \times \sigma(Y)}$$

where $(X \cup Y)$ refers to transactions that contain both X and Y. E.g.:

$$\lambda(\{\text{fatigue}\}, \{\text{drowsiness}\}) = \frac{\sigma(\{\text{fatigue,drowsiness}\})}{\sigma(\{\text{fatigue}\}) \times \sigma(\{\text{drowsiness}\})}$$

We applied ARM to each of the acute stage and the late stage (T1.2,T3.4), and empirically chose to illustrate the top 20 rules yielded by this approach, because only a small number of rules were of clinical interest. We chose minimum values for the *support* and *lift* metrics that were suitable for frequent and interdependent symptoms.

From the many possible encodings of ARMs [38], we selected a node-link representation (Fig. 1.A), which was deemed by clinicians to be more friendly to broader audiences (A3), and a good fit for the relatively small number of nodes. Graphs are laid out using a force-directed layout algorithm based on statistical multidimensional scaling [39,79], which results in nodes with high degree being placed centrally. Consistent with this encoding, which is closest to humans arranging nodes manually [92] or when locating connected clusters [81], the layout is fixed. Other layouts have been tested: the dot layout output a tree-like representation, deemed less desirable, whereas the Distributed Recursive Layout and the Fruchterman-Reingold Layout [19] resulted in cluttered diagrams. We followed established design principles for network visualization [71]: circles encode rules, with larger size and deeper shade denoting higher rule support and lift, respectively, whereas rectangles encode symptoms. Incoming edges for a node indicate which item(s) appear in the antecedent of an association rule, whereas outgoing arrows indicate item(s) in the consequent. Because the rule directionality is meaningful, rules containing the same sets of symptoms are treated as separate nodes in the graph. Clicking on a rule highlights the antecedents and consequents of the rule, whereas clicking on a symptom highlights the rules containing that symptom and all the other symptoms in those rules (Fig. 3.C). Rules can be further filtered out based on *support* and *lift* levels (Fig. 3.A).

4.4.2 Symptom Trajectory Panel

Designing an appropriate encoding for the symptom longitudinal data (A2) turned out to be particularly challenging, primarily due to the nature and richness of the temporal data, the acknowledged variability in ratings across patients, and the missing or uneven time points, which were expected in this context. The design process explored a wide range of possible temporal encodings, many of which suffered from scalability issues and, after several sessions, focused on a promising
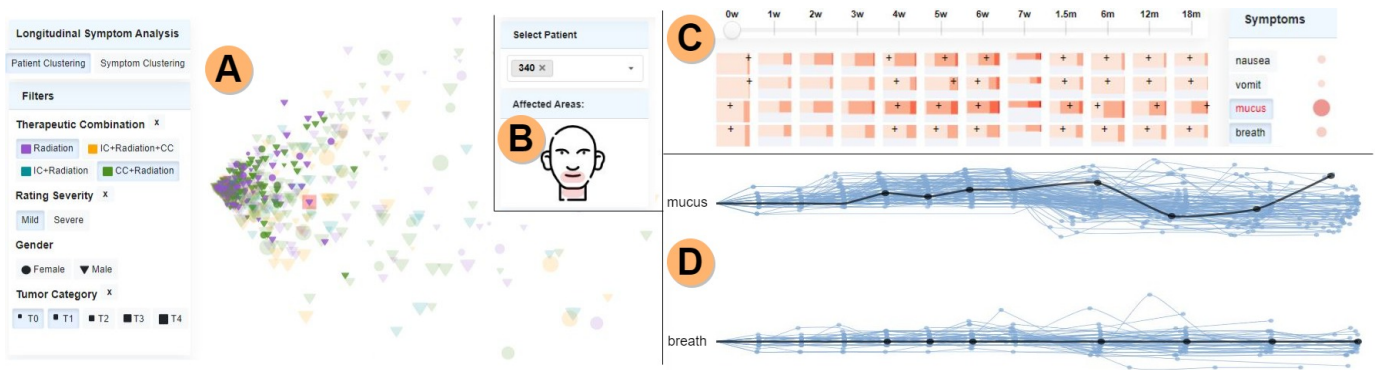
Fig. 4. Symptom burden analysis. A) Patients in the mild symptom burden cluster, having tumor categories T0 and T1 (current patient, 340, is highlighted in red), with all other patients faded. B) The anatomical sketch shows that the mouth and neck areas are affected by the selected symptoms (mucus, breath) for the current patient. C) The patient's ratings are shown by black marks. In this case, the patient had a low rating for mucus at the first assessment (0 weeks), while at the end of the observation period (18 months post-treatment) the rating increased. D) Filament plots encoding symptom trajectories for the selected symptoms, for the patients filtered in the scatterplot. One filament per patient shows the temporal development for that symptom; black filaments mark the current patient, confirming the mucus rating increase in the late stage.

encoding called a "tendril plot" [51]. A tendril plot is a visual summary of the incidence, significance, and temporal aspects of adverse events in clinical trials, in which individual temporal threads, one per each patient, emanate from a common root and shoot upwards and curl either to the left or to the right depending on whether the next event in the timeline was adverse or an improvement. For clinical trial data, tendrils were shown to create beautiful, compact, naturally clustering pathlines illustrating the positive or negative evolution of each group of patients. The clinicians had also seen this representation and thought it could work (T1.2, T2.2). Whereas promising on paper, unfortunately, the tendril implementation did not yield similarly clean illustrations for the symptom data, because of the much smaller number of time points, the variability in therapeutic sequences, and the variability in patient outcomes, which are not typical of clinical trials.

Numerous design variations yielded a new custom temporal encoding, which we call a filament plot (Fig. 4.D). Filament plots also emanate from a common root, then proceed in a left-to-right direction aligned with the time sequence. Wider timesteps, typical for late stage, are accordingly more widely spaced. Each filament represents the full observation period for a specific patient, with dots along the filament to indicate time stamps. To account for inter-patient rating variability, the curvature degree for the filament at each time step encodes the relative change from the previous rating, where upward rotation indicates worsening symptoms (rating increase), and downward rotation shows symptom amelioration (rating decrease).

To calculate the rotation, if patient p is located at position (x_t, y_t) at timestep t for a symptom with rating r, we compute the next position (x_{t+1}, y_{t+1}) at timestep $t+1$ by first calculating the horizontal rotation angle as:

$$\theta = \frac{\theta_{max} \cdot \Delta r_{t+1}}{2 \cdot \Delta r_{max}}$$

where θmax is the total maximum rotation allowed, whose value is set to $\frac{3\pi}{4}$; Δr_{t+1} is the rating difference between $t+1$ and t:

$$\Delta r_{t+1} = r(t+1) - r(t);$$

and Δr_{max} is the maximum difference between two rating values which is 10 in our case. Negative differences between ratings (i.e., rating decreases) produce negative angle values for θ.

