

# Chapter 69

## The Role of Big Data in Radiation Oncology: Challenges and Potentials

Issam El Naqa  
McGill University, Canada

### ABSTRACT

*More than half of cancer patients receive ionizing radiation as part of their treatment and it is the main modality at advanced stages of disease. Treatment outcomes in radiotherapy are determined by complex interactions between cancer genetics, treatment regimens, and patient-related variables. A typical radiotherapy treatment scenario can generate a large pool of data, “Big data,” that is comprised of patient demographics, dosimetry, imaging features, and biological markers. Radiotherapy data constitutes a unique interface between physical and biological data interactions. In this chapter, the authors review recent advances and discuss current challenges to interrogate big data in radiotherapy using top-bottom and bottom-top approaches. They describe the specific nature of big data in radiotherapy and discuss issues related to bioinformatics tools for data aggregation, sharing, and confidentiality. The authors also highlight the potential opportunities in this field for big data research from bioinformaticians as well as clinical decision-makers’ perspectives.*

### INTRODUCTION

Cancer is a leading cause of mortality in the United States and worldwide. It remains the second most common cause of death in the United States, accounting for nearly 1 of every 4 deaths. It is projected that a total of 1,665,540 new cancer cases and 585,720 cancer deaths are to occur in the United States in 2014 (Siegel, Ma, Zou, & Jemal, 2014). Radiation therapy (radiotherapy) is one of three major treatment modalities of cancer beside surgery and chemotherapy and it remains the main option at locally advanced stages of the disease. More than half of all cancer patients receive radiotherapy as a part of their treatment. Despite radiotherapy proven benefits, it comes with a Damocles’ sword of benefits and risks to the exposed patient. A key goal of modern radiation oncology research is to predict, at the time

DOI: 10.4018/978-1-4666-9840-6.ch069

of radiation treatment planning, the probability of tumour response benefits and normal tissue risks for the type of treatment being considered. Although recent years have witnessed tremendous technological advances in radiotherapy treatment planning, image-guidance and delivery, efforts to individualize radiotherapy treatment doses based on *in vitro* assays of various biological endpoints has not been clinically successful (IAEA, 2002; C. M. West, 1995). Conversely, several groups have shown that dose-volume factors play an important role in determining treatment outcomes (Bentzen et al., 2010; Blanco et al., 2005; J. Bradley, Deasy, Bentzen, & El-Naqa, 2004; J. O. Deasy & El Naqa, 2008; I. El Naqa et al., 2006; Hope et al., 2005; Jackson et al., 2010; Marks, 2002a; Tucker et al., 2004), but these methods may suffer from limited predictive power when applied prospectively. The lack of progress or major breakthroughs in radiotherapy outcomes over the past two decades demands fundamentally new insights into the methodologies used to better exploit the power of this unique non-invasive high-energy source, as well as it requires a new vision to guide the analysis of radiotherapy response and the design of new therapeutic strategies.

Outcomes in radiotherapy are usually characterized by tumour control probability (TCP) and the surrounding normal tissues complications (NTCP) (Steel, 2002; Webb, 2001). Traditionally, these outcomes are modeled using information about the dose distribution and the fractionation scheme (Moissenko, Deasy, & Van Dyk, 2005). However, it is recognized that radiation response may also be affected by multiple clinical prognostic factors (Marks, 2002b) and more recently, inherited genetic variations have been suggested as playing an important role in radiation sensitivity (J. Alsner, C. N. Andreassen, & J. Overgaard, 2008; C. M. L. West, Elliott, & Burnet, 2007). Moreover, evolution in imaging and biotechnology have provided new extraordinary opportunities for visualizing tumours *in vivo* and for applying new molecular techniques for biomarkers discovery of radiotherapy response, respectively (Bentzen, 2008; Jain, 2007; Nagaraj, 2009; C. M. L. West et al., 2007). However, biological assays, which can be performed on either tumour or normal tissues, may not be the only determinant of tumour control or risk of radiotherapy adverse reactions. Therefore, recent approaches have utilized data driven models using advanced informatics tools in which dose-volume metrics are mixed with other patient or disease-based prognostic factors in order to improve outcomes prediction (Issam El Naqa, 2012). Accurate prediction of treatment outcomes would provide clinicians with better tools for informed decision-making about expected benefits versus anticipated risks.

In this chapter, we will provide an overview of recent advances in radiotherapy informatics and discuss current challenges to interrogate big data as it appears in radiotherapy using top-bottom and bottom top-approaches. We will describe the specific nature of big data in radiotherapy, and the role of the emerging field of systems radiobiology for outcomes modeling. We will provide examples based on our and others experiences. Finally, we will discuss issues related to bioinformatics tools for data aggregation, sharing, and confidentiality.

## **BACKGROUND**

### **Radiotherapy of Cancer**

Radiotherapy is targeted localized treatment using ablative high-energy radiation beams to kill cancer cells. More than half of all cancer patients, particularly patients with solid tumours such as in the brain, lung, breast, head and neck, and the pelvic area receive radiotherapy as part of their curative or palliative

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/the-role-of-big-data-in-radiation-oncology/150228](http://www.igi-global.com/chapter/the-role-of-big-data-in-radiation-oncology/150228)

## Related Content

---

### Analyzing Twitter User-Generated Content Changes

Luca Cagliero, Luigi Grimaudo and Alessandro Fiori (2013). *Data Mining in Dynamic Social Networks and Fuzzy Systems* (pp. 87-109).

[www.irma-international.org/chapter/analyzing-twitter-user-generated-content/77524](http://www.irma-international.org/chapter/analyzing-twitter-user-generated-content/77524)

### Influence of Domain and Model Properties on the Reliability Estimates' Performance

Zoran Bosnic and Igor Kononenko (2009). *International Journal of Data Warehousing and Mining* (pp. 58-76).

[www.irma-international.org/article/influence-domain-model-properties-reliability/37405](http://www.irma-international.org/article/influence-domain-model-properties-reliability/37405)

### Discovering Frequent Embedded Subtree Patterns from Large Databases of Unordered Labeled Trees

Yongqiao Xiao and J. F. Yao (2005). *International Journal of Data Warehousing and Mining* (pp. 70-92).

[www.irma-international.org/article/discovering-frequent-embedded-subtree-patterns/1752](http://www.irma-international.org/article/discovering-frequent-embedded-subtree-patterns/1752)

### Analyzing Social Emotions in Social Network Using Graph Based Co-Ranking Algorithm

Kani Priya, Krishnaveni R., Krishnamurthy M. and Bairavel S. (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 329-341).

[www.irma-international.org/chapter/analyzing-social-emotions-in-social-network-using-graph-based-co-ranking-algorithm/308495](http://www.irma-international.org/chapter/analyzing-social-emotions-in-social-network-using-graph-based-co-ranking-algorithm/308495)

### The Healthcare Cost Dilemma: What Health Insurance Companies Can Do to Mitigate Unsustainable Premium Increases

Russ Danstrom and Jeff Nicola (2004). *Managing Data Mining: Advice from Experts* (pp. 124-145).

[www.irma-international.org/chapter/healthcare-cost-dilemma/24782](http://www.irma-international.org/chapter/healthcare-cost-dilemma/24782)