Operationalizing Healthcare Big Data in the Electronic Health Records using a Heatmap Visualization Technique

Don Roosan  
*Western University of Health Sciences*

Mazharul Karim  
*Western University of Health Sciences*

Jay Chok  
*Keck Graduate Institute*

Moom Roosan  
*Chapman University, roosan@chapman.edu*

Follow this and additional works at: [https://digitalcommons.chapman.edu/pharmacy_articles](https://digitalcommons.chapman.edu/pharmacy_articles)

**Recommended Citation**


This Conference Proceeding is brought to you for free and open access by the School of Pharmacy at Chapman University Digital Commons. It has been accepted for inclusion in Pharmacy Faculty Articles and Research by an authorized administrator of Chapman University Digital Commons. For more information, please contact laughtin@chapman.edu.
Operationalizing Healthcare Big Data in the Electronic Health Records using a Heatmap Visualization Technique

Comments
This paper was originally published in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: HEALTHINF* in 2020. https://doi.org/10.5220/0008912503610368

Copyright
SCITEPRESS – Science and Technology Publications, Lda.
Operationalizing Healthcare Big Data in the Electronic Health Records using a Heatmap Visualization Technique

Don Roosan1, Mazharul Karim1, Jay Chok2 and Moom R. Roosan3

1College of Pharmacy, Western University of Health Science, Pomona, California, U.S.A.
2School of Applied Life Science, Keck Graduate Institute, Claremont, California, U.S.A.
3School of Pharmacy, Chapman University, Irvine, California, U.S.A.

Keywords: Electronic Health Record, Big Data, Visualization, Heatmap, Data Science.

Abstract: Background: The majority of the electronic health record (EHR) contains a wealth of information, including unstructured notes. Healthcare professionals may be missing substantial portions of essential diagnostic and treatment information by not focusing on unstructured texts. The objective of this study is to present progress notes data using heatmap visualization. Methods: In this study, the research team used the unstructured text from the progress notes of deidentified patient data. The research team conducted qualitative content-coding based on the clinical complexity model and developed a heatmap based on the processed frequency data. Result: The researchers developed a color-coded heatmap focusing on the severity and acuity of patients’ status accumulated through multiple previous patient’s visits. Conclusions: Future research into creating an automated process to generate the heatmap from an unstructured dataset can open up opportunities to operationalize big data in healthcare.

1 INTRODUCTION

The electronic health record (EHR) contains vital information about patients’ overall health. Much of this information is found in the unstructured notes taken by doctors, nurses, and other practitioners, making it easy to overlook. By ignoring the unstructured text, healthcare professionals may be missing a substantial amount of essential diagnostic and treatment information. Due to heavy workloads, healthcare professionals cannot afford to take the time to analyze and incorporate all the data available in a patient’s EHR from previous visits and admissions (Ben-Assuli, Shabtai, & Leshno, 2013; Lanham et al., 2014). Currently, more than the 80% of information in the EHR is disjointed and incoherent and not in a structured format, making it difficult for healthcare professionals to decipher and integrate it into their decision-making process (Thyvalikakath et al., 2014; Islam, Weir, & Del Fiol, 2014). Moreover, data are reaching “critical mass” in EHRs and should be reused in other ways, including in “quality improvement,” in the healthcare settings.

Several visualization techniques have been incorporated with decision-support systems to facilitate healthcare decision-making during treatment. However, visualization techniques for unstructured data are not widely used (Hersh, 2014). Also, medical and diagnostic errors are threats that the medical community cannot afford to ignore (Medford-Davis et al., 2015). Moreover, the lack of timely attention to diagnostic error can have dire implications for public health, as exemplified by the widely reported diagnostic error regarding Ebola virus infection in a Dallas hospital emergency department (ED) (Mandl, 2014). Diagnostic error is likely to be one of the most common types of errors in ED settings (Berner, 2009; Medford-Davis et al., 2015). The ED environment is high-paced and high-volume. It carries low-certainty in a multi-agent, dynamic and complex environment. These factors compound and may lead to diagnostic errors and

https://orcid.org/0000-0003-2482-6053
https://orcid.org/0000-0003-4412-1819
https://orcid.org/0000-0002-2082-0079
https://orcid.org/0000-0002-5318-5120
adverse events due to information loss. Thus, in an environment prone to interruptions like the ED, vital patient information and cues are often lost during information collection and integration among physicians, residents, nurses, and other healthcare providers (Carter, Davis, Evans, & Cone, 2009). This data loss is significantly due to a lack of time to adequately review the previous progress notes or visits, information that could potentially provide essential information.

