Predicting Xerostomia induced by IMRT Treatments

A logistic regression approach

Inês Soares
Department of Computer Engineering
Faculty of Sciences and Technology
Inesc-Coimbra
University of Coimbra
Coimbra, Portugal

Joana Dias
Inesc-Coimbra and Faculty of Economics
University of Coimbra
Coimbra, Portugal

Humberto Rocha
Inesc-Coimbra
Coimbra, Portugal

Brígida Ferreira
I3N
University of Aveiro
Aveiro, Portugal

Maria do Carmo Lopes
IPOC-FG, EPE, Coimbra
I3N, University of Aveiro
Aveiro, Portugal

***Abstract*— Radiotherapy is one of the main treatments used against cancer. Radiotherapy uses radiation to destroy cancerous cells trying, at the same time, to preserve healthy tissues. The planning of a radiotherapy treatment is patient dependent, resulting in a lengthy trial and error procedure until a treatment complying as most as possible with the medical prescription is found. Intensity Modulated Radiation Therapy (IMRT) is one type of radiation treatment that allows the achievement of a high degree of conformity between the area to be treated and the dose absorbed by tissues. Nevertheless, it is still not possible to eliminate completely the potential treatments' side effects. In this paper we use a database of already treated head-and-neck cancer patients in the Portuguese Institute of Oncology of Coimbra, and explore the possibility of classifying new and untreated patients according to the probability of presenting xerostomia 12 months after the beginning of IMRT treatments by using a Logistic Regression approach. The results obtained show that the classifier presents a high discriminative ability in predicting the binary response "aptness to xerostomia at 12 months".**

Keywords— Radiotherapy; IMRT; logistic regression predictors; ROC curves; AUC.

I. INTRODUCTION

Cancerous cells are characterized by having a fast reproduction and, at the same time, being less capable of repairing themselves than healthy cells if damaged by radiation. This makes radiation therapy one of the main treatments against cancer, being delivered to about 50% of all cancer patients sometime during the illness. The main goal of radiation therapy is to deliver enough radiation to kill target cells, maintaining always the compromise between the local control of the tumor and the collateral effects, i.e., minimizing the damages on the surrounding healthy organs and tissues. The treatment of each patient is personalized and planned based on computed tomography (CT) images, where the target volume(s) (PTV) and organs at risk (OAR) are delineated [1]. With the patient immobilized on a coach, in the same position he was when the CT scans were performed, the radiation is delivered by a linear accelerator (LINAC), mounted on a gantry that can rotate along a central axis. The rotation of the coach combined with the rotation of the gantry allows the irradiation from almost any angle around the tumor.

In this paper we focus on a particular type of radiation therapy: Intensity Modulated Radiation Therapy (IMRT). IMRT allows the achievement of a high degree of conformity between the delivered dose and the shape of the PTV [1,2]. In fact, the head of the LINAC is composed by a number of movable leaves (multileaf collimator - MLC), which can block part of the radiation beam during portions of the radiation delivery time, controlling the intensity and incidence of radiation. In IMRT treatments, these leaves are set to open a desired aperture, during a specific period of time (Multiple Static Collimation), so that the radiation delivered has a specific intensity profile.

To plan the treatment for a new patient, the planner should make three decisions: (1) Determine the minimum number of incidence angle to satisfy the treatment aims – Beam Angle Optimization Problem ; (2) Determine the best intensity map to the set of determined angles – Fluence Map Optimization; (3) Define the efficient way to produce the defined intensity profiles – Beam Segmentation Optimization [see, for instance, 3-6]. The treatment planning for a new patient is a complex

task, being frequently done through a trial and error time consuming procedure until a treatment complying with the medical prescription is found. This process is known as forward planning, being a time consuming practice and without guarantees of an optimal treatment. An inverse planning is preferable, consisting in the use of optimization models and algorithms to automatically determine the best treatment, given the medical prescribed doses. Nevertheless, inverse planning technique is far from being a resolved problem [see, for instance, 7].