Next, we want to rotate θ degrees relative to the horizontal line $P_1 P_2$ defined by the points $P_1 = (x_t, y_t)$ and $P_2 = (x_t + l, y_t)$ where l quantifies the time elapsed between $t+1$ and t. A higher l indicates that more time passed between $t+1$ and t (i.e., late vs acute). Finally, we rotate P_2 around P_1 by θ degrees.

For missing data during the observation period, the associated points are not represented, and we consider no rating change from the previous

time points; the surveillance period is represented on each filament until the last recorded time point for each patient. We account for the time ratio between the acute (1 week) and late (months) stages, so the distances illustrated for the acute time points are smaller as opposed to the late time points. Hovering over a filament greys out all the other filaments in the plot. This interaction helps in the comparison of symptom trajectories for the same patient, and via brushing and linking with the other views, in highlighting the additional patient data (T3.1).

This compact representation helps in the analysis of symptom evolution trends, by clearly indicating the overall symptom burden (low/high). The representation also helps in spotting outlier trajectories that should be further evaluated and facilitates the discovery of steady vs. variable progression of symptoms. The panel includes two such filament plots, supporting the side-by-side comparison of different symptoms for selected patient groups. To further enhance visual support, during the evaluation of the acute period in the entire THALIS environment, the acute time periods are highlighted in the filament plots, and vice versa for the late period (Fig. 3.B and Fig. 3.D).

In order to better support activities A1 and A2, an additional option uses the same filament encoding, this time with the color mapped to the therapy type, to capture the mean trajectory per each therapeutic combination (Fig. 1.B). Since in the therapy case the symptom mean ratings across the population bear meaning, the filaments are spread out according to the mean ratings per therapy (T2.1). This therapy-analysis option helps estimate what treatment plans are less symptomatic, or on the contrary, conduct to high symptom burden. In addition, to satisfy activity A3, the current patient's filament is highlighted in black in each plot (Fig. 4.D). Whereas reliable automated symptom prediction is an unsolved problem in symptom research, THALIS supports human-machine analysis via trajectory views of similar patients.

4.4.3 Cohort Symptom Panel

The last panel explicitly supports activities A1 and A3, and provides an abstract summary of the entire temporal symptom data. As in other fields [61], and as indicated by our activity analysis, this summary provides context for a specific datapoint, but does not lead the investigation. The panel comprises a percentile heatmap, a correlation matrix, and an anatomical sketch (Fig. 1).

The percentile heatmap (Fig. 1.D) is a custom representation showing the rating distribution of individual symptoms over time, for the entire patient cohort (T2.3). We arrived at this representation after exploring a variety of alternatives such as stacked line plots, parallel coordinates plots, and radar charts, guided by feedback from collaborators. We settled on a matrix-based layout due to its compactness and to its ability to support small multiple plots. Each row corresponds to a symptom, with rows grouped by symptom category, and each column
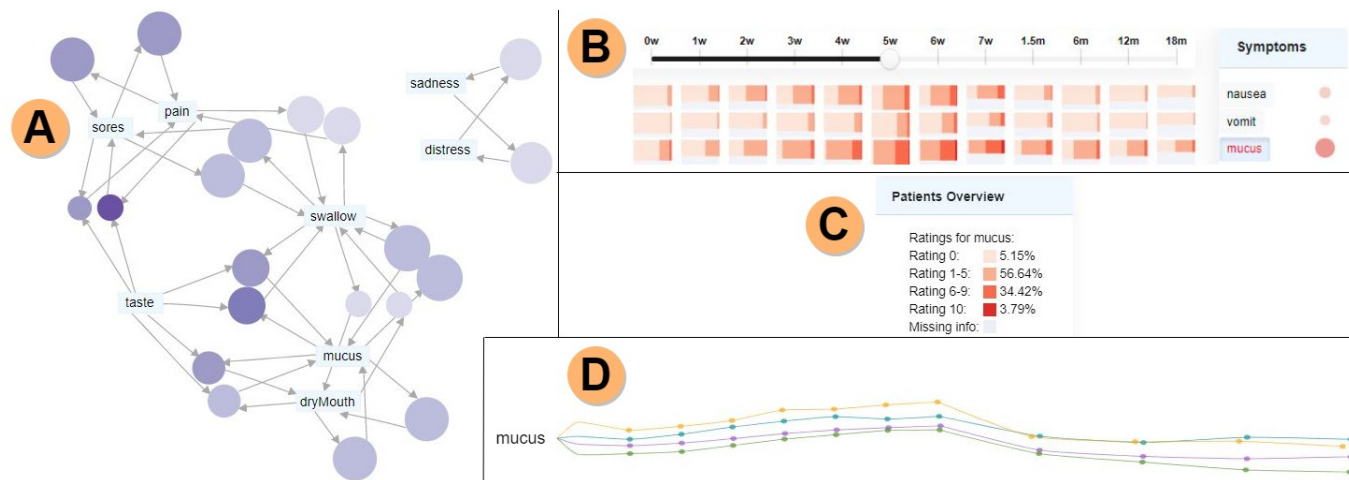
Fig. 5. Symptom cluster diversity analysis. A) Symptom association graph for the acute phase showing mucus and swallow correlate with many symptoms. Note that the network layout is fixed, and that by construction it places centrally nodes with high degree. B) The percentile heatmap shows a spread of high ratings for mucus along the whole observation period. C) Summary panel for mucus showing that among patients who reported ratings for week 5 during treatment, more than 95% noted mucus as a present symptom. D) Mean rating filament plot emphasizing rising ratings at the end of the acute phase, especially for the IC+Radiation+CC treatment.

corresponds to a time point. Each cell in this matrix is a horizontal bar graph showing via shade the percentage of patients reporting within a specific range (0, 1-5, 6-9, or 10) for that symptom, at that time point. The bar height maps the percentage of individuals from the entire cohort who reported the symptom ratings at that time point. The current patient is indicated in this heatmap by cross markers (Fig. 4.C) (T3.1). This encoding proved to be an intuitive way of showing what symptoms produce a higher burden for patients, and when, as well as to indicate how many patients were affected by these symptoms from the entire cohort (T1.2, T2.3).