Several attempts have been made to alleviate the burden posed by the amount and complexity of information available within EHR systems. Informatics and analytics have been proven to improve decision-making with the help of EHR data (Roosan, Law, Karim, & Roosan, 2019). To remedy problems such as documentation redundancy, neglect of crucial data, and difficulty navigating EHR software, prototype visualization tools have been tested to be effective (Carroll et al., 2014; Shneiderman, Plaisant, & Hesse, 2013). Various visualization tools, as simple as bar graphs and pie charts, can aggregate data visually. The problem at hand, however, is that large-scale multidimensional data are difficult to aggregate into these types of visualization tools. Therefore, researchers have been using more complex visualization tools such as parallel coordinates or heatmaps to assist with visualizing complex data (Islam, Weir, & Del Fiol, 2016).

To understand healthcare data complexity, it is essential to assess the factors related to both objective properties of the task and perceived task complexity (Liu & Li, 2012; Roosan et al., 2016). The objective properties of the task involve specific task characteristics, such as the number of decision steps or competing goals. On the other hand, perceived task complexity refers to the conjunct properties of the task and the characteristics of the task performer. When the task overwhelms the cognitive capacity of the task performer, the task is perceived to be complex by the task performer. Models of task complexity have been created in other research domains such as aviation and military to influence and predict human performance and behavior. In a previous study, the research team developed and validated a clinical complexity measurement model. This model integrates the patient CCFs proposed by Schaink et al. (2012) and task CCFs outlined by Liu and Li (2012). In the clinical complexity model, task complexity is conceptualized as having seven dimensions. Each dimension is then broken down into a subset of factors. For example, the dimension “ambiguity” (i.e., unclear, vague, or less specific clinical task components) consists of the factors “confusing information” (missing, ambiguous, or contradictory information cues) and “unclear goals” (objective is unclear or vague or less transparent or lacks specific goals). The patient complexity factors are divided into five dimensions, each of which is then broken down into several factors. For example, Mental Health relates to issues dealing with psychological stress, addiction/substance abuse, and related conditions.

Our research team applied this model to identify the specific complexity-contributing factors of clinical decision tasks to find the frequencies of particular complexity factors in the progress notes.

To build the clinical complexity model, Roosan et al. (2016) used the transcripts from a previous observational study to iteratively construct the measurement model. This model integrates the patient CCFs proposed by Schaink et al. (2012) and task CCFs outlined by Liu and Li (2012). In the clinical complexity model, task complexity is conceptualized as having seven dimensions. Each dimension is then broken down into a subset of factors. For example, the dimension “ambiguity” (i.e., unclear, vague, or less specific clinical task components) consists of the factors “confusing information” (missing, ambiguous, or contradictory information cues) and “unclear goals” (objective is unclear or vague or less transparent or lacks specific goals). The patient complexity factors are divided into five dimensions, each of which is then broken down into several factors. For example, Mental Health relates to issues dealing with psychological stress, addiction/substance abuse, and related conditions.

In this study, the research team used the same clinical complexity model to construct a heatmap of the progress notes data to highlight the severity and acuity of patients. We hypothesize that by using unstructured texts in the EHR, researchers can operationalize a significant proportion of currently available but unused healthcare big data. The objective of this study was to explore the feasibility of creating a heatmap to visualize data from the progress notes dataset.

2 METHOD

The research team consisted of pharmacy students, pharmacists, and academic researchers. The team conducted a secondary chart review using a large healthcare dataset. The research team decided to use progress notes for content-coding from the Neehr Perfect® program (“EHR Go,” 2019), which is an EHR that is built on VistA, the most widely used EHR in the world. Neehr Perfect® provided deidentified data for the pharmacy students so that they could get a realistic experience using the EHR. The program includes 170 patients’ charts ranging in complexity...
Table 1: Clinical complexity-contributing factors.

