International Journal of Engineering Research ISSN:2319-6890)(online),2347-5013(print)
Volume No.3, Issue No.5, pp : 309-314 01 May 2014

Optimization of Lung Cancer using Modern Data Mining Techniques

T. Sowmiya, M. Gopi, M. New Begin, L.Thomas Robinson
Department of Master of Computer Applications
Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College
Avadi,Chennai,Tamil Nadu

sowmismailbox@gmail.com, mailtogopim@gmail.com, newbegin_m@yahoo.com, son.mca@gmail.com

Abstract :
Now a day the most dangerous diseases in the world are Cancer. Lung cancer is one of the most dangerous cancer types in the world. These diseases can spread worldwide by uncontrolled cell growth in the tissues of the lung. Early detection of the cancer can save the life and survivability of the patients who affected by this diseases. In this paper we survey several aspects of data mining procedures which are used for lung cancer prediction for the patients. Data mining concepts is useful in lung cancer classification. We also reviewed the aspects of ant colony optimization (ACO) technique in data mining. Ant colony optimization helps in increasing or decreasing the disease prediction value of the diseases. This case study assorted data mining and ant colony optimization techniques for appropriate rule generation and classifications on diseases, which pilot to exact Lung cancer classifications. In additionally to, it provides basic framework for further improvement in medical diagnosis on lung cancer. Our Proposed idea for the lung cancer optimization on data mining is by using the (ROCO) method. We use reduced-order constrained optimization (ROCO) to create clinically acceptable IMRT plans quickly and automatically for advanced lung cancer patients Diagnosis. Our new ROCO implementations works with the treatment planning system and full dose calculations used at Memorial Sloan-Kettering Cancer Center for diagnosis, and we have implemented the mean dose hard-constraints on cancer, along with the point-dose and dosage-volume constraints that we used for our previous work on the prostate.

Keywords:
ACO, data mining, rule pruning, ROCO

INTRODUCTION

Lung cancer is a most dangerous disease which is due to uncontrolled cell growth in tissues of the lung. If the Lung cancer is not treated in the earliest stage, this growth can spread beyond the lungs in a process called metastasis into nearby tissue cells and, eventually, into various parts of the body rapidly. Most lung cancers which are in the primary stage are carcinomas that derive from epithelial cells of the body. Common causes of lung cancer are tobacco and smoking. It is the main cause of lung cancer deaths, and it is so difficult to detect in its starting stages because symptoms can show their properties at advanced stages sometimes in the final stages. There is several research and prediction methods suggest that the early detection of lung cancer will decrease the mortality rate. Decision classification is the most important task for mining any data set. The problems are classified as mainly collaborated with the assignment of an object to an object oriented parameters that is class and its parameter [1], [2]. There are several decision tasks which we observe in several industries like engineering, medical, and management's related science can be considered as classification problems on these issues. Best examples are pattern classifications for the problems, speech recognition patterns, character recognition patterns, medical diagnosis and credit Scoring.

But in our case view classification alone is an insufficient for classifying the lung cancer dataset. If we consider the data mining for frequent patterns of classification then it will be the better tool for classifying relevant data types from the raw datasets. The best performance of association rule is directly depend on the frequent pattern mining, to balance the core problems of the mining association rules [3]. With the developing and more numbers of detailed research on frequent item pattern mining, it is elaborately used in the field of data mining concepts, for example, mining association rules, correlation analysis, classifications, clustering data 4],vector machine[5] and positive Association rule classification [6].

The main goal of data mining is to extract the importantdetails from huge amount of raw source data. We emphasized to mine lung cancer data to discover knowledge that is never to be only correct, but also comprehensible for lung cancer detections [7], [8], [9].

Comprehensibility is most important whenever discovered knowledge will be always used for supporting a human'sfinal decision. After all, if discovered knowledge is not comprehensible for a user data, it will be not possible to interpret and validates the knowledge data. So hence we can say that having trust in discovering the rule on knowledge is more important. In Final decision making, this can lead to wrong decisions. We provide here an overview on medical data mining techniques. The rest of this paper is arranged as follows: Section 2 introduces and describes about the optimization technique ANT Colony optimization; Section 3 describes about related works; section 4 discuss about the Theoretical extraction. ; Section 5 Proposed idea on the optimization of lung cancer Section 6 describes Conclusion.