To support exploration driven by a specific patient (A3), a dropdown selection box is also provided (Fig. 4.B). A selection in this box highlights the patient data across panels (Fig. 4). A timeline selector further allows the selection of a particular time point in the data (Fig. 4.C), and further interface elements allow selection and analysis of sets of similar patients. Additionally, a compact correlation matrix (Fig. 1.E), along with the percentile heatmap, supports T1.2, by showing the strength of the correlation between a selected symptom and all other symptoms, with circles encoding Spearman's coefficient via color and size. Finally, because a discussion of task T3.3 revealed that patients tend to point to the location of their symptoms, an anatomical sketch (Fig. 1.C) supports visual anchoring based on anatomy. Regions in the head and neck affected by the selected symptoms are highlighted in this sketch.

5 EVALUATION AND RESULTS

Because no design approach is failproof, although ACD has higher success rates than HCD (63% compared to 25%) [68], we evaluated THALIS through a combination of multiple demonstrations and case studies involving domain experts, namely a senior data mining specialist and three senior clinical radiation oncology experts. Whereas we recognize these experts as co-authors, not all of them were involved in the development process at all stages. Two case studies were completed during separate, dedicated sessions, in addition to regular feedback sessions. Because the designers and evaluators were in different locations, and due to COVID-19 constraints, these sessions were conducted remotely using screen sharing and note-taking. The oncology experts directed the exploration using the think-aloud method, while the first author was driving the interface according to their instructions. Both case studies analyze a set of 699 HNC patients, which was significantly larger than prior clinician analyses, and span all activities, A1-A3. Qualitative feedback was also provided during weekly design-driven sessions and was used to improve the overall design of THALIS.

5.1 Case I: Symptom-Burden Analysis in Radiotherapy

The study seeked to assess the impact of therapy on symptom burden on this set, and took place before our development of the associative rule model. The oncologists were originally hoping to replicate published analysis results obtained on significantly smaller cohorts of 80 to 270 patients [28, 50, 85]. Using the system over the course of several sessions showed, however, that those clustering results were not generalizable to the larger cohort, and so the investigation shifted focus to discovering and analyzing outliers in terms of patient characteristics and symptom trajectories. The study workflow started directly with the therapy scatterplot panel (Fig. 4.A) (T1.1, T3.2). At first glance, most patients were visibly grouped in the left-center part of the plot, suggesting strong similarity. Filtering the patients (T3.1) based on their rating severity revealed that this group corresponded to a mild-rating severity cluster. Further filtering by therapy and tumor category, the experts noted that most of these patients were treated with radiation with or without concurrent chemotherapy (CC) and, not surprisingly, presented a small tumor size and a low symptom burden at the end of the observation period. They concluded that for this group, the therapy plan did not effectively impact the quality of life. Next, the oncologists examined whether a smaller set of symptoms, as in their prior studies, would correlate with patient groupings (T1.1, T1.3). To this end, they filtered data by daily interference symptoms, including, for example, {mood, enjoyment, and work} (Fig. 2). This time, they found that almost a third of the patients suffered from high symptom burden in this symptom group.

Encouraged by this finding, the analysis moved swiftly to the filament plots (Fig. 4.D), to examine the symptom trajectories (T2.2). The plots captured a general trend in most symptom trajectories, namely, a rating decrease post-treatment, with the exception of {numbness, memory, breath}. Moreover, these three symptoms, along with nausea and vomit, exhibited a steady symptom development, with fewer patient outliers or drastic rating changes over time (T1.2). There was, in fact, no correlation between the temporal outliers in the filament plots and the therapy scatterplot outliers. This finding indicated that patients experienced steady ratings for these five symptoms over time, regardless of overall symptom burden or therapy treatment. This observation was of notable interest, and so the analysis moved to examine the cohort context (T2.3). Using the percentile heatmap (Fig. 1.D) and the correlation matrix, our collaborators noted that groups of symptoms such as {swallow and dry mouth}, or {taste, appetite, constipation, and sores} showed higher ratings over time, suggesting possible interrelationship or causative factors between these symptoms. For example, when
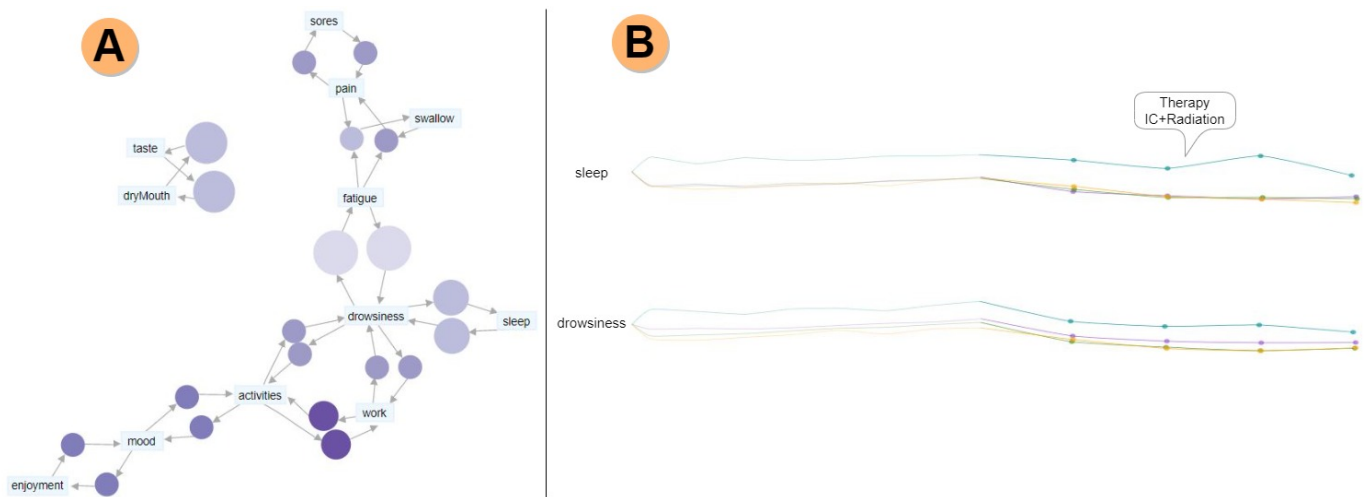
Fig. 6. Late symptom cluster analysis. A) Symptom association graph showing drowsiness as a central symptom for the late phase. The connection between sleep and drowsiness is expected, as these two symptoms are known to be a factor in dangerous muscle-mass loss. B) The filament plots show mean rating values, with the late phase highlighted, and the acute phase faded. Notably, in the case of sleep and drowsiness, IC+Radiation is the therapy associated with higher symptom ratings, and it is noticeably different from the other therapy plans.

selecting dry mouth, the panel indicated strong correlations between dry mouth and {mucus, choking, and swallow}, but also with {taste, drowsiness, and fatigue} as well. Finally, the anatomical sketch layout (Fig. 1.C) emphasized which head and neck locations are affected by the selected symptoms (T3.3). In this case, we noted that both dry mouth and taste affected the mouth area. The oncologists are planning studies to verify this set of symptom cluster hypotheses.