<table>
<thead>
<tr>
<th>Complexity contributing factors (CCFs)</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclear goals</td>
<td>The objective is ambiguous or vague, less clear or lacks specific goals</td>
</tr>
<tr>
<td>Large number of goals</td>
<td>Multiple goal elements, higher or larger number of goals</td>
</tr>
<tr>
<td>Conflicting goals</td>
<td>Achieving one goal has a negative effect or outcome on another goal</td>
</tr>
<tr>
<td>Confusing information</td>
<td>Unclear, missing, ambiguous or contradictory information cues</td>
</tr>
<tr>
<td>Unnecessary information</td>
<td>Large quantity of not useful information</td>
</tr>
<tr>
<td>Changing information</td>
<td>Unpredictable events, high rate of information change</td>
</tr>
<tr>
<td>Urgent information</td>
<td>Information about very acute patient situation</td>
</tr>
<tr>
<td>Multiple decision-making options</td>
<td>Large number of options to make a decision</td>
</tr>
<tr>
<td>Large number of decision steps</td>
<td>More than two steps or actions to attain the objective</td>
</tr>
<tr>
<td>Decision conflict</td>
<td>Two or more actions that are incompatible or competing, conflict between task components</td>
</tr>
<tr>
<td>Lack of expertise</td>
<td>Unique situation requiring additional knowledge, novel and non-routine decisions, treatment or disease uncertainty</td>
</tr>
<tr>
<td>Lack of team coordination</td>
<td>Coordinating activities and creating shared decision-making within and between healthcare teams</td>
</tr>
<tr>
<td>Time pressure</td>
<td>Situations that need immediate attention due to scarcity of time</td>
</tr>
<tr>
<td>Polypharmacy</td>
<td>Patient receiving medications from more than one pharmacy</td>
</tr>
<tr>
<td>Significant physical illness</td>
<td>Multiple chronic conditions, loss of physical functioning</td>
</tr>
<tr>
<td>Mental anxiety</td>
<td>External factors creating cognitive stress (e.g., job, culture, family)</td>
</tr>
<tr>
<td>Psychological illness</td>
<td>Depression, mood disorders, losing self-consciousness</td>
</tr>
<tr>
<td>Addiction/substance abuse</td>
<td>Drug or substance abuse in the past or present</td>
</tr>
<tr>
<td>Older age</td>
<td>Patient age 75 and older</td>
</tr>
<tr>
<td>Health disparity</td>
<td>Patients with a different ethnic background or cultural barrier with limited access to healthcare</td>
</tr>
<tr>
<td>Noncompliant patient</td>
<td>Patient not following medication or treatment regimen, difficulty communicating with providers</td>
</tr>
<tr>
<td>Poverty and low social support</td>
<td>Poor social support, low quality of life due to economic strains and lower social status</td>
</tr>
<tr>
<td>Heavy utilization of healthcare resources</td>
<td>Complex chronic patients with multiple care providers and institutions require more resources</td>
</tr>
<tr>
<td>Difficulty with healthcare system navigation</td>
<td>Low understanding of healthcare system, limited healthcare literacy</td>
</tr>
</tbody>
</table>

and type. For this study, the research team initially selected three complex patient charts, finally selecting the most complex chart. Researchers identified charts in which patients had more than three diagnoses for inpatient admission and at least 20 or more visits to inpatient settings. Once the three patients’ charts were selected, the team selected one complex case that included more than 20 visits from the same patient. Two clinical pharmacists verified that the case was complex. Researchers used the complex patient’s chart and transferred the data to a The researchers used Atlas.ti Version 8.0 software to
code the data. The study was exempted by the Claremont College IRB board as the dataset team used was acquired from the Neehr Perfect® program, which included only deidentified information. We conducted qualitative content coding of the dataset based on the clinical complexity factors from the clinical complexity model (Islam, Mayer, & Clutter, 2016; Islam et al., 2015). The factors are listed in Table 1.

The data analysis was based on content analysis (Roosan et al., 2016; Stemler, 2001). Specifically, the team members followed the “emergent coding” process of content analysis (Haney, Russell, Gulek, & Fierros, 1998). In this process, researchers independently review a subset of the data and form a checklist for coding. After independently coding, the research team meets to discuss and reconcile the differences. Once the coding has reached the desired level of reliability, it is applied to the remainder of the data. For the transcriptions of the interviews, the research team used the RATS (relevance of the study, appropriateness of qualitative method, transparency of procedure, and soundness of interpretive approach) protocol for qualitative data analysis (Clark, 2003). This protocol provides standardized guidelines for qualitative research methods.

Two students parsed the sentences to meaningful content (Table 2), and three other students coded the only one code was applied to each parsed sentence. After each coding session, the three students met to examine coding disagreements and to revise codes and code definitions. The interrater reliability, Cohen’s kappa, was calculated to be 0.83. A final Excel file was developed consisting of the frequencies of the total of 49 CCFs from the 21 visits recorded on a complex patient chart. Utilizing these frequency tables, researchers plotted and visualized the data in R “pheatmap V0.2” package to develop the heatmap.