2. ANT COLONY OPTIMIZATION

The Ant Colony Optimization (ACO) algorithm is a more meta-heuristic which is a grouping of the distributed environments,

positive feedback systems, and systematic greedy approaches to find an optimal solution value for combinatorial optimization problems on lung cancer. The Ant Colony Optimization algorithm is mainly inspired by the various types of experiments & treatment plans run by Goss et al. [19] which using a grouping the real ants in the real environments. They studied and observed the behaviors of those real ants and suggest that the real ants were having capability to choose and select the shortest path between their shelter and food products resource, in the existence of alternate paths between the two. The above searching for food products resource is possible through an indirect communications known as stigmergy amongst of the ants. When ants are travelling for the food Resources, ants deposit a new type of chemical substances, called pheromone. When they arrive at a closing point; ants make a probability on choices, biased by the intensity of Pheromone they smell. This behavior has an autocatalytic effect because of the very fact that An ant choosing a path will increase the probability that the corresponding path will be chosen Again by other ants in the next move of the. After finishing the search ant's returns back, the Probability on choosing the same path is higher because of increasing pheromone quantity. So the pheromone will released on the chosen way, it provides the new way to the ants. We can say that, all ants will select the shortest path. Figure 1 shows the behavior of ants in a double bridge experiment [20]. If we analyze the case then we observed that because of the same pheromone laying the shortest Path will be taken. It will be starts with the first ants which arrive at the food source are those that took the two shortest branches of the path. After approaching the food destination point these ants start. Ants return trip, was more pheromone is present on the shortest branch is the possibility for choosing the shortest one than the one on the Long Branch. This ant behavior was first formulated and arranged as Ant System (AS). Based on the AS algorithm, the Ant Colony Optimization (ACO) algorithm was proposed [22]. In ACO algorithm, the optimization problem can be expressed as a formulated graph G = (C; L), where C is the setoff components of the problem which is given, and L is the set of main possible connections or transitions among the element values of C. The proposed solution is mainly represented in terms of feasible paths on the given graph G, with respect to a given constraints and predicate the values of the population. The population numbers of ants that is also been called agents collectively solves the problem under consideration using the graph representation terms. We have assumed that the ants are only probably poor of finding a perfect solution, good fine quality solutions can emerge as a final result of collective interaction among the ants. Pheromone trailsthe encode, big long-term memory about the whole ant search process from the starting to the food source destination point. The values basically depend on problem formulation functions, representation and the optimization objective values which is different in case to case.

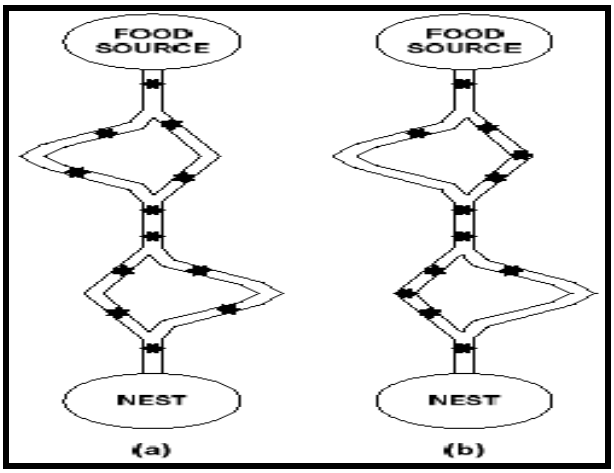


Figure 1: Double bridge experimentation. (a) Ants start exploring the dual bridge. (b)Usually most of the ants choose the shortest path [20].

The algorithm plagiaristic by ScholarDorito was given below:

Algorithm ACO Metaexperiential ();
While (termination criterion not satisfied)
Ant generation and action ();
Pheromonevanishing ();
Inspirationactions ();
"Optional "
End while
End Algorithm

3. LITERATURE REVIEW

In 2011, Shyi-Ching Liang et al. [28] suggest Cataloging rule is the most common representation of the rule in data mining. It is based on controlled learning process which causes rules from drill data set. The main goal of the cataloging rule mining is the prediction of the predefined class based on the collection. Based on ACO procedure, Ant-Miner solved the arrangementlawproblematic. According to the author, Ant-Miner shows good presentation in many dataset. In this research paper author future, an extension of Ant-Miner is proposed to integrate the concept of parallel dispensation and alliance. In this paper intercommunication is provided via pheromone among ants is a critical part in ant colony optimization's pointeddevice. The algorithm design in such a way, with a slight adjustment in this part which removes the parallel pointed capability. Based on Ant-Miner, they propose an addition that modifies the algorithm design to incorporate parallel processing. The pheromone trail deposited by ants during the searching technique affected each other. With the help of pheromone, ants can have better decision making while searching. They provide a possible direction for researches toward the grouping rule problem.