5.2   Case II: Symptom Cluster Diversity

This study aimed primarily to explore the value of associative rule mining in longitudinal symptom analysis (T1.2). Examining the association diagrams, the oncologists were stunned to find surprising symptom clusters during and post-treatment; in particular, 8 common symptoms for the acute stage (Fig. 5.A), with two strongly coupled subgroups: {distress, sadness}, and {swallow, pain, sores, taste, mucus}; and respectively, 12 frequent symptoms during the late part of the treatment (Fig. 1.A), showing symptom clusters such as {taste, dry mouth}, and {sores, pain}. The experts were impressed to see that the {sores, pain} cluster is strongly associated with {taste} in the acute phase, while in the late phase, there is a connection between {drowsiness, sleep} (Fig. 6.A), which is known to be a factor in dangerous muscle-mass loss. The {taste, dry mouth} cluster in the late phase supported our collaborators' previous findings. However, the connection between {fatigue, drowsiness} in the late phase and the centrality of {mucus} (Fig. 5.A), as well as the {taste, sores} connection within the acute graph was unexpected. *"In our group, we have established this arc from taste to dry mouth in late stage, but we haven't thought of the taste to sores link in the acute phase. That is striking."*

The ability to highlight a particular symptom (T3.4) or rule and to filter the rules based on their support and lift (Fig. 3.A, Fig. 3.C) were found essential during the exploration, by helping our collaborators to figure out which symptoms were more persistent or more dependent on each other. For instance, {fatigue, drowsiness} were the most common symptoms (based on their support) and {activities, work} the most dependent on each other (based on their lift) in the late phase (Fig. 1.A). Insights observed from the symptoms association graphs were further extended using the percentile heatmap (Fig. 1.D), revealing the spread of high ratings for {taste} and {fatigue} over the whole patient supervision period (T2.3). Moreover, because {mucus} was usually perceived as an acute symptom, the experts found it remarkable that a large number of patients experienced {mucus} during the late period as well (Fig. 6). The mean value filament plots were used to show the mean ratings per time point for each therapy while high-

lighting the treatment phase of interest (acute/late) (Fig. 3.B, Fig. 3.D) (T2.1). The plots showed that the trends were remarkably conserved over time between therapies, even though their magnitudes might differ. To achieve a better understanding, the option to separate the filaments based on the starting mean rating (baseline) was used (Fig. 1.B, Fig. 5.D) which showed a difference in the symptom burden between therapies for the association-identified symptom groups. For example, in the case of {taste} and {mucus}, in both acute and late phases, the highest rated treatments were IC+Radiation+CC (induced chemotherapy, radiation, and concurrent chemotherapy) and IC+Radiation, while CC+Radiation and Radiation alone were rated lower. Noticeably, in the case of {drowsiness, sleep}, IC+Radiation was remarkably separated from the other treatment plans (Fig. 6.B). The oncologists concluded this case study and the associative approach were a gold mine for their symptom research, by highlighting the diversity of symptom clusters over time.

5.3   Expert Feedback

Because THALIS used participatory design, feedback from the domain experts is implicitly reflected in the final design choices we report. Here, we focus instead on expert feedback related to the current system. The current version of THALIS yielded excellent feedback from the oncology team, often indicating a shift in thinking about their work. We report sample feedback, in relation to our activity analysis A1-A3:

(A1, A3) Quote from the most senior oncologist: *I gotta be honest, every time I meet with you guys and we see these visualizations, I get so much material for future research. In general, to be fair, my focus in clinical practice [and in helping patients] tends to be on dry mouth and swallowing. I say "We're going to talk about dry mouth and swallowing, cause these two are really bad", and "then there's all the other stuff". And then I see this [the ARM and heatmap and filaments], and here's this other stuff, that is usually at my periphery, but I don't focus on, although patients do mention it. If I were sitting with a patient and I'd look at this interface and ARMs—I get it, hey, there's actually a LOT of moving parts here [beyond dry mouth and swallowing], and they're related, and they have different time sources. It's sobering.*

(A1, A2) Both case studies had the team exclaim, on multiple occasions, about being *"blown away", "surprised by that", "that [symptom] spread over time just jumps out at you", "This entire ARM approach is so different [from the approach we've followed in our past research on symptom clusters]. I want to stick a flag in the ground with the ARM work, and look at dose to organs and use ARM to see dose-to-swallowing correlation, based on this spatial structure underneath",*

"This interface and the ARM provide great preliminary data for so many grants [projects] right off the bat!", "Really impressed", and "[This relationship] is not intuitive, so it's very interesting. And I wouldn't have thought about it. But now, it makes perfect sense. Duh!", "The [filament view] is such a great asset for the interface."

(A3) The clinician oncologists: *"[THALIS's] ability to go from patient to population is fantastic, I really love it, it's exactly what I need", "I like that when a patient is with [CDF], they want percentages, e.g., 66% of patients have normal appetite after 12 months, and [THALIS] shows that", "When I see a patient, this [taste-dry mouth] association in the late phase is the default picture I have in my mind. But here I see that also fatigue connects to drowsiness, and that these symptoms show up in the acute phase as well, and that I really need to discuss these issues with my patients." "I can share [this view] with my patients, to explain that pain and swallowing and fatigue are really tightly related— we don't know if it's causation, but they definitely show up together, so could you please, please, take your pain and anti-inflammatory meds, and could you please do the swallowing exercises we've talked about?"*

6 DISCUSSION

The case studies and the domain expert feedback demonstrate THALIS's value in bridging the gap between machine and human analysis, and its ability to help generate novel insights. Our integrated approach is able to capture longitudinal differences between acute and late stages, while detecting outliers and trends in the symptom and therapy data. More importantly, our approach supports individual patient analysis, while handling a large cohort both computationally and visually. Through an ACD approach, and as indicated by the expert feedback, THALIS successfully serves the core interests of its audience. In conjunction with the clustering panel, the symptom association rule view, the filament plots, and the cohort symptom panel enabled discovering interesting relationships in the data, and in several cases lead to unexpected but insightful results. Furthermore, THALIS couples multiple customized novel visual-encodings with symptom clustering algorithms in the background, enabling the domain experts to explore multiple scenarios and test their hypotheses in real-time. Its use of a multi-view paradigm supports flexible analytical workflows that leverage computational power and human expert knowledge.