Despite the fact that a treatment is planned trying to minimize the probability of induced complications, it is not possible to eliminate completely all side effects. When treating head-and-neck cancer patients, one of the most frequent side effects is xerostomia, the medical term for dry mouth due to lack of saliva. Xerostomia reduces drastically the quality of life of patients due to difficulties in swallowing and in feeding and they usually lose weight. It is a side effect of the exposure of salivary glands to radiation.

In this study we use an available database of head-and-neck cancer patients already treated in the Portuguese Institute of Oncology of Coimbra (IPOC-FG). Our aim is to be able to predict whether a given patient subject to a given IMRT treatment will or not experience xerostomia 12 months after the beginning of the treatments. The approach developed to address this problem consists in applying Logistic Regression, a well-known type of probabilistic statistical classification model, to predict the binary response "xerostomia problem after 12 months of IMRT treatments". The already treated patients stored in the database are used to train the predictor, which is then used to classify new patients. The Receiver Operating Characteristics (ROC) curve and the Area Under the ROC Curve (AUC) are then used to visualize the performance of the classifier and to measure the discriminative ability of the model in making the predictions for new patients. As far as the authors know, this is the first time that this methodology is applied with the aim of determining a potential problem of xerostomia after 12 months of radiation treatments.

The paper is organized as follows: In the next section we describe the database. The classification model and the performance measure that describes the discriminative ability of the model used are presented in section 3. In section 4, we describe the clinical examples of head-and-neck cases used and the computational results. The conclusions are drawn in the last section.

II. DATABASE

The existing database was provided by IPOC-FG, consisting of a total of 458 head-and-neck anonymized cancer patients, including several patients' features and medical registers. For each patient we have a set of attributes describing the patient's characteristics, another set characterizing the tumor and also information regarding the treatments that have been or are being delivered to the patient. There is also a set of attributes related to follow-up medical consultations, including information describing complications that were experienced by the patients.

The features registered by the medical oncologist for each patient before the IMRT treatment starts include:

(1) the patient's data: age and gender;

(2) description of the tumor: type, location, stage (T – the size and/or extent of the primary tumor, N – the amount of spread to nearby lymph nodes, M – the presence of metastasis or secondary tumors formed by the spread of cancer cells to other parts of the body – and AJCC – general lymphomas stage), histology;

(3) the treatments applied before or concomitantly with the radiotherapy: post-operative (namely, if the patient was submitted to surgery or not before the radiation treatment), type of quimiotherapy, type of radiotherapy;

(4) attributes related to the radiation treatment: prescribed dose, the technical plan applied on the first phase of the treatment;

(5) target volumes: identification of the PTV and its volume and the lymphatic nodes with disease (in the present database, a maximum of 4 lymphatic nodes were registered) and the corresponding volumes.

Regarding information registered after the beginning of the radiation treatment, the following attributes are considered:

(1) the starting date of the radiation treatment and the date of the last follow-up consultation;

(2) the overall and interruption treatment time (more precisely, the total number of days from the first session of radiotherapy to the last one and the total number of days that the patient did not attend to the treatment among the planned sessions);

(3) total delivered dose in Gy, total number of plans, fractions, beams, segments and monitor units on overall treatment, , mean dose on GTV – Gross Target Volume, which is the known tumor – and PTV – Planning Target Volume – which consist on the GTV and a microscopic spread (a volume known as Clinical Target Volume - CTV) plus a marginal volume around CTV) [1], D98 (the dose received by 98% of the total volume of the GTV), D2 (the dose received by 2% of the total volume of the GTV) and standard deviation of the doses delivered on GTV as well as when considering fractions of 2Gy on GTV);

(4) the response of the tumor to the treatment (complete/partial/progression) and the response type (persistence or recidive);

(5) location of local and regional disease/recidive, neoplasies, metastases (in particular on bone, brain, liver and lung);

(6) in case of death, cause of death;

(7) complications severity on the OARs in each follow-up (the complications considered in the present database were on the ears, pain, skin, larynx, pharynx, esophagus, mucous membrane, salivary glands and the

weight loss), ranked from 0 to 5, where 0 means no complication and 5 death;

(8) if there are other treatments being done concomitantly, identification of the treatment (in the present database, up to 4 alternative treatments for each patient);

(9) new follow-up institution (if the patient was transferred from IPOCFG to another medical institution);

(10) for every follow-up consultation, the number of days between the beginning of the radiation treatment and the corresponding register.