3 RESULTS

The research team constructed the visual heatmap from the aggregated data, as described in Figure 1. Researchers plotted patient visits on the X-axis and clinical complexity variables on the Y-axis. The unique feature of this visualization tool is the use of color-coding based on severity: dark blue means fewer frequencies and dark red exemplifies higher frequencies. Values range from 0 to 1, with 0 indicating the patient displayed baseline clinical complexity variables and 1 when the frequencies of the different complexity variables were high. The research team selected a “spectral” visualization color scheme to show variation in the frequency with multiple colors with adequate depth. The complexity factors on the Y axis provide the information complexity within the heatmap on which the practitioner should focus. The dark red color in the sentences based on the clinical complexity factors. Heatmap represents the high presence of complexity factors in a visit. For example, the team was looking for vital patient could information for this complex patient in the charts that provide better insight into the severity and acuity of the case.

Looking at this heatmap, one can find that several visits had significant physical illnesses and changing information. These instances corresponded to the progress notes when the patient developed sepsis several times during previous visits. As a result, the patient became resistant to several antibiotics.

Table 2: Unitized texts of the transcript from the patient’s clinical notes.

<table>
<thead>
<tr>
<th>Unitized texts</th>
<th>Associated codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>“The patient has several immediate needs such as stabilizing high blood pressure, taking care of blood transfusion high blood glucose levels and mental health issues. I am not sure where researchers should focus more. I think the blood pressure should be a priority but I am still confused.”</td>
<td>Conflicting goals</td>
</tr>
<tr>
<td>“Researchers’ kind of think using Vancomycin should be able to take care of most of the infections even though team do not have the lab results. Researchers may wait for it but patient’s situations may get worse.”</td>
<td>Decision conflict</td>
</tr>
<tr>
<td>“There are quite a few other options as well. For example, azithromycin or clindamycin.”</td>
<td>Multiple decision-making options</td>
</tr>
<tr>
<td>“The patient was readmitted from a previous infection in this thigh. I am not sure if the patient received appropriate antibiotics during discharge and if he actually received it or not.”</td>
<td>Confusing information</td>
</tr>
<tr>
<td>“But the cellulitis in his thigh is getting worse and that is more what I would be worried about. I don’t know if it is from his previous wound or not.”</td>
<td>Changing information</td>
</tr>
</tbody>
</table>
Such information can be vital and time-sensitive in a moment of urgency, allowing clinicians to focus on finding more appropriate antibiotics for the multidrug resistance organisms rather than using first-line antibiotic, which otherwise could have failed. The lack of access to this information can lead to worse patient prognosis or result in patient death.

In this example, the team normalized the score of frequencies and used the values of 0 to 1 for visualization purposes. Complexity factors with 0 values do not indicate that the frequencies do not exist, only that the frequencies were the same as reference or baseline. A value of 1 shows the highest presence of clinical complexity factors for the patient across all visits.

The color-coded heatmap can assist clinicians in determining a more focused plan for the patient’s next visit. In addition, the researchers hope to incorporate filtering tools into the heatmap to further assist clinicians. Being able to filter the heatmap by time or area of interest can help clinicians understand the severity and acuity of the patient. In the heatmap in Figure 1, researchers focused on the different complexity factors to understand the specific activities that occurred on a particular visit for the patient. For example, the dark red on visits 11, 12, 16, and 17 for changing information refers to the many activities occurring during these visits. Examinations of the patient chart for those days revealed that the patient had recurrent infections and was admitted several times to the hospital. Obtaining this vital life-saving information early in treatment may help the admitting clinician choose between different antibiotics or therapy options. Moreover, knowing in advance about a patient’s previous recurrent infection can also help with assessing risk for readmissions and determining alternative antibiotics as appropriate.

4 DISCUSSION

Previous studies have used different visualization techniques for specific datasets. For example, some studies have visualized public health datasets to predict the progression of infection or individual disease states (Elliott et al., 2012). However, due to the digitization of healthcare data, a robust technique is needed to understand the meaningful information hidden in different visits for the patient. Specifically, the complexity of the information can help us understand patient readmission to the hospital. In this study, researchers have contributed by developing an innovative heatmap technique to understand the complexity of clinical progress notes.
The study adds a unique perspective in the EHR design for future designers and researchers. Currently, very few mechanisms exist to help health professionals utilize large amounts of unstructured texts in the EHR. Visualizing such information can help clinicians focus on crucial pieces of information that otherwise might be ignored. In this study, the heatmap researchers created provides a unique overview of 21 visits in a very complex patient case consisting of hundreds of pages. Using the heatmap, researchers can easily visualize multiple visits and identify more critical visits.