Through close collaboration with domain experts, our solution introduces compact, customized visual encodings for the symptom data: a filament encoding and a percentile heatmap. The percentile heatmap scales well with the number of subjects, by design, at the cost of summarization. Whereas the inherent scalability of filaments with the number of items shown is limited, these encodings successfully abstract the cohort data with the help of similarity-based filtering operations, which are appropriate in this context; for hundreds of dense observations, as common in other problems, tendrils [51] offer a better solution. In further terms of scalability, the ARM graph can provide rules for any number of time points in the late and acute time periods. Still, the graph representation for association rules is suited for a smaller number of rules (less than 100 [39]). The scatterplot and correlation matrix are time point specific, so any number of plots could be generated. On the other hand, some views are prone to clutter. Some of these encodings may have limited generalizability beyond this application domain. In the case of filaments, they work in this application because there is a significant correlation between similar patients' trajectories and because our application emphasizes relative trajectory changes as opposed to absolute values. This type of correlation and relativity may not be true across application domains. However, our custom encodings can be repurposed for other longitudinal problems that feature missing data, as in astronomy or biology [40, 62, 63, 87]. Future work includes longitudinal clustering, applying the ARM approach on sequential data, and interactively changing ARM metrics and the number of rules.

Reflecting upon this successful design experience, we extract three main lessons for designers dealing with similar problems:

L1. *Use an activity-centered design process, in particular in remote collaborations.* In our experience, following a design process focused on activities as opposed to humans, from requirements engineering to the evaluation against these activities, allowed us to align this project with the core client interests. Because of this alignment with their core activities, we had significant buy-in from clients, providing us with the ability to stay on task and make steady progress, propelled by activity-relevant insights in several meetings. The approach furthermore resulted in a successful remote collaboration, and an eagerness to adopt THALIS in the clinic.

L2. *Use visual scaffolding to introduce custom, novel visual encodings.* Through many design iterations, our solution converged towards custom encodings, such as filaments and the percentile heatmap. These compact encodings are scalable (in the case of filaments, through filtering), and effectively serve the original design aims. We were able to introduce these encodings through visual scaffolding [67] over many meetings with domain experts: small, gradual changes from one iteration to the next. From the other end of the spectrum, established encodings such as scatterplots and node-link diagrams have greater adoption chances in low visual literacy environments. A mix of novel and standard encodings, when following visualization design principles, may facilitate encoding adoption.

L3. *In XAI (explainable AI), emphasize domain sense and actionability.* In healthcare applications like THALIS, that blend visual encodings with alternative AI methodology such as ARM, we found that transparency in the AI model [59] was not enough to make the model trustworthy. Beyond transparency, our model outputs, in their node-link representation, made sense to the domain experts because THALIS did not contradict their clinic knowledge, and although it did not confirm earlier findings on smaller cohorts, the experts appeared to gain trust in it. At the same time, it was essential to make this AI model actionable: cohort-based analyses are useful in symptom research, but in the clinic, the emphasis is on the individual patient, their therapy, and their likely symptom trajectory. Building an integrated machine-human system that explicitly supports the need to act on the patient's care served our project well.

7 CONCLUSION

In this work we described the activity-centered design of THALIS, a novel environment to support the integrated human-machine analysis of longitudinal symptom clusters as a function of cancer therapy. We described the application domain data and activities with an emphasis on the multidisciplinary development of clustering tools for symptom data in cancer therapy. We also introduced a novel blend of data mining and visual encodings to predict and explain longitudinal symptom development based on an existing cohort of patients and described customized interactive encodings: interactive association-rule graphs, filaments, and percentile heatmaps. The evaluation of the resulting mixed workflows and encodings over an existing head and neck cancer symptom repository with domain experts proves the value of this integrated approach for both symptom research and work in the clinic. Last but not least, we summarized the design lessons learned from this successful, multi-site, remote collaboration. We hope these lessons will help other designers who tackle similar design problems and challenges in human-machine integrated visual analysis.

REFERENCES

[1] S. S. Abdullah, N. Rostamzadeh, K. Sedig, A. X. Garg, et al. Visual analytics for dimension reduction and cluster analysis of high dimensional electronic health records. *Inform.*, 7(2), 2020.

[2] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules in Large Databases. In *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, p. 487–499. Morgan Kaufmann Publishers Inc., 1994.

[3] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 2011.

[4] A. Aktas, D. Walsh, and L. Rybicki. Symptom clusters: myth or reality? *Palliative Med.*, 24(4):373–385, 2010.

[5] P. Angelelli, S. Oeltze, J. Haász, C. Turkay, et al. Interactive Visual Analysis of Heterogeneous Cohort-Study Data. *IEEE Comp. Graph. and Appl.*, 34(5):70–82, 2014.

[6] M. Y. Ansari, A. Ahmad, S. S. Khan, G. Bhushan, et al. Spatiotemporal clustering: a review. *Artif. Intel. Rev.*, pp. 1–43, 2019.

[7] L. Bartram and M. Yao. Animating causal overlays. In *Comp. Graph. Forum*, vol. 27, p. 751–758. Wiley Online Library, 2008.

[8] T. Baumgartl, M. Petzold, M. Wunderlich, M. Hohn, et al. In Search of Patient Zero: Visual Analytics of Pathogen Transmission Pathways in Hospitals. *IEEE Trans. Vis. Comp. Graph.*, 27(2):711–721, 2021.

[9] J. Bernard, D. Sessler, T. May, T. Schlomm, et al. A visual-interactive system for prostate cancer stratifications. In *Proc. IEEE VIS Workshop Visualizing Electronic Health Record Data*, 2014.

[10] M. Biggs, C. Floricel, L. Van Dijk, et al. Identifying Symptom Clusters from Patient Reported Outcomes through Association Rule Mining. *19th Int. Conf. Artif. Intel. in Med. (AIME)*, 2021.

[11] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Trans. Vis. Comp. Graph.*, 17(12):2301–2309, 2011.

[12] D. Bruzzese and C. Davino. Visual mining of association rules. In *Visual Data Mining*, p. 103–122. Springer, 2008.