The aim of this work is to be able to predict whether patients that begin their treatment today will or will not have xerostomia 12 months from now. This means that the only information that we can use is the one that is available prior to the beginning of the radiation therapy treatment or, at most, during the first weeks of treatment. If, at an early stage of their treatment, we are able to detect patients that will probably have xerostomia later on, it will still be possible to adjust treatment plans to try to avoid this complication.

If we look at all attributes that are available before the beginning of the treatment, we can immediately observe that some of them will be irrelevant considering our objective. This is why we decided to consider only a subset of all the existing attributes. The attributes that were chosen were those with an expected strong connection with the target response according to the medical physicists. They are the following: the age of the patient, the gender, a binary variable indicating the post-operative (namely, if the patient was submitted to surgery or not before the radiation treatment), the type of quimiotherapy used, the type of radiotherapy used, the overall treatment time planned, the technical plan applied on the first phase of the treatment, the interruption treatment time, the mean dose on the GTV considering fractions of 2Gy, the severity of xerostomia problem on the first week of radiation treatments and the mean dose on all salivary glands and in each of them in particular (more precisely, the contralateral and ipsilateral parotids, the oral cavity and the contralateral and ipsilateral submandibular glands). We ended up with a total of 16 attributes.

Although the database has 458 patients, and even considering only a subset of 16 attributes, there are still many missing values. Although there are many techniques described in the literature that propose ways of dealing with missing values, taking into account the type of attributes that we are working with [8], we felt that it was better not to consider registers with missing values. This drastically diminishes the number of available patients to 16, mainly because of lack of information regarding xerostomia in the first week. This is, in fact, one of the most important attributes to take into consideration if we want to predict xerostomia in the long run. Xerostomia predominance will be highly dependent on the radiosensitivity of the patients' tissues. There is no known way of measuring this radiosensitivity, and this attribute can be a good proxy for that.

We have thus worked with a set of 16 patients with complete registers for the 16 independent variables and for the dependent variable "xerostomia problem 12 months after the beginning of IMRT treatments". Ten out of these 16 patients did not presented xerostomia after 12 months (belonging to class "0"), and 6 presented xerostomia (belonging to class "1").

On the following sections, we will use interchangeably the words patient, sample, observation, element and instance with the exact same meaning.

III. METHODOLOGICAL APPROACH

The problem of predicting a target response for a new patient based on a dataset of previously classified patients can be seen as a machine learning problem, namely, a classification learning problem. In a classification problem, a training data set consisting of n elements is available. Each element is characterized by a p-dimensional attribute vector x, belonging to a suitable space, and a class label (also known as response) $y \epsilon \{0,1,\ldots\}$. The objective is to construct a decision or classification rule that would accurately predict the class labels of elements for which only the attribute vector is observed.

We intend to apply supervised classification algorithms to classify new patients according to the possibility of having or not xerostomia 12 months after the beginning of the radiation treatment. The available database of existing patients is used as training set to define the classification model, which is then used to assign new patients to a given class, according to the target response. Our approach consists in applying a well-known technique, namely Logistic Regression. To assess the suitability of the model, we use a cross-validation procedure. The cross-validation procedure involves the partitioning of the available data sample into complementary subsets, performing the analysis on one subset (called training set) and validating the analysis on the other subset (called the validation set or testing set) [9]. We have chosen to use the leave-one-out cross-validation (LOOCV), that uses a single observation from the original sample as the validation data, and the remaining observations as the training data. So, all observations with exception of one are used to train the model. The trained model is then used to predict the class of the remaining observation. This procedure is repeated such that each element in the dataset is used once as the validation data. The ROC curve and the AUC are then used to assess the performance of the classifier and to measure the discriminative ability of the model in making the predictions for new patients. On the following, we will briefly describe the methodologies used.

