A Smart Healthcare Portal for Clinical Decision Making and Precision Medicine

Joseph J. Nalluri, Khajamoinuddin Syed, Pratip Rana, Paul Hudgins, Ibrahim Ramadan, William Nieporte, William Sleeman IV, Jatinder Palta, Rishabh Kapoor, Preetam Ghosh[*]

ABSTRACT

There has been an unprecedented generation of healthcare data at clinical practices. With the availability of advanced computing frameworks and the ability to electronically mine data from disparate sources (e.g. demographics, genetics, imaging, treatment, clinical decisions, and outcomes) *big data research* in medicine has become a very active field of interest. In this paper, we discuss the challenges associated with designing clinical decision support systems that try to leverage such disparate data sources and create smart healthcare tools to aid medical practitioners for better patient care and treatment plans. We next propose an integrated data curation, storage and analytics portal, called HINGE (the *Health Information Gateway and Exchange* application), that can effectively address many of the outstanding challenges in this domain. HINGE specifically caters to healthcare data from radiation oncology patients however, the underlying formalisms and principles, as discussed here, are readily extendible to other disease types making it an attractive tool for the design of next generation clinical decision support systems.

CCS CONCEPTS

• **Information systems → Decision support systems**; • **Applied computing** → *Health care information systems*; Health informatics;

1 INTRODUCTION

The adoption of Electronic Health Record (EHR) systems and Health Information Exchanges (HIEs) by healthcare providers has generated massive amounts of data. This data is entered on a daily basis with an intent to improve patient care. Most of this information is a byproduct of routine care delivery. EHR systems store all information of a patient from different departments of clinical care into a centralized repository leading to *Big Data* in healthcare. The ability to use this data for clinical decision support at the point of care and to effectively generate new knowledge constitutes the major promise of big data in precision medicine and smart healthcare. However, poor, incomplete and mis-representative data will generate poor results ("garbage in, garbage out"). An important component of EHRs are the electronic medical records (EMRs) which are usually unstructured, and entered into the EMR database with an intent of patient management and not for the purposes of data analytics. Hence, it is often difficult to access or share this data because of privacy/information security concerns. As such, *big data* research requires a lot of preprocessing before it can lead us to the clinically relevant answers for complex diseases. This is particularly true for oncology, as cancer is the second leading cause of death globally, being responsible for 8.8 million deaths in 2015 [1]. In 2016, an estimated 1,685,210 new cases of cancer were diagnosed in the United States alone with 595,690 deaths [2]. Implementation of smart healthcare systems needs a balance between the clinical domain, the clinical need/question and the organizational mandate[3].

Big data in the domain of oncology is divided into four categories: clinical, dosimetry, imaging and biological [4]; they characterize all the five V's of *big data*, namely volume, velocity, variety, veracity and value. The main challenges with oncological healthcare data is the lack of robust standards to describe the nuances and complexity of cancer data and EHR interoperability. In the field of radiation therapy (RT), there is an urgent need to obtain population based comparative effectiveness data on all radiotherapy techniques [5, 6]. Most patients with cancer receive treatment based on results of controlled randomized clinical trial (RCT) studies performed on hundreds of similar patients [7]. Randomized clinical trials are used to generate the highest level of evidence and outcome evaluation. Data from these clinical trials has been internally validated in a controlled environment to the extent that we can trust the relationship between the treatment and outcome. However, since these clinical trials are specific to a particular study, they take years to accrue patients, are expensive to run, and tend to be done at a smaller number of academic institutions and hence their practicality is very limited [5, 8]. Because of the heterogeneous nature of patients, varying comorbidities and socioeconomic status, it becomes practically impossible for these clinical trials to address population based treatment effectiveness. Population based studies are designed to include all samples and evaluate effectiveness of real-world clinical practice. There are many large databases such as Surveillance, Epidemiology and End Results (SEER) program established by the National Cancer Institute (NCI) in 1973 [9] and Center of Medicare and Medicaid Services (CMS) that collect data on a large number of cancer patients treated over time. The data in this database include demographics, cancer incidence, clinical and survival factors but fail to include detailed clinical and treatment information. Data analysis from the SEER database suggests that the database lacks information on the radiation dose, technique and radiation therapy receipt [10].