[13] R. Cava, C. M. D. S. Freitas, and M. Winckler. ClusterVis: Visualizing Nodes Attributes in Multivariate Graphs. In *Proc. Symp. App. Comp.*, p. 174–179. ACM, 2017.

[14] M. Cavallo and Ç. Demiralp. Clustrophile 2: Guided Visual Clustering Analysis. *IEEE Trans. Vis. Comp. Graph.*, 25(1):267–276, 2019.

[15] Ç. Demiralp. Clustrophile: A Tool for Visual Clustering Analysis. *arXiv:1710.02173*, 2017.

[16] K. M. Christopherson, A. Ghosh, A. S. R. Mohamed, M. Kamal, et al. Chronic radiation-associated dysphagia in oropharyngeal cancer survivors: Towards age-adjusted dose constraints for deglutitive muscles. *Clin. & Trans. Rad. Onco.*, 18:16–22, 2019.

[17] K. K. H. Chui, J. B. Wenger, S. A. Cohen, and others. Visual Analytics for Epidemiologists: Understanding the Interactions Between Age, Time, and Disease with Multi-Panel Graphs. *PLoS one*, 6(2):1–8, 2011.

[18] C. S. Cleeland, T. R. Mendoza, X. S. Wang, C. Chou, et al. Assessing symptom distress in cancer patients: The M.D. Anderson Symptom Inventory. *Cancer*, 89:1634–46, 11 2000.

[19] G. Csardi, T. Nepusz, et al. The igraph software package for complex network research. *InterJournal, comp. sys.*, 1695(5):1–9, 2006.

[20] D. Dingen, M. van't Veer, P. Houthuizen, E. H. Mestrom, et al. RegressionExplorer: Interactive exploration of logistic regression models with subgroup analysis. *IEEE Trans. Vis. Comp. Graph.*, 2018.

[21] S. T. Dong, D. S. Costa, P. N. Butow, M. R. Lovell, et al. Symptom Clusters in Advanced Cancer Patients: An Empirical Comparison of Statistical Methods and the Impact on Quality of Life. *Pain and Symptom Manag.*, 51(1):88–98, 2016.

[22] S. P. Dow, A. Glassco, J. Kass, M. Schwarz, et al. Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy. *ACM Trans. Computer-Human Inter.*, 17(4), 2011.

[23] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. EventAction: Visual analytics for temporal event sequence recommendation. In *IEEE Conf. Vis. Anal. Sci. and Tech. (VAST)*, p. 61–70, 2016.

[24] A. K. Dubey, U. Gupta, and S. Jain. Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. *Int. J. Comp. Ass. Rad. and Sur.*, 11(11):2033–2047, 2016.

[25] B. Elgohari, A. Wentzel, P. Hanula, T. Luciani, et al. Cohort-Based Spatial Similarity Can Predict Radiotherapy Dose Distribution. *Int. J. Rad. Onco. • Bio. • Phys.*, 105(1):E416–E417, 2019.

[26] H. Elhalawani, A. S. Mohamed, et al. Matched computed tomography segmentation and demographic data for oropharyngeal cancer radiomics challenges. *Scientific data*, 4:170077, 2017.

[27] N. Elmqvist and P. Tsigas. Growing squares: Animated visualization of causal relations. In *Proc. ACM Symp. Software Vis.*, p. 17–ff, 2003.

[28] S. A. Eraj, M. K. Jomaa, C. D. Rock, A. S. R. Mohamed, et al. Long-term patient reported outcomes following radiation therapy for oropharyngeal cancer. *Rad. Onco.*, 12(1):150, 2017.

[29] G. Fan, L. Filipczak, and E. Chow. Symptom Clusters in Cancer Patients: A Review of the Literature. *Curr. Onco.*, 14(5):173–179, 2007.

[30] K. A. Ferreira, M. Kimura, M. J. Teixeira, et al. Impact of Cancer-Related Symptom Synergisms on Health-Related Quality of Life and Performance Status. *Pain and Symp. Manage.*, 35(6):604–616, 2008.

[31] A. G. Forbes, A. Burks, et al. Dynamic influence networks for rule-based models. *IEEE Trans. Vis. Comp. Graph.*, 24(1):184–194, 2017.

[32] K. Furmanová, N. Grossmann, L. P. Muren, O. Casares-Magaz, et al. VAPOR: Visual Analytics for the Exploration of Pelvic Organ Variability in Radiotherapy. *Comp. & Graph.*, 91:25–38, 2020.

[33] D. Gotz and H. Stavropoulos. DecisionFlow: Visual Analytics for High-Dimensional Temporal Event Sequence Data. *IEEE Trans. Vis. Comp. Graph.*, 20(12):1783–1792, 2014.

[34] G. B. Gunn, T. R. Mendoza, C. D. Fuller, I. Gning, et al. High symptom burden prior to radiation therapy for head and neck cancer: a patient-reported outcomes study. *Head & neck*, 35(10):1490–1498, 2013.

[35] S. Guo, F. Du, S. Malik, E. Koh, et al. Visualizing uncertainty and alternatives in event sequence predictions. In *Proc. CHI Conf. Human Factors in Comput. Sys.*, p. 1–12, 2019.

[36] S. Gupta and R. Mamtora. A survey on association rule mining in market basket analysis. *Int. J. Info. Comp. Techn.*, 4(4):409–414, 2014.

[37] C. K. Gwede, B. J. Small, P. N. Munster, M. A. Andrykowski, et al. Exploring the differential experience of breast cancer treatment-related symptoms: a cluster analytic approach. *Support. Care in Canc.*, 16(8):925–933, 2008.

[38] M. Hahsler. arulesViz: Interactive Visualization of Association Rules with R. *The R J.*, 9(2):163–175, 2017.

[39] M. Hahsler and S. Chelluboina. Visualizing association rules: Introduction to the r-extension package arulesviz. *R proj. mod.*, pp. 223–238, 2011.

[40] P. Hanula, K. Piekutowski, J. Aguilera, and G. E. Marai. Darksky halos: use-based exploration of dark matter formation data in a hybrid immersive virtual environment. *Front. in Robo. and AI*, 6:11, 2019.

[41] R. L. Harris. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, Inc., USA, 1999.

[42] C. W. Huang, R. Lu, U. Iqbal, S. H. Lin, et al. A richly interactive exploratory data analysis and visualization tool using electronic medical records. *BMC Med. Infor. and Dec. Making*, 15:92, 2015.