A. Logistic Regression Model

Logistic regression is a renowned probabilistic statistical classification model. However, the name is somewhat misleading. Despite of, in the terminology of statistics, this model is known as logistic regression, it really is a technique for classification rather than regression [10]. The logistic regression classifier, also known as logit model, is used to predict a target response, which is a dependent variable, based on a set with one or more independent attributes. More precisely, the probabilities describing the possible values that the dependent variable could take are modeled, as a function of the explanatory variables, using a logistic function that gives outputs between 0 and 1 [10,11]. Logistic regression measures thus the relationship between a dependent variable and one or

more independent variables by using probability scores as the predicted values of the dependent variable.

Regarding the possible values of the outcome, the classifier can be of two types, binomial or multinomial. The first type deals with situations where the observed outcome can have only two possible categories; the second type considers cases where the number of available classes is higher than two. In the present work, our interest is focused on the binomial approach since the dependent variable will only take one out of two possible values: 1 if the patient presents xerostomia and 0 otherwise. Therefore, the target response falls into one of two categories, "0" or "1".

Rather than modeling the response directly, logistic regression models the probability of belonging to a particular category [11]. With this type of output, we can then apply any value as threshold to make the predictions. Thus, considering a cutoff equal to α, if the probability obtained by the logistic classifier is higher than α, the class assigned should be "1", otherwise it should be "0". In fact, the threshold value represents a decision boundary in the feature space. The most used threshold is the value 0.5.

We have used the *R* software for implementing our approach, using the *R* command *glm*. The construction of the logistic classification model is presented in algorithm 1.

Algorithm 1: Logistic Regression model

INPUTS
L: the set of observations
L_{12}: *vector(column) with the corresponding target responses*

1: $p \leftarrow$ *matrix(# observations,1)*
2: for i in L:
3: $L_{train} \leftarrow L \backslash L[i,]$
4: $L_{test} \leftarrow L[i,]$
5: *LogRegModel* $\leftarrow$ *glm(*L_{12}*~.,family="binomial"(link="logit"),data=* L_{train}*))*
6: *p[i]* $\leftarrow$ *predict(LogRegModel,* L_{test}*,type="response")*

OUTPUT
p: vector with the predicted probabilities for each observation

B. Receiver Operating Characteristic (ROC) Curve

A key question when interpreting the results of a classification model is "how well does the model discriminate between the observations with and without the outcome?". For a binary outcome, the ROC curve is the most commonly used performance measure to judge the discriminative ability of a model [12]. The logistic classifier yields a probability consisting in a numerical value that represents the degree to which an observation is a member of a class. Such score can be used as a threshold to produce a discrete (binary) classifier [13]: if the classifier output is above the cutoff, the classifier produces "1", else it produces "0". Then, and since we are working with a binary classification model, it is possible to build a specific table layout that allows visualization of the performance of the algorithm for the applied threshold, namely the confusion matrix (also known as contingency table). This structure is a table with two rows and two columns that reports the number of:

- false positives (FP): number of positive instances ("1") classified as negatives ("0");
- false negatives (FN): number of negative instances classified as positives;
- true positives (TP): number of positive instances classified as positives;
- true negatives (TN): number of negative instances classified as negatives.

Each column of the matrix represents the instances in a predicted class, while each row represents the instances in the actual class. The sum of TP and FN is the total number of patients with the outcome (P), while the sum between FP and TN is the total number of patients without the outcome (N). The accuracy of the model could be estimated as the percentage of correct predictions for a taken threshold (the most usually chosen is 0.5). However, the simple computing of the accuracy cannot be a highly reliable metric for the real performance of a classifier, because it will yield misleading results if the data set is unbalanced. In our particular case, for instance, the accuracy of predicting always "0" would be equal to 62.5%, since 10 patients out of 16 will not present xerostomia, better than any random classifier.

The ROC curve is most welcome, allowing more detailed and reliable analyses. The ROC curve is a plot of sensitivity (also known as recall and True Positive Rate (TPR)) against (1-specificity) (also known as False Positive Rate (FPR)) for consecutive cutoffs for the probability of an outcome. The sensitivity is the ratio between the TP classifications and P; the specificity is the fraction of TN classifications among N and so, the FPR is given by the ratio between FP classifications and N. The confusion matrix can be constructed for the whole range of cutoffs, from 0 to 1, and the sensitivity and specificity can also be examined over the whole range of thresholds and thus the results can be plotted in a ROC curve. Each threshold value produces a different point in the ROC curve [13]. Sorting by decreasing order the probability values produced by the classification model, an observation that is classified as positive for a given cutoff will be classified as positive for all other lower cutoffs. Thus, moving down on the sorted values and processing one observation at a time and updating the TP and FP accordingly, we can obtain the list of points that create the ROC curve. This process starts in the point (0,0) and ends at (1,1), taking a linear execution time (see Algorithm 2) [13].