Pros of ACO:-

- Fast, resolutions of practical quality

Cons of ACO:-

- Solution may be far from optimal
- Make only limited number of different solutions
- Choices made at early stages reduce a set of possible steps at latter stages

PROPOSED SOLUTION:

Introduction of ROCO (Reduced Order Constrained Optimization)

Intensity-modulated radiotherapy (IMRT) has revolutionized the treatment of cancers in the last decade: it allows a higher dose to be delivered to a tumor while protecting nearby radiation-sensitive normal local tissues, yielding a better local control and fewer post-treatment complications than previous techniques However, the process of obtaining a clinically acceptable IMRT plan for a difficult behavior site is often slow and intensive, requiring treatment hours of expert period in a physical trial-and-error hoop in which the parameters of the optimization score function are continuously adjusted. Big long planning times place a severe stress on available resources in a busy medical clinic, and can result in to a treatment delays, acceptance of a neighbor sub-optimal plans or, in the given worst case, errors due to the time pressure. In this paper, we have apply a new method called reduced-order constrained optimization (ROCO) to greatly reduce the amount of time required to obtain a clinically main acceptable IMRT plan. By minimizing the trial-and-error method effort representative of recent IMRT design, it allows the medical treatment planners to focus on clinical tradeoffs between tumor coverage and normal body organ doses. We have already previously applied ROCO to prostate gland cancer cases4; in this paper, we have improved our applications of ROCO and reports new results on a more challenging treatment area, the lung.

Lung cancer accounts for the most cancer-related deaths in both men and women in the United States of America. An estimated calculation nearly 157,300 deaths, accounting for about 28% of all cancer deaths, are expected to occur in 2010. Radiation therapy is the one of the most main curative treatment for inoperable non-small cell lung cancer (NSCLC), but it remains a technically challenging procedure with very low 5-year survival rates (<10%) 6. IMRT is promising for treatment of NSCLC compared to traditional radiotherapy or 3D-CRT since it may enable dose escalation to the tumor7; however, the organs at risk (OARs) are sensitive to the radiation, including lungs, esophagus, and the spinal cord. Since the exact sizes andcorrect locations of lung cancers are diverse, unlike prostate cancer, a standard multi-field class solution for IMRT application is not used. Typical treatment plans for the locally advanced (stage III) lung e cancer feature prescription doses of 1.8–2 Gy/fraction delivered by 3–5 coplanar treatment beams of 6 MV photons, occasionally with the addition of the non-coplanar beams and rays. Hence for the locally main advanced NSCLC cases we examine in this paper, we have estimate that it takes an perfect expert planner around 3 hours to create a clinically acceptable IMRT plan (not counting time spent contouring structures and selecting beam directions). In this paper, we describe our implementation of ROCO, which we have incorporated with the medical treatment planning system at Memorial Sloan-Kettering Cancer Center (MSKCC), and our results from retrospective application of ROCO to 12 locally-advanced lung cancer cases. The anonymized clinical data (image sets, structure contours, and clinical treatment plans) for these patients were provided by MSKCC under IRB approval. ROCO consists of various steps, after beam ray directions have been selected. Firstly, random set value of score function parameters are chosen via Latin hypercube sampling or trialing, and then these plans are optimized for using the clinical score-function-based optimization. Second, principal component analysis (PCA) isolates the important modes of variation in the intensity matrices values, which helps to shifts the independent variables of the given problem to the few dominant PCA modes. Sampling and PCA models are mainly generated for each patient individually, not as group solutions. Hence the third 50 step is hard-constrained optimization. Naturally reduction by PCA makes it feasible to rapidly and automatically locate plans with clinically acceptable PTV coverage and normal tissue protection in the space spanned by the trial plans. Using the MSKCC planning system, the full or overall process takes approximately 30 minutes per patient. Advanced lung cancer cases present new challenges when compared to our previous work on the prostate gland. For prostate gland cases, because the relationship between the PTV and the rest of the anatomy varies relatively little from the one patient to other patient, the same beam directions were used for each patient. Stage III lung cancer tumors, on the other hands, show extremely and exactly variable geometries and can growth to the considerable size, growing outside of the lung parts proper and into the mediastinum; additionally, single or multiple lung tumors can appear or present in a various variety of geometries near OARs such as the heart, the esophagus, the spinal cord, and the brachial plexus. Because of this, ROCOapplication was be used the clinical beam directions which were chosen by the planner in every case. The current implementation is integrated part with the medical clinical MSKCC treatment planning system in order to make it flexible and stretchable enough to deal with the different type of treatment sites besides the prostategland, Hence whereas the software's were previously described used the data which is exported from, and performed calculations to outside of, the treatment planning system4 (which caused difficulty because of discrepancies in dose calculation).