[43] Q. R. Huang, Z. Qin, S. Zhang, and C. M. Chow. Clinical Patterns of Obstructive Sleep Apnea and Its Comorbid Conditions: A Data Mining Approach. *Clin. Sleep Med.*, 04(06):543–550, 2008.

[44] J. Illi, C. Miaskowski, B. Cooper, et al. Association between pro- and anti-inflammatory cytokine genes and a symptom cluster of pain, fatigue, sleep disturbance, and depression. *Cytokine*, 58(3):437–447, 2012.

[45] A. Inselberg. Multidimensional detective. In *Proc. VIZ: Vis. Conf., Info. Vis. Symp. and Parallel Rendering Symp.*, p. 100–107, 1997.

[46] W. Jentner and D. A. Keim. Visualization and visual analytic techniques for patterns. *High-Utility Patt. Min.*, p. 303–337, 2019.

[47] Z. Jin, S. Cui, S. Guo, et al. CarePre: An Intelligent Clinical Decision Assistance System. *ACM Trans. Comput. Healthcare*, 1(1), 2020.

[48] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., USA, 2002.

[49] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[50] M. Kamal, M. P. Barrow, J. S. Lewin, A. Estrella, et al. Modeling symptom drivers of oral intake in long-term head and neck cancer survivors. *Supp. Care in Cancer*, 27(4):1405–1415, 2019.

[51] M. Karpefors and J. Weatherall. The Tendril Plot—a novel visual summary of the incidence, significance and temporal aspects of adverse events in clinical trials. *Amer. Med. Info. Assoc.*, 25(8):1069–1073, 2018.

[52] M. Kaur and S. Kang. Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Proc. Comp. Sci.*, 85:78–85, 2016.

[53] H.-J. Kim, I. Abraham, and P. S. Malone. Analytical methods and issues for symptom cluster research in oncology. *Curr. Opi. in Supp. and Pall. Care*, 7(1):45–53, 2013.

[54] L. Kim and S. Myoung. Comorbidity Study of Attention-deficit Hyperactivity Disorder (ADHD) in Children: Applying Association Rule Mining (ARM) to Korean National Health Insurance Data. *Iranian J. Pub. Health*, 47(4):481–488, 2018.

[55] J. Kirkova, A. Aktas, D. Walsh, and M. P. Davis. Cancer Symptom Clusters: Clinical and Research Methodology. *Palliative Med.*, 14(10):1149–1166, 2011.

[56] P. Klemm, S. Oeltze-Jafra, K. Lawonn, K. Hegenscheid, et al. Interactive Visual Analysis of Image-Centric Cohort Study Data. *IEEE Trans. Vis. Comp. Graph.*, 20(12):1673–1682, 2014.

[57] R. Kost, B. Littenberg, and E. S. Chen. Exploring generalized association

rule mining for disease co-occurrences. *AMIA Symp. Proc.*, p. 1284–1293, 2012.

[58] B. C. Kwon, M. Choi, J. T. Kim, et al. RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records. *IEEE Trans. Vis. Comp. Graph.*, 25(1):299–309, 2019.

[59] Z. C. Lipton. The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery. *Queue*, 16(3):31–57, 2018.

[60] M. H. Loorak, C. Perin, N. Kamal, M. Hill, et al. Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients. *IEEE Trans. Vis. Comp. Graph.*, 22(1):409–418, 2015.

[61] T. Luciani, A. Burks, et al. Details-first, show context, overview last: supporting exploration of viscous fingers in large-scale ensemble simulations. *IEEE Trans. Vis. Comp. Graph.*, 25(1):1225–1235, 2018.

[62] T. B. Luciani, B. Cherinka, D. Oliphant, Myers, et al. Large-scale overlays and trends: Visually mining, panning and zoomingthe observable universe. *IEEE Trans. Vis. Comp. Graph.*, 20(7):1048–1061, 2014.

[63] C. Ma, T. Luciani, A. Terebus, J. Liang, and G. E. Marai. Prodigen: visualizing the probability landscape of stochastic gene regulatory networks in state and time space. *BMC bioinform.*, 18(2):1–14, 2017.

[64] T. Ma and A. Zhang. Integrate multi-omic data using affinity network fusion (ANF) for cancer patient clustering. In *IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, p. 398–403, 2017.

[65] N. S. Madiraju. *Deep temporal clustering: Fully unsupervised learning of time-domain features*. PhD thesis, Arizona State University, 2018.

[66] S. Malik, F. Du, M. Monroe, E. Onukwugha, et al. Cohort Comparison of Event Sequences with Balanced Integration of Visual Analytics and Statistics. In *Proc. 20th Int. Conf. Intel. UI (IUI)*, p. 38–49. ACM, 2015.

[67] G. E. Marai. Visual Scaffolding in Integrated Spatial and Nonspatial Analysis. In *EuroVis Workshop Visual Analytics (EuroVA)*. The Eurographics Association, 2015.

[68] G. E. Marai. Activity-Centered Domain Characterization for Problem-Driven Scientific Visualization. *IEEE Trans. Vis. Comp. Graph.*, 24(1):913–922, 2018.

[69] G. E. Marai, C. Ma, A. T. Burks, F. Pellolio, et al. Precision Risk Analysis of Cancer Therapy with Interactive Nomograms and Survival Plots. *IEEE Trans. Vis. Comp. Graph.*, 25(4):1732–1745, 2019.

[70] G. E. Marai and T. Möller. The Fabric of Visualization. In *Foundations of Data Visualization*, pp. 5–14. Springer, 2020.

[71] G. E. Marai, B. Pinaud, K. Bühler, A. Lex, et al. Ten simple rules to create biological network figures for communication, 2019.

[72] A. Maries, N. Mays, M. Hunt, K. F. Wong, et al. GRACE: A Visual Comparison Framework for Integrated Spatial and Non-Spatial Geriatric Data. *IEEE Trans. Vis. Comp. Graph.*, 19(12):2916–2925, 2013.

[73] T. Metsalu and J. Vilo. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*, 43(W1):W566–W570, 2015.

[74] C. Miaskowski, A. Barsevick, A. Berger, et al. Advancing Symptom Science Through Symptom Cluster Research: Expert Panel Proceedings and Recommendations. *Nat. Cancer Inst. (JNCI)*, 109, 2017.

[75] C. Miaskowski, B. A. Cooper, S. M. Paul, et al. Subgroups of patients with cancer with different symptom experiences and quality-of-life outcomes: a cluster analysis. *Onco. Nurs. Forum*, 33(5):E79–89, 2006.