C. Area Under the Curve

The ROC curve allows a clear visualization of the performance of a classifier. However, when the aim is to compare different classifiers or simply the evaluation of the performance of a single classification model, the visualization mode is not the best approach. Therefore, we have to reduce the ROC performance to a single value that represents the expected performance of the model. The most recommended method for this purpose is the AUC [12,14], which produces a value belonging to the interval [0,1]. By definition, the AUC

represents the probability that a randomly chosen positive observation is correctly ranked with a greater suspicion than a randomly chosen negative one [12-14]. Thus, in our study, it can be interpreted as the probability that a patient with the outcome is given a higher probability of the outcome by the model than a randomly chosen patient without the outcome.

A random classifier generates a ROC curve equal to the diagonal line that links the points (0,0) and (1,1), and thus produces an AUC of 0.5 [13]. Therefore, an uninformative model has an AUC lower than or equal to 0.5 and, hence, no realistic classifier will have an AUC smaller than 0.5, whereas a perfect discriminating model produces an AUC of 1 [12]. The script behind the computation of the AUC is shown in algorithm 2.

Algorithm 2: ROC and AUC (adapted from [13])

INPUTS
L: the set of test observations
p[i]: probability of observation i is positive, obtained by the classification model

1: $L_{sorted} \leftarrow$ *L sorted by decreasing order of probability values*
2: $FP \leftarrow 0$
3: $TP \leftarrow 0$
4: $R \leftarrow \{\}$
5: $FP_{prev} \leftarrow 0$
6: $TP_{prev} \leftarrow 0$
7: $A \leftarrow 0$
8: $p_{prev} \leftarrow -\infty$
9: for i in L_{sorted}*:*
10: *if* $p[i] \neq p_{prev}$*:*
11: $R \leftarrow R + (FP/N, TP/P)$
12: $base \leftarrow |FP - FP_{prev}|$
13: $height \leftarrow TP + TP_{prev}$
14: $A \leftarrow A + base \cdot height / 2$
15: $p_{prev} \leftarrow p[i]$
16: $FP_{prev} \leftarrow FP$
17: $TP_{prev} \leftarrow TP$
18: *if i is a positive observation:*
19: $TP \leftarrow TP + 1$
20: *else:*
21: $FP \leftarrow FP + 1$
22: $R \leftarrow R + (FP/N, TP/P)$
23: $base \leftarrow |1 - FP_{prev}|$
24: $height \leftarrow 1 + TP_{prev}$
25: $A \leftarrow A / (P \cdot N)$

OUTPUTS
R: the list of points that create the ROC curve
A: the AUC

IV. Results

In this section, we present the results of testing the logistic regression model to predict the complications in the salivary glands, 12 months after the beginning of IMRT treatments. Our goal concerned the ability of making correct predictions for new and unclassified patients, given a training data set containing patients already classified. Summarizing the steps followed, and thoroughly described in the previous section, we started by constructing the logistic regression model, predicting then the classes for new patients ("0" or "1") using the LOOCV technique. Once all patients were classified (notice that in the present case the test set coincides with the original data set, due to the use of the LOOCV procedure), we traced the ROC curve and determined the AUC, to evaluate the prediction ability of the model. This methodology was applied to different subsets of attributes, among the total of 16 variables described on the section 2, always considering a total of 16 patients. The best results were attained when considering the attributes: age, gender, post-operative, type of quimiotherapy, type of radiotherapy, overall treatment time, technical plan applied on the first phase of the treatment, interruption treatment time, mean dose on the GTV considering fractions of 2Gy, severity of xerostomia problem on the first week of radiation treatments and the mean dose on the contralateral and ipsilateral parotids and on the contralateral and ipsilateral submandibular glands.