To address the need for an accurate, comprehensive dataset and to determine clinical practice variations, outcomes, gaps in

[*] JJN and KS contributed equally.

JJN, WS, JP and RK are affiliated with the Department of Medical Physics, Virginia Commonwealth University. Email: {joseph.nalluri, william.sleemaniv, jatinder.palta, rishabh.kapoor}@vcuhealth.org.

KS, PR, PH, IR, WN and PG are affiliated with the Department of Computer Science, Virginia Commonwealth University. Email: {lnusk,ranap,hudginspj,ramadinig,nieportewm,pghosh}@vcu.edu.

PRE-PRINT
Conference'18, January 2018, Varanasi, India

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

treatment quality, and to compare the effectiveness of various treatment modalities, we propose the *Health Information Gateway and Exchange (HINGE)* application. HINGE collects patient-specific radiotherapy data in electronic form to visualize the following: (i) gaps in patient care, (ii) comparison between treatment modalities, (iii) effectiveness of treatment and (iv) linking treatment outcomes. It provides benchmark data and quality improvement tools for individual providers.

In the following section, the role of decision support systems for various clinical domains is reviewed in Section 2. Section 3 describes the proposed architecture and framework for developing HINGE for the field of oncology; along with the challenges encountered and solutions adopted. Section 4 summarizes the benefits of the proposed methodology and lists the challenges that need to be addressed. Section 5 reviews the extendibility of the application to other areas and discusses the necessity and the adoptability of future clinical decision systems in various healthcare domains.

2 BACKGROUND AND RELATED WORK

Several research groups and vendors of the healthcare industry have worked towards developing such integrated systems/platforms.This system would serve as — an aid to the physicians in their clinical decision-making process, allow for healthcare institutes to assess their practices and treatment outcomes, help the research and development teams to create predictive algorithms, and provide opportunities to policy makers for new learning and decisions. Broadly, we can divide such expert systems into two groups — knowledge-based and non-knowledge based[11].

Knowledge based systems often follow the principle of creating a knowledge base (data curation and modeling), creating an inference engine (logic rules, connection rules, business intelligence metrics) and mechanism of user interactions (displays, user-input systems). These systems are very domain-specific and require a close participation from domain experts in the process of creating computerized inference rules that resemble human reasoning. Esposito et al. developed a fuzzy logic based decision system that would assess the diagnosis and disease progression of patients suffering from multiple sclerosis over their life time. The domain knowledge was quantified and represented in terms of linguistic variables, linguistic values and membership functions, in cooperation with clinical experts. Thereafter, the medical decision-making process was formalized in terms of *if-then* rules and deployed in a differential evolution model to determine the highest correct classification rate [12]. A similar system was also designed to support the diagnosis of meningitis by representing the domain knowledge into a graphical model and fuzzy relationships between concepts [13]. Riano et al. designed an ontology-based decision support system for patients suffering from chronic illness. This comprised of a formal translation of chronic care related concepts and relationships into an ontology. A personalized ontology was created by adapting the ontology to the specifics of the health record of the patient. This personalized ontology was used to further personalize intervention plans consisting of general treatments, into patient-specific intervention plans [14].

On the other hand, non-knowledge based systems often deploy a hypothesis-driven artificial learning or machine learning based method to learn from the clinical information consisting of raw data. Lin et al. designed a system using artificial neural networks to create an inference model to predict the outcomes of kidney transplantation. This would allow the physician to statistically provide the patient with a predicted survival rate in kidney transplantation [15]. Amaral et al. used several machine learning algorithms — linear Bayes, K nearest neighbors, decision trees and support vector machines (SVMs) and compared them to determine the best classifier for the diagnosis of chronic obstructive pulmonary disease [16]. Cho et al. used linear SVMs to predict the onset of diabetic nephropathy among patients with diabetes. The proposed classifier was built using features from an irregular and unbalanced dataset and could predict the onset of diabetic nephropathy 2-3 months before the actual diagnosis with high accuracy [17]. SVMs were also used by Chi et al. to design a hospital referral system which recorded several parameters relating to patients' demography, disease complication, and risk factors; and predicted a recommendation where the patient's probability of achieving desired outcome is maximized [18].