Currently, very few healthcare programs utilize a heatmap to visualize patient data across visits. Problems need to be accurately represented via a heatmap in order to craft proper policy (Ulmer, McFadden, & Nerenz, 2009). The goal of this study was to provide a tool that can help clinicians visualize data from patient progress notes, allowing them to identify and access specific visits to understand the severity and acuity of a patient’s illness or injury. The functionality of heatmaps includes the filtering of aggregated data that can be used to help clinicians narrow the possible sources of a problem that a patient may have. This filtering can help clinicians focus on what can be improved to ensure the patient receive high-quality care.

Workload issues are causing critical problems in the healthcare industry. Provider burnout and fatigue due to the digitization of healthcare are causing new errors (Kwekkeboom, Abbott-Anderson, & Wanta, 2010; Saber Tehrani et al., 2013). Providers are overburdened with the extra work of using digital health tools when they should be taking care of patients (Huang, Tobin, & Tompane, 2012). Many clinicians are leaving the field or moving to part-time jobs due to the extra workload (Rahman, 2016). Therefore, system designers need to use innovative visualization techniques that are not disruptive of workflow and that support the clinician’s cognition. The constant addition of new information into the EHR makes it difficult for clinicians to realize where the most critical information is buried. Our technique sheds light on dealing with this problem using this innovative heatmap visualization. This visualization improves the overall understanding of healthcare information for patients as well as their clinicians (Roosan et al., 2019).

The heatmap visualization of EHR data has several implications for big data. Currently, the amount of data generated in an EHR is voluminous. This poses a challenge for clinicians who need to review this data in a short time. Each admission and subsequent visit generate more than 100 data points. During readmissions, hospital staff commonly review the patient’s previous visits. However, the information may be buried among hundreds of lines of data, and clinicians often have no clue about which visits they should focus. Using the heatmap approach, they may be able to identify a specific visit that holds the information they need. The approach may help administrators prioritize patients for discharge and focus on the more complex patients for better care.

Center for Medicare and Medicaid Services (CMS) currently does not reimburse for 30-day readmission for patients. However, if a patient has a history of readmissions and the clinician learns of this history by focusing on the pertinent information with the help of the heatmap, then he or she will be able to prioritize therapy for the patient.

Researchers assume that the analytics of the heatmap need to be integrated with the EHR. For example, clinicians need to be able to click on the specific heatmap to view the days related to the visit. Also, specific search options, such as using a text search, can help clinicians. Many EHRs already have text search options. Previous studies have used heatmaps mostly for understanding multidimensional genomics datasets (Gu, Eils, & Schlesner, 2016; Rahman et al., 2017; Ramirez, Dündar, Diehl, Grüning, & Manke, 2014; Shen, Olslen, & Ladanyi, 2009; Zhu et al., 2009). However, using a heatmap to visualize this unstructured text from clinical notes is a new concept introduced in this study.

In this study, the research team has created a heatmap of a single patient visit using qualitative content-coding and operationalizing the clinical complexity model. Future software or algorithms using machine learning and artificial intelligence can learn the content-coding process and automate the visualization process to create the heatmaps. Such a process can help not only clinicians but also patients who want to make sense of information in their health records. In the current age of health information digitization, meaningful and life-saving relevant information must be at the fingertips of clinicians at the point of care. Future studies with actual EHR data can further validate the process outlined in this study.

The study has several limitations. Researchers created the heatmap based on deidentified data from Neehr Perfect®, Data in the real world may be missing data points and may create more noise in the heatmap. Also, the heatmap was not validated by clinicians for usefulness. However, the assumption is that clinicians will benefit from such visualization-based decision support within the EHR. In this study, the research team did not do any usability evaluations of the heatmap visualization to understand how clinicians
may perceive such tools. The heatmap was enthusiastically received by most of the providers, but it is not known how well it will be integrated within the clinical workflow. The manual coding may have introduced another form of bias. However, to reduce this bias, two distinct coders coded independently, and the inter-rater reliability was high.

5 CONCLUSIONS

The large amount of unstructured text data in the EHR provides a challenge for clinicians to focus on the information necessary for diagnosis and optimal therapy options. In this study, we used qualitative content-coding to visualize progress notes information to focus on patient visits that have crucial information. Using a clinical complexity model in this study, the research team visualized unstructured data from the EHR. By focusing and shifting attention for providers to the right information, the heatmap visualization technique may have the potential to reduce providers’ cognitive fatigue and information overload. Future research into creating a machine learning approach to automate this process can support and operationalize big data in healthcare.

ACKNOWLEDGEMENTS

This work was supported by an internal grant from Western University of Health Sciences, College of Pharmacy.

REFERENCES