The ROC curve obtained for this dataset is illustrated in Figure 1.

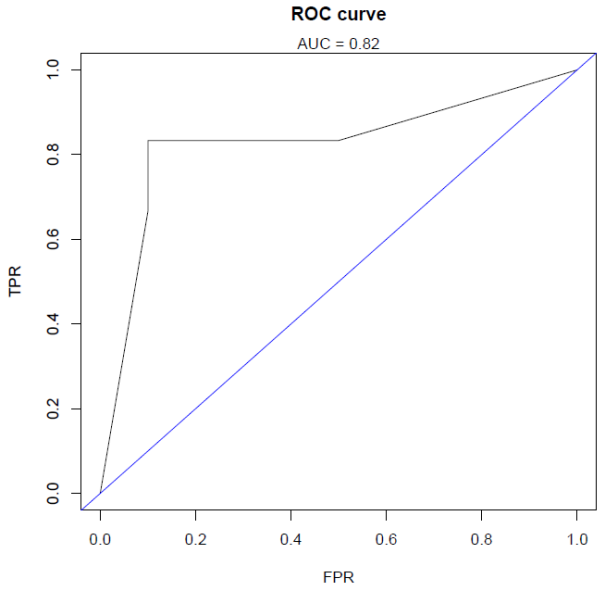

Fig. 1. ROC curve generated by Logistic Regression predictor when applied to our dataset by a LOOCV technique. The AUC produced is 0.82. The diagonal line that links the points (0,0) and (1,1) produces an AUC of 0.5.

Figure 1 shows a high performance of the logistic regression model in making the predictions of existence of xerostomia 12 months after the beginning of radiation treatments. In fact, the ROC curve traced corresponds to an AUC equal to 0.82. This value evidences that the model is capable of making predictions highly consistent with the true classifications.

Table I depicts the results produced in each iteration of algorithm 2. The logistic classifier yields a probability consisting in a numerical value that represents the degree to which a patient is a member of class "1". Such probability score can be used as a threshold to produce a classifier and, consequently, a ROC point. The column identified as "Thresholds" in table I stores these probabilities sorted by decreasing order, to be then sequentially used as threshold values in the construction of the ROC curve and also on the computation of the AUC. The TP and FP values represent the

TP and FP classifications accomplished for each threshold value. Each line of table 1 leads to the generation of a point in the ROC curve.

TABLE I. ROC CURVE PHASES.

Thresholds	TP	FP
1	0	1
1	1	1
1	2	1
1	3	1
1	4	1
0.99998	5	1
0.70626	5	2
0.10669	5	3
0.00720	5	4
1.61 x10^{-6}	5	5
2.22x10^{-16}	5	6
2.22x10^{-16}	5	7
2.22x10^{-16}	5	8
2.22x10^{-16}	5	9
2.22x10^{-16}	5	10
2.22x10^{-16}	6	10

Table I suggests that logistic regression model is able of correctly predicting the classes for new patients efficiently. Looking at the table, we are able to identify different compromises between the degree of specificity and sensibility of the classifier. Looking at the existing compromises, we can define an adequate threshold value to improve the predictions. For instance, if we consider a cutoff equal to 0.75, we are able of correctly predicting the outcome for 14 patients in a total of 16 (see the confusion matrix depicted on table II). This value produces an accuracy of 0.875. In the case of considering the most commonly used threshold, 0.5, we correctly estimate the output for 13 samples among the total of 16, obtaining an accuracy of 0.8125 (Table III). The threshold value identified as the break down in the accuracy by ROC graph produced better results than the most frequently used cutoff of 0.5. The same occurs when comparing with a random classifier, which produces an accuracy of 0.5, since the probability of a sample belongs to class "1" is 0.375 and the number of elements in this class is 6 from a total of 16 (see table IV). In fact, the random classifier is that which produces poorer results.

TABLE II. CONFUSION MATRIX FOR A THRESHOLD EQUAL TO 0.75.

		Predicted Values	
		0	**1**
True Values	**0**	9	1
	1	1	5

TABLE III. CONFUSION