Clinical notes are one of the largest venues of healthcare data on patient symptoms and can be found in electronic health records (EHR). A substantial portion of patient information and clinical information is stored as narrative text (dictated or transcribed). Linguistic String Project was one of the earliest medical NLP system that was developed [19] and later evolved into the Medical Language Processor that included healthcare lexicon, medically tagged lexicon, and mapping of medical information into formalized structures. Pakhomov et al. designed an NLP architecture in collaboration between Mayo Clinic and IBM, that could identify clinically relevant entities among clinical notes [20]. Matheny et al. designed a hybrid model consisting of rule based algorithms and NLP to detect infectious symptoms from clinical notes. Their model could identify symptoms for tuberculosis, acute hepatitis and influenza which were previously unidentified [21].

3 METHODOLOGY

With the goal of improving cancer care by collecting reliable information, we have designed the HINGE data analytics platform that collects data from the radiotherapy delivery systems, such as the linear accelerators, treatment management systems (TMS), patient reported outcome systems and electronic medical record systems. A high-level architecture reflecting the basic framework of the proposed system is shown in Figure 1.

1. **Clinical practice tier**: It encompasses all the IT and medical device systems pertaining to different phases of the patient treatment. Treatment planning systems (TPS) contain information about a prescribed radiation therapy treatment plan by physicians and dosimetrists. The TMS systems use these plans as input and deliver the radiation to the patient. These individual systems are often proprietary softwares that record and document this information
2. **Aggregation tier**: It collects the various pieces of patient information from the clinical practice tier and organizes it into a standardized template by integrating the information coherently; as per the defined data taxonomy and data dictionary created by the clinical domain experts.

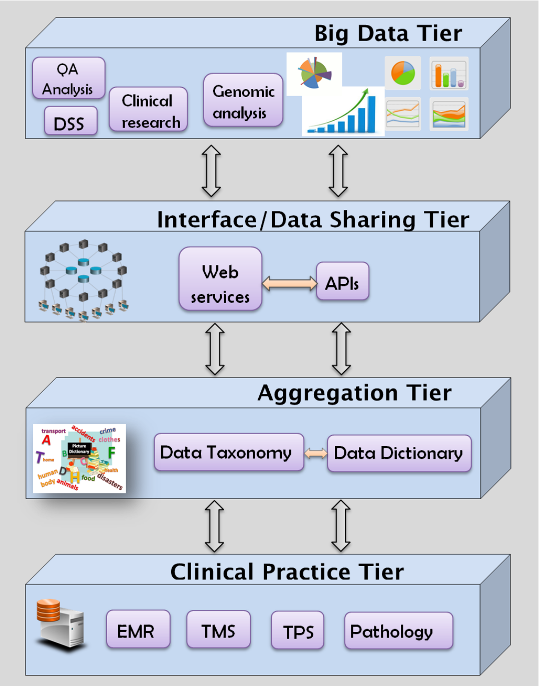

Figure 1: Smart tools for *Big Data* research in Healthcare

(3) **Interface/data sharing tier**: It contains resources for end-users to interact with the patient information. Herein, the physicians and clinical staff have a venue to view, edit, verify and share the entire patient information comprising all records.

(4) **Big data tier**: It contains several analytical tools, such as decision support system (DSS), QA analysis, genomic analysis and so on for model based predictions. These tools can perform smart and predictive analyses on single patient information, as well as across a cohort of several patients or disease cases. These tools are also able to extract clinical parameters, define data elements/features and construct statistical models to answer various treatment outcomes and research hypotheses.

The implementation of the aforementioned tiers and architecture pose several computational and technical challenges which are discussed in the following section; along with their adopted solutions.

3.1 Defining quality measures and data elements

With the aim to improve the quality of cancer care, we worked with our professional society (American Society of Radiation Oncology) to establish clinical measures by which individual care can be assessed and compared with national practice. Such clinical measures were based on established clinical guidelines, expert consensus, results from clinical trials and formed the basis for assessing the quality of treatments and practice variations and identification of the care gaps. These clinical measures are based on the work done by the American College of Radiology's Quality Research in Radiation Oncology program [22].