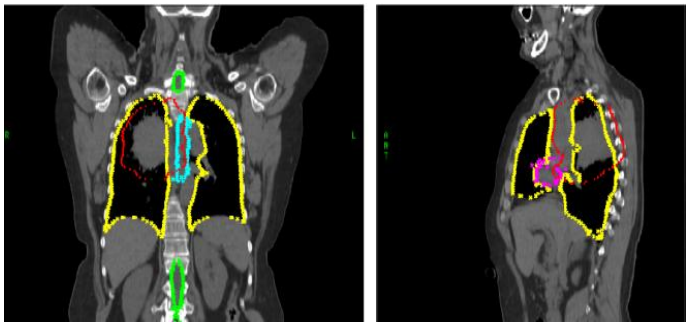


Fig. 1 shows a single CT image slice of a representative lung cancer patient, with contours for the various OARs, together

ISSN:2319-6890)(online),2347-5013(print)
01 May 2014

with a 3D representation the CT images showing the tumor wrapping around the esophagus. The dimension of the space of possible 4.

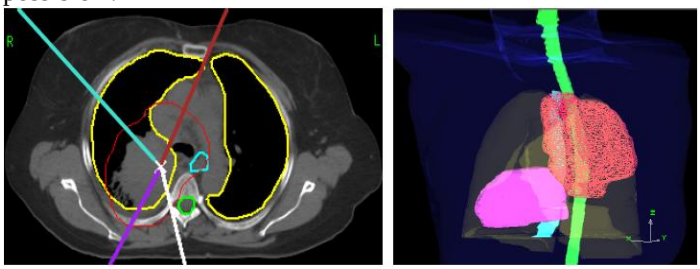


FIG. 1. The left three panels show CT image slices in the treatment plane for patient #8 in our study; the rightmost panel shows a beam's eye view of the same patient's case. The PTV is shown in red color, the lungs in yellow color, and the spinal cord in green color, the heart in pink color, and the esophagus in cyan. The solid lines in the third panel show the beam directions. Treatments are so larger for these locally-advanced lung cases than for prostate gland cases, because the larger treatment fields contain a greater big total number of beamless. For prostate gland cases there are on the order of the 103 beamless, and for the lung cancer cases that we consider, there are about 104beamless. Finally, IMRT for NSCLC often includes "rind" structures to prevent hot spots in nonspecific normal local tissues. Table I summarizes the major differences pertaining to treatment planning between the prostate gland cases we had to considered previously and the stage III NSCLC cases considered in this paper.

In practice, the unconstrained optimizations require a deal of heuristic trial and error method to arrive at the parameter settings such that the resulting plan is medical clinically acceptable. The given planner will uses the weights (or "importance factors") in the objective function to try to "steer" the optimization algorithm to more clinically desirable solutions12, but this can be difficult since the process of adjusting these weights is inherently imprecise and unintuitive. The role of dose limits in IMRT optimization is also puzzling, since it has been pragmatic that in

Criterion	Prostate Case	Lung Case
Beams (geometry)	5 (class solution)	4–9
Beamlets	$\sim 10^3$	$\sim 10^4$
Median PTV volume	~ 160 cm^3	~ 380 cm^3
PTV/OAR relationships	Similar	Variable
Non-specific normal tissue sparing	Beam arrangement	"Rind" structures
Optimization parameters	~ 30	~ 50
OARs	3–5	5–10

TABLE I. Evaluation of IMRT management planning complication in prostate and lung treatments. Hot spots in non-specific normal tissues around the prostate are avoided by beam arrangements and little PTV size, so rinds are not generally required.