[76] J. Müller, V. Zebralla, S. Wiegand, and S. Oeltze-Jafra. Interactive visual analysis of patient-reported outcomes for improved cancer aftercare. In *EuroVA@ EuroVis*, pp. 78–82, 2019.

[77] Multidisciplinary Larynx Cancer Working Group. Conditional Survival Analysis of Patients With Locally Advanced Laryngeal Cancer: Construction of a Dynamic Risk Model and Clinical Nomogram. *Sci. Reports*, 7(1):43928, 2017.

[78] D. Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[79] S. C. North. Drawing graphs with neato. *NEATO User man.*, 11(1), 2004.

[80] C. Plaisant, B. Milash, A. Rose, S. Widoff, et al. LifeLines: Visualizing Personal Histories. In *Proc. SIGCHI Conf. Hu. Fact. in Com. Sys.*, p. 221–227, 1996.

[81] M. Pohl, M. Schmitt, and S. Diehl. Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In *Computational Aesthetics'09: Proceedings of the Fifth Eurographics conference on Comp. Aesthetics in Graph., Vis. and Imaging*, pp. 49–56, 2009.

[82] R. Raidou, O. Casares-Magaz, A. Amirkhanov, V. Moiseenko, et al. Bladder runner: Visual analytics for the exploration of rt-induced bladder toxicity in a cohort study. *Comp. Graph. Forum*, 37(3):205–216, 2018.

[83] J. Rogers, N. Spina, A. Neese, et al. Composer—visual cohort analysis of patient outcomes. *App. Clin. Info.*, 10(2):278, 2019.

[84] D. I. Rosenthal, T. R. Mendoza, M. S. Chambers, J. A. Asper, et al. Measuring head and neck cancer symptom burden: The development and validation of the M. D. Anderson symptom inventory, head and neck module. *Head & Neck*, 29(10):923–931, 2007.

[85] D. I. Rosenthal, T. R. Mendoza, C. D. Fuller, K. A. Hutcheson, et al. Patterns of symptom burden during radiotherapy or concurrent chemoradiotherapy for head and neck cancer: A prospective analysis using the University of Texas MD Anderson Cancer Center Symptom Inventory-Head and Neck Module. *Cancer*, 120(13):1975–1984, 2014.

[86] H. M. Skerman, P. M. Yates, and D. Battistutta. Multivariate methods to identify cancer-related symptom clusters. *Res. in Nursing & Health*, 32(3):345–360, 2009.

[87] A. M. Smith, W. Xu, Y. Sun, J. R. Faeder, et al. RuleBender: integrated modeling, simulation and visualization for rule-based intracellular biochemistry. *BMC Bioinform.*, 13(8):1–16, 2012.

[88] M. Steenwijk, J. Milles, M. van Buchem, J. Reiber, et al. Integrated Visual Analysis for Heterogeneous Datasets in Cohort Studies. *IEEE VisWeek Workshop Vis. Analy. in Health Care*, 3, 2010.

[89] N. Steinhauer, M. Hörbrugger, A. D. Braun, T. Tüting, S. Oeltze-Jafra, et al. Comprehensive visualization of longitudinal patient data for the dermatological oncological tumor board. In *EuroVis 2020 Short Papers*. The Eurographics Association, 2020.

[90] J. Tosado, L. Zdilar, et al. Clustering of largely right-censored oropharyngeal head and neck cancer patients for discriminative groupings to improve outcome prediction. *Scientific reports*, 10(1):1–14, 2020.

[91] C. Turkay, A. Lundervold, A. J. Lundervold, and H. Hauser. Hypothesis Generation by Interactive Visual Exploration of Heterogeneous Medical Data. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, p. 1–12. Springer, 2013.

[92] F. Van Ham and B. Rogowitz. Perceptual organization in user-generated graph layouts. *IEEE Trans. Vis. Comp. Graph.*, 14(6):1333–1339, 2008.

[93] T. D. Wang, C. Plaisant, A. J. Quinn, et al. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, p. 457–466, 2008.

[94] T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, et al. Temporal summaries: Supporting temporal categorical searching, aggregation and comparison. *IEEE Trans. Vis. Comp. Graph.*, 15(6):1049–1056, 2009.

[95] C. Ware, E. Neufeld, and L. Bartram. Visualizing causal relations. In *Proc. IEEE Info. Vis.*, vol. 99, p. 39–42, 1999.

[96] A. Wentzel, G. Canahuate, L. van Dijk, A. Mohamed, et al. Explainable spatial clustering: Leveraging spatial data in radiation oncology. In *IEEE Vis. Conf.*, p. 281–285, 2020.

[97] A. Wentzel, P. Hanula, et al. Precision toxicity correlates of tumor spatial proximity to organs at risk in cancer patients receiving intensity-modulated radiotherapy. *Radiother. & Onco.*, 148:245–251, 2020.

[98] A. Wentzel, P. Hanula, T. Luciani, B. Elgohari, et al. Cohort-based T-SSIM Visual Computing for Radiation Therapy Prediction and Exploration. *IEEE Trans. Vis. Comp. Graph.*, 26(1):949–959, 2020.

[99] A. Wentzel, T. Luciani, L. V. van Dijk, et al. Precision association of lymphatic disease spread with radiation-associated toxicity in oropharyngeal squamous carcinomas. *Radiotherapy & Onco.*, 161:152–158, 2021.

[100] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and Intel. Lab. Sys.*, 2(1-3):37–52, 1987.

[101] K. Wongsuphasawat and D. Gotz. Outflow: Visualizing patient flow by symptoms and outcome. In *IEEE VisWeek Workshop Vis. Anal. in Health.*, p. 25–28. American Medical Informatics Association, 2011.

[102] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, et al. LifeFlow: Visualizing an Overview of Event Sequences. In *Proc. SIGCHI Conf. Hu. Fact. in Comp. Sys.*, p. 1747–1756, 2011.

[103] L. Zdilar, D. M. Vock, et al. Evaluating the effect of right-censored end point transformation for radiomic feature selection of data from patients with oropharyngeal cancer. *JCO Clin. Cancer Informatics*, 2:1–19, 2018.

[104] V. Zebralla, J. Müller, T. Wald, Boehm, et al. Obtaining patient-reported outcomes electronically with "oncofunction" in head and neck cancer patients during aftercare. *Front. in Onco.*, 10:2502, 2020.

[105] Z. Zhang, D. Gotz, and A. Perer. Iterative cohort analysis and exploration. *Info. Vis.*, 14(4):289–307, 2015.