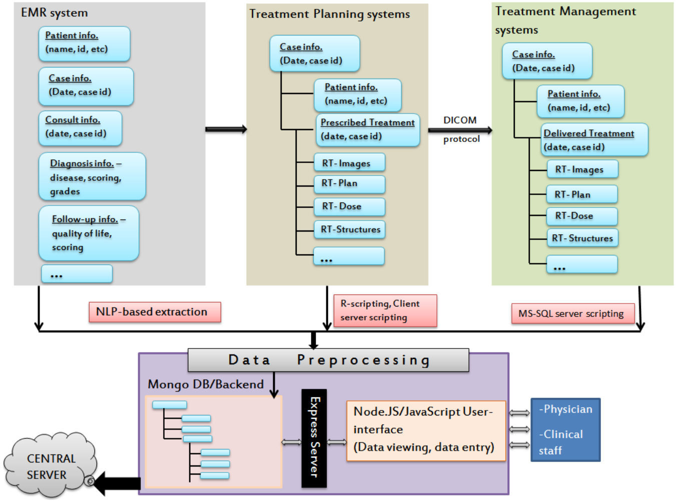

Figure 2: HINGE system design

Evaluation of data from the radiation oncology clinics based on these clinical measures identified the potential areas of improvement for the practitioners. It also provides benchmark data that allows the practitioners to assess the quality of care in their own practice and comparing individuals with national practice. In simple terms, these clinical measures provide a means to quantify and improve quality of care by providing a feedback loop back to the practice using their own data. Based on these clinical measures we defined the corresponding data elements and their sources. A brief description of these data elements across different systems are shown in Figure 2

While the lack of quality measures and the subsequent data elements has been addressed in Radiation Oncology, domain experts from other medical fields must develop their own metrics.

3.2 Standardization of data syntax and semantics

Clinical systems that currently collect clinical, treatment and process data are disparate. The lack of uniformity in data syntax and semantics makes it extremely difficult for data aggregation and analysis. Unfortunately, there are not many ontologies with radiation oncology specific terms. One of the major tasks we addressed was to develop a common syntax to provide the semantics associated with data collection and interpretation. The NCI-sponsored Common Data Element Dictionary is a source of standard terminology used in clinical trials but does not contain many of the radiation oncology specific data elements [23]. This standardization for the fields and terms used in the EMR, treatment procedures and workflow processes have the potential to increase the quality and comparability of the data used to create future models. We have used the most commonly used ontologies, such as AJCC [24], CTC AE [25, 26], AAPM TG-263 [27].

3.3 Extraction of data from clinical notes

Radiation Oncology providers use a spectrum of different documentation methods such as handwritten notes, dictating note contents into a speech dictation device and transcribing notes on a computer in the electronic medical record systems. In the process of recording clinical notes the providers record observations, impressions, toxicities, plans of care and describe the episodes of care, and these records are generally created at each interaction between the patient and the healthcare provider. Despite various NLP based approaches to structure this data and learn from the free text, none of the approaches can be applied uniformly across the board to all types and styles of free-text data [28]. NLP has been widely used as a filter to augment manual data abstraction efforts but fully automated extraction with high levels of accuracy is still an area of active research [29, 30].

The challenge here is twofold. The first challenge is in setting up data entry templates at the clinics for prospective data entry and to simplify the subsequent data extraction process for analytics. The second challenge is to deploy NLP-based techniques to extract data elements from historical notes for the existing patient data at the local clinics. We designed standardized disease site specific clinical note templates (CNT) in the EMR for the management of patients receiving radiotherapy at any of the radiation oncology services. All clinical data such as staging, risk group, diagnostic test results, performance scores and toxicities are recorded in a structured format. We used all predefined radiotherapy ontologies and defined additional ones where no standard data definitions existed. Automatic calculation of assessment scores and graphical indication of treatment progress are rendered in these templates. The software provides functionality to export abstracted data elements from the templates to a central database server. These templates also have free-text areas where the physicians can dictate their assessments. With this approach, our templates support the hybrid documentation model where relevant structured data and free-text narrative data are entered for appropriate sections in the template. With the help of these templates, clinical data entered by our physicians in the EMR is abstracted into a HINGE aggregating application.