An unconstrained optimization, dose limits more stringent than the clinical limits are required to obtain convergence to an acceptable plan (see, e.g., 10, 14, 15). The inverse planning process of obtaining a clinically acceptable IMRT plan for a difficult site can take several hours, normally96 due to the manual process of adjusting the parameters in the objective function10, 13, and 16. In our previous work17, we implemented sensitivity analysis to identification key parameters of an unconstrained IMRT objective function that have a strong impact on the resultant dose allocation. We then applied an external loop over the perceptive parameter set to find the parameters such that the minimizer of the corresponding objective function gave the best score of a scalar function of sketch quality. While this method hurriedly produced plans that generally satisfied the clinical constraints, it still suffered from (1) using a scalar-valued objective function to approximate a fundamentally hard-constrained problem, and (2) re1quiring training data to identify the susceptible set, assuming a generalizable class solution for the management site. The ROCO algorithm has neither limitation.

While hard-constrained optimization for IMRT planning has been proposed previously (e.g., using mixed-integer programming18), it is usually prohibitively time-consuming due to the huge dimensionality of the problem and the difficulty in implementing the dosage-volume constraints. Another recent focus of interestis Multi Objective (MO) optimization technique , which allows the planner to choose from a family of Pareto-optimal plans (that is, plans in which no criterion can be improved without worsening the others) The ROCO algorithm makes constrained optimization computationally tractable using four steps:

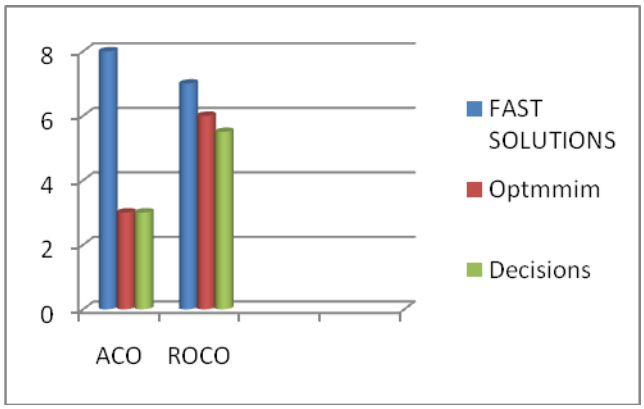
1. Select the targets and OARs to be included in the score function, and then select the beams whose strengths are to be optimized.
2. Randomly sample sets of score job constraints, apply the medical optimization to both set, and store the resultantstrengthdesigns.
3. Apply principal component analysis (PCA) to this set of passion summaries. The ensuingmajormechanisms form a basis for the space of plans that contains the optimal plan.
4. Compute the coefficients of the basis vectors that optimize goal analysis, subject to medical restrictions.

Pros of ROCO Algorithm: -

- Solution may be far from optimal
- The Explanation will be best from Optimum,
- It will helps to work in more type of limitless numbers in different solutions
- Conclusions made at early stages reduce a set of possible steps at latter stages

Statistics of ACO & ROCO:-

ISSN:2319-6890)(online),2347-5013(print)
01 May 2014



CONCLUSION:

The usage of data excavating methods in Lung cancer classification growths the chance of making a correct and first finding, which could prove to be vigorous in struggling the disease? In this paper, we offer a review on lung cancer revealing. We also examines the effectiveness of data mining by which we can find the effective lung cancer detection techniques. Next exploration we find several orderings algorithm and their result by which we can find the future visions. As the area of Lung cancer is very puzzling and the researchers are continuing their research progress in effective recognition, there is lot of scope in the case of capable recognition. As per our surveillance there are some upcoming suggestions which are listed below:

1) We can apply neural system and Fuzzy established method to trainCancer data set for finding better grouping and correctness.

2) We can apply optimization system like Ant Colony Optimization to improve the classification [33] for improving the detection.

3) Machine knowledge setting or Care Vector machine [32] is alsoAn insight for better detection.

4) We can use some similarity based algorithm to find over fitting andOvergeneralizationFeatures. It can be applied by clustering algorithm like Means.