In order to extract data from existing and historical patient data, we used NLP techniques based on *rule-based extraction*

Rule Based Extraction

Based on the quality measures described in 3.1, four data elements were considered for initial NLP analysis. The data elements were prostate-specific antigen (PSA) level, clinical tumor stage (T stage), gleason score (GS) and National Comprehensive Cancer Network (NCCN) risk group.

Each of these data elements were mentioned with high variability in clinical notes. Rule-based extraction methods require a lot of manual work nevertheless, for extraction of complicated, structured templates, formal rule-based approach is best suited as they produce more reliable results. Following rules (regular expressions) were applied to extract the four fields mentioned above for specifically prostate cancer patients; similar rules were then formed for other cancer types.

(1) **Prostate-Specific Antigen (PSA) level**: It measures the amount of prostate-specific antigen in the blood. PSA is released into the blood by the prostate gland
Keyword: PSA
Extraction rules: (a) strings followed by keyword; (b) numeric string followed by "ng/mL"
Post-processing check: Value is not in date format, "undetectable" or "0,0.10"
Ignore: punctuations, "value, score, rose, rising, of, to, was, is, = , at"
Instruction: Use maximum value if multiple scores were found

(2) **Gleason score**: Indicates how likely a tumor would spread on a range of 2 to 10. A low Gleason score means the cancer tissue is similar to normal prostate tissue and the tumor is less likely to spread; a high Gleason score would mean the vice-versa
Keyword: gleason, gleason score, gleason grade, gs
Extraction rules: (a) strings followed by keyword; (b) strings "primary + secondary =" or "primary + secondary = total" or "primary + secondary equals total" or,"primary plus secondary equals total"
Post-processing check: Value is numeric; score to be added when only "primary" and "secondary" is found
Instruction: Use maximum value if multiple scores found

(3) **T stage**: Its the doctor's best estimate of the extent of disease spread, based on the results of the physical exam (including DRE), lab tests, prostate biopsy, and any imaging tests
Keyword: T
Extraction rules: Occurrence of keyword and "1a" or "1b" or "1c" or "2a" or "2b" or "2c" or "3b" or "3c" or "4"

(4) **NCCN risk**: Physicians calculate the NCCN Risk using PSA, T stage and Gleason score into four categories, very low risk, low risk, intermediate and high
Keyword: risk
Extraction rule: strings "very low", "low", "intermediate" and "high" followed by keyword

Similar techniques would be used to extract further measures from the notes.

3.4 Abstraction of data from Treatment Management and Planning systems

Existing treatment management and planning systems are primarily designed to optimize the clinical workflow and therefore lack the utilities required to facilitate *big data* applications. Information from current TMSs and TPSs cannot be stored in EMRs because there is no complete provision for such data fields in the EMR; hence it needs to be specifically curated. Creating a complete picture of a patient's treatment requires the aggregation of data from multiple systems including data formats such as DICOM-RT[31, 32], free text notes, and SQL database entries. Once data is extracted from these systems, the relevant data elements must be identified which is especially difficult with free text notes as they can vary greatly between different physicians, facilities, or software systems. Each vendor has its own proprietary software which has to be taken into consideration while writing scripts to fetch the data. The calculation of a composite radiation treatment dose is something often non-trivial with the existing software systems [33–36]. If a patient is treated with two or more courses of treatment, or if the single course is modified during treatment, the resulting dose delivered from these courses must be combined to provide the total dose that was actually delivered. This compose dose requires the planning images for each treatment course and the corresponding calculated doses. Delivery information from the treatment management system must be combined with calculated doses by the planning system in order to determine how much radiation was actually delivered in total. One of the features of the HINGE application is to aggregate this treatment information to generate the complete treatment dose.

Python scripts with MS-SQL server and R-scripts were written to extract the information from the treatment management and planning systems, respectively. These scripts would extract the relevant DICOM data (images and RT-treatment plans) and compute the dose calculations and dose-volume histograms (DVH). An *R* package called *RadOnc*[37] was used for this purpose.

3.5 Application development

The design of HINGE application is shown in Figure 2. There are three major data sources for this application - EMR, TPS and TMS. After performing 3.1, 3.2, data was extracted via programming scripts into a JSON style document model. The following were the characteristics of the data,

1. the nature of the data was unstructured and inconsistent
2. the extracted data elements had an inherent hierarchical structure and multiplicity to it. Many of the data elements, such as, *follow-up*, *quality of life*, *staging information* had interlinked dependencies among them as well as one-to-many type relationships.
3. Since the majority of the data elements were extracted from clinical notes using NLP techniques, they required a manual verification process during which editing/deleting data would be the norm.
4. The existence of data elements among every system was not assured, resulting in unbalanced data.
5. The database had to have an adaptability for adding new fields, forms or additional information with ease.

Owing to these challenges, we designed a *smart* database schema that allows data manipulation easily across the entire database. We created built-in utilities that perform data consistency check (hierarchical structure), data validation (correct information type), data merge and data import functionalities. In case of importing data, the data import utility can automatically determine the functionality of either an *insert*, *update* into the database based on the set of schema. The schema also enforces the interlinked dependencies between data elements during import of data (new and old). The current schema also allows a piecemeal style of data import in real time while retaining the consistency.

MongoDB was selected as the database for HINGE. The user interface was designed using JavaScript and Node.js technologies with *Express* as the server on a Linux operating system.

3.6 Deployment of data extraction tools and decision support systems

The HINGE application is to be deployed at several clinics. Concerns of information security must be addressed and each clinic may have a completely different combination of software platforms and hardware resources. Even the same health system may have different EMR, treatment management, and planning systems. The following aspects needs to be addressed,

3.6.1 *Anonymization module.*

The HINGE application will act as primary data aggregators and would anonymize patient data before sending them to the main central server for analysis. If there is a change in the software systems at a clinic, only the aggregator needs to be updated and the central analysis database and tools will remain untouched.

3.6.2 *Storage.*

Every remote clinic annually generates approximately 400 gigabytes of data. The central enterprise server generates about 20 terabytes of data, annually. This is dependent upon the cancer types to be diagnosed, the profile of patients addressed (e.g., tumor vs palliative), the frequency of the visits, etc. An additional concern regarding the infrastructure has to be addressed when data backup/redundancy is considered over a period of time.

3.6.3 *Client-server setup.*

Schedules will be set up during appropriate times during the day for data transfer between remote clinics to central server. Database triggers can be set up at the central server to recompute the analyses based on newly added data.

3.6.4 *Quality assurance.*

After data is extracted from the relevant clinical systems, it must meet a quality assurance standard. Some data elements, such as dates, names, and cancer stage, may exist in some form over multiple data sources and the correct values must be correctly identified. In other cases, expected values may be missing and it is important to determine if the missing data was never entered or if it was likely missed by the data extraction system. High quality, reliable data is critical because it is required for any decision support system.

4 DISCUSSION AND FUTURE WORK

By collating population based treatment delivery and health outcome information, our proposed HINGE platform serves as a great tool for the physicians, healthcare providers, vendors, policy makers, researchers and hospital administrators. A platform such as HINGE is uniquely poised and capable to address some of the unanswered questions in predicting radiation therapy outcomes of cancer patients. The data from HINGE can be used to create decision support models in the clinical systems that enable clinicians to improve the quality and safety of care rendered to the patients. To realize this goal, other than the above-mentioned challenges, there are various other technical challenges that need to be addressed.

- Design of unique interfaces where clinician and patient reported data is captured with minimal interaction with the capturing database.
- Definition and utilization of disease specific ontologies in medicine.

- Use of innovative tools to capture digital data directly from the patients. Standardization and establishment of trust models to share data between multiple healthcare providers, i.e., in other words "make data free".
- Reduction of data housed in disparate databases and linking databases with genomic, treatment, claims and clinical databases.
- Develop clinical decision support and machine learning tools that enable medical researchers and clinicians to analyze data and perform predictive analytics.

5 CONCLUSION

The proposed HINGE platform is an endeavor towards integrating the clinical practice, data acquisition, automating and streamlining clinical workflow in order to derive clinically meaningful measures of care rendered to patients. Although, the HINGE platform is a service to the domain of radiation oncology, the principles adopted and deployed as mentioned in 3, can be readily extended to different clinical domains. For big data and smart healthcare techniques to succeed in precision medicine, it is imperative that all stakeholders - physicians, physicists, nurses, clerks and commercial vendors work together on *how* and *what* data needs to be collected. Every clinical domain has to define their own workflow templates, clinical measures, domain-specific models, metrics, data features, standardization schemes and data analysis tools that can be used as a feedback to the practice for quality improvement and answer the challenging questions of their healthcare domain.