5) We can use the work on ROCO in several vital ways. First, we functional ROCO to a more problematical treatment position: the lung rather than the prostate, and displayed that the same universal algorithmic strategy produced clinically suitable plans. We examined compromises in selection and dimensionality reduction and showed that acceptable plans could be obtained in approximately 30 minutes, a main time savings over the physical trial and-error process of unrestricted optimization. ROCO strategies satisfy all of the clinical restrictions that were satisfied by the planner's plans; with the same PTV *D*95, there were no significant differences between the OAR sparing achieved by ROCO and the organ sparing achieved by the medical plans. From these results, we are assured that ROCO will be flexible enough for general external beam radiation remedy preparation, and is not confined to simpler treatments such as prostate cancer. A major improvement we made to ROCO in our current work is our incorporation of ROCO into MSKCC's clinical treatment scheduling system. ROCO is now capable of evaluationand inscription beam and dose information directly to/from the treatment scheduling system. Most significantly, ROCO uses the clinical full dose calculation to evaluate the dose dis tributions corresponding to each PCA method. Using an estimatedreduceddosagekernel resulted in an inaccurate dosagescheming, which proved to be a main trouble in our previous work. Ideally, ROCO would return a solution satisfying the specified hard constraints if any such feasible resolutions and anacceptable plan would consequence. In medical exercise, some iterative alteration of parameters is inevitable: the notion of clinical acceptability which varies from clinic to clinic or even planner to planner — is extremely difficult to pose either as an objective function or a solid restriction. In the future, we need to progress newrestrictions (e.g., rind-type structures to suppress hot spots in normal tissue) or objective function terms (e.g., to try and bias the solution towards more uniform PTV coverage). The key advantage of ROCO with respect to the trial-and-error loop typical of conventional soft constrained IMRT is that such constraints can be posed and a solution found within a tiny minutes. This is true because the inefficient parameter-sampling step to generate the PCA vectors is only completed after, independent of the restrictions; the controlled optimization is performed quickly in the low-dimensional space, and the new explanation, if one happens, is assured to satisfy the restrictions. This makes any trial-and-error far less tedious and the control over the solution much more direct. Improving planner time savings is one of the primary goals of our future effort with ROCO. We design to apply ROCO to skull and neckline cancer, which remains a puzzling site for current IMRT planning techniques: because of the complexity of dose-painting and the large number of OARs in behavior fields, skull and neckline plans can need days of designer time, and even then the space of medical adjustments between OAR sparing and target coverage may not have been entirely discovered. ROCO will be able to progress these restrictions by reducing the time it takes to obtain a plan that satisfies clinical constraints.

References

i. *JunzoWatada, Keisuke Aoki, Masahiro Kawano, Muhammad SuzuriHitam, Dual Scaling Approach to Data M Journal of Advanced Computational Intelligence Intelligent Informatics (JACIII), Vol. 10, No. 4, pp. 441-447, 2006.*

ii. *Jiawei Han and MichelineKamber, "Data Mining Concepts and Techniques." San Francisco, CA: Elsevier Inc, 2006.*

iii. *U. M.Piatetsky-Shapiro, G.& myth P. &Uthurusamy, R. Fayyad, "From Data Mining toKnowledge Discovery: An Overview," in Advances in Knowledge Discovery and DataMining, 1996.*

iv. *S.-C. Liao & M. Embrechts I. -N. Lee, "Data mining techniques applied to medical information," Med. Inform , pp. 81-102, 2000.*

v. *E, Donald, "Introduction to Data Mining for Medical Informatics," Clin Lab Med, pp. 9-35, 2008.*

vi. *R. Zhang, Y, Katta, "Medical Data Mining," Data Mining and Knowledge Discovery, pp. 305-308, 2002.*

vii. *Irene M. Mullins et al., "Data mining and clinical data repositories: Insights from a667,000 patient data set," Computers in Biology and Medicine, vol. 36, pp. 1351-1377, 2006.*

*viii. Deborah A. Kuban, Susan L. Tucker, Lei Dong, George Starkschall, Eugene H. Huang, M. Rex Cheung, Andrew K. Lee, and Alan Pollack, "Long-term results of the M. D. Anderson ran domized dose-escalation trial for prostate cancer," International Journal of Radiation Oncology*Biology*Physics, **70**, 67 – 74 (2008).*

*ix. Sue S. Yom, Zhongxing Liao, H. Helen Liu, Susan L. Tucker, Chao-Su Hu, XiongWei, Xuanming Wang, Shulian Wang, Radhe Mohan, James D. Cox, and Ritsuko Komaki, "Initial evaluation of treatment-related pneumonitis in advanced-stage non-small-cell lung cancer patients treated with concurrent chemotherapy and intensity-modulated radiotherapy," International Journal ofRadiation Oncology*Biology*Physics, **68**, 94 – 102 (2007).*

*x. R Lu, RRadke, L Happersett, J Yang, Chui, E Yorke, and A Jackson, "Reduced-orderconstrained optimization in IMRTplanning," Physics in Medicine and Biology, **53**, 6749–6766 (2008).*

xi. American Cancer Society, "Cancer facts & figures 2010," http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/ acspc-026238.pdf (2010).

xii. C. C. Ling et al., A Practical Guide to Intensity-Modulated Radiation Therapy (Medical Physics Publishing, 2004).