REFERENCES

[1] http://www.who.int. World Health Organization, 2017.

[2] https://www.cancer.gov/. national Cancer Institute, year = 2017, url = https://www.cancer.gov/about-cancer/understanding/statistics/, urldate = 11-15-2017.

[3] Paolo Fraccaro, Panagiotis Plastiras, Chiara Dentone, Antonio Di Biagio, Peter Weller, et al. Behind the screens: Clinical decision support methodologies–a review. *Health Policy and Technology*, 4(1):29–38, 2015.

[4] Issam El Naqa. Biomedical informatics and panomics for evidence-based radiation therapy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(4):327–340, 2014.

[5] Jatinder R Palta, Jason A Efstathiou, Justin E Bekelman, Sasa Mutic, Carl R Bogardus, Todd R McNutt, Peter E Gabriel, Colleen A Lawton, Anthony L Zietman, and Christopher M Rose. Developing a national radiation oncology registry: From acorns to oaks. *Practical radiation oncology*, 2(1):10–17, 2012.

[6] Charles S Mayo, Marc L Kessler, Avraham Eisbruch, Grant Weyburne, Mary Feng, James A Hayman, Shruti Jolly, Issam El Naqa, Jean M Moran, Martha M Matuszak, et al. The big data effort in radiation oncology: Data mining or data farming? *Advances in Radiation Oncology*, 1(4):260–271, 2016.

[7] Rebecca Siegel, Deepa Naishadham, and Ahmedin Jemal. Cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, 62(1):10–29, 2012.

[8] Justin E Bekelman, Anand Shah, and Stephen M Hahn. Implications of comparative effectiveness research for radiation oncology. *Practical radiation oncology*, 1(2):72–80, 2011.

[9] Seer.cancer.gov. Epidemiology, and end results program, 2017.

[10] Reshma Jagsi, Paul Abrahamse, Sarah T Hawley, John J Graff, Ann S Hamilton, and Steven J Katz. Underascertainment of radiotherapy receipt in surveillance, epidemiology, and end results registry data. *Cancer*, 118(2):333–341, 2012.

[11] Robert A Greenes. *Clinical decision support: the road to broad adoption*. Academic Press, 2014.

[12] Massimo Esposito, Ivanoe De Falco, and Giuseppe De Pietro. An evolutionary-fuzzy dss for assessing health status in multiple sclerosis disease. *International journal of medical informatics*, 80(12):e245–e254, 2011.

[13] Vijay K Mago, Ravinder Mehta, Ryan Woolrych, and Elpiniki I Papageorgiou. Supporting meningitis diagnosis amongst infants and children through the use of fuzzy cognitive mapping. *BMC medical informatics and decision making*, 12(1):98, 2012.

[14] David Riaño, Francis Real, Joan Albert López-Vallverdú, Fabio Campana, Sara Ercolani, Patrizia Mecocci, Roberta Annicchiarico, and Carlo Caltagirone. An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients. *Journal of biomedical informatics*, 45(3):429–446, 2012.

[15] Ray S Lin, Susan D Horn, John F Hurdle, and Alexander S Goldfarb-Rumyantzev. Single and multiple time-point prediction models in kidney transplant outcomes. *Journal of biomedical informatics*, 41(6):944–952, 2008.

[16] Jorge LM Amaral, Agnaldo J Lopes, José M Jansen, Alvaro CD Faria, and Pedro L Melo. Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease. *Computer methods and programs in biomedicine*, 105(3):183–193, 2012.

[17] Baek Hwan Cho, Hwanjo Yu, Kwang-Won Kim, Tae Hyun Kim, In Young Kim, and Sun I Kim. Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods. *Artificial intelligence in medicine*, 42(1):37–53, 2008.

[18] Chih-Lin Chi, W Nick Street, and Marcia M Ward. Building a hospital referral expert system with a prediction and optimization-based decision support system algorithm. *Journal of biomedical informatics*, 41(2):371–386, 2008.

[19] Naomi Sager, Margaret Lyman, Christine Bucknall, Ngo Nhan, and Leo J Tick. Natural language processing and the representation of clinical data. *Journal of the American Medical Informatics Association*, 1(2):142–160, 1994.

[20] Serguei Pakhomov, James Buntrock, and Patrick Duffy. High throughput modularized nlp system for clinical text. In *Proceedings of the ACL 2005 on Interactive poster and demonstration sessions*, pages 25–28. Association for Computational Linguistics, 2005.

[21] Michael E Matheny, Fern FitzHenry, Theodore Speroff, Jennifer K Green, Michelle L Griffith, Eduard E Vasilevskis, Elliot M Fielstein, Peter L Elkin, and Steven H Brown. Detection of infectious symptoms from va emergency department and primary care clinical documentation. *International journal of medical informatics*, 81(3):143–156, 2012.

[22] Jean B Owen, Julia R White, Michael J Zelefsky, and J Frank Wilson. Using qrroâĎć survey data to assess compliance with quality indicators for breast and prostate cancer. *Journal of the American College of Radiology*, 6(6):442–447, 2009.

[23] nim.nih.gov. Common Data Element Dictionary, 2017.

[24] American Joint Committee on Cancer. American joint committee on cancer.

[25] Common Terminology Criteria for Adverse Events (CTCAE). Common terminology criteria for adverse events (ctcae).

[26] Ethan Basch, Stephanie L Pugh, Amylou C Dueck, Sandra A Mitchell, Lawrence Berk, Shannon Fogh, Lauren J Rogak, Marcha Gatewood, Bryce B Reeve, Tito R Mendoza, et al. Feasibility of patient reporting of symptomatic adverse events via the patient-reported outcomes version of the common terminology criteria for adverse events (pro-ctcae) in a chemoradiotherapy cooperative group multicenter clinical trial. *International Journal of Radiation Oncology* Biology* Physics*, 98(2):409–418, 2017.

[27] AAPM TG 263 Standardizing Radiation Therapy Nomenclature. Aapm tg 263 - standardizing radiation therapy nomenclature.

[28] S Trent Rosenbloom, Joshua C Denny, Hua Xu, Nancy Lorenzi, William W Stead, and Kevin B Johnson. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *Journal of the American Medical Informatics Association*, 18(2):181–186, 2011.

[29] David S Carrell, Scott Halgrim, Diem-Thy Tran, Diana SM Buist, Jessica Chubak, Wendy W Chapman, and Guergana Savova. Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. *American journal of epidemiology*, 179(6):749–758, 2014.

[30] Kaihong Liu, William R Hogan, and Rebecca S Crowley. Natural language processing methods and systems for biomedical ontology learning. *Journal of biomedical informatics*, 44(1):163–179, 2011.

[31] Maria YY Law and Brent Liu. Dicom-rt and its utilization in radiation therapy. *Radiographics*, 29(3):655–667, 2009.

[32] Maria YY Law, Brent Liu, and Lawrence W Chan. Dicom-rt–based electronic patient record information system for radiation therapy. *Radiographics*, 29(4):961–972, 2009.

[33] Joel Poder, Johnson Yuen, Andrew Howie, Andrej Bece, and Joseph Bucci. Dose accumulation of multiple high dose rate prostate brachytherapy treatments in two commercially available image registration systems. *Physica Medica*, 43:43–48, 2017.

[34] Di Yan, Frank Vicini, John Wong, and Alvaro Martinez. Adaptive radiation therapy. *Physics in medicine and biology*, 42(1):123, 1997.

[35] Samuel B Park, James I Monroe, Min Yao, Mitchell Machtay, and Jason W Sohn. Composite radiation dose representation using fuzzy set theory. *Information Sciences*, 187:204–215, 2012.

[36] EH Balagamwala, T Djemil, SA Koyfman, ST Chao, L Angelov, JH Suh, and P Xia. 3d composite dose is necessary to assess cumulative spinal cord dose for retreatment of spinal tumors with stereotactic body radiotherapy. *International Journal of Radiation OncologyâĂć BiologyâĂć Physics*, 81(2):S647, 2011.

[37] Reid F Thompson. Radonc: an r package for analysis of dose-volume histogram and three-dimensional structural data. *Journal of Radiation Oncology Informatics*, 6(1):98–110, 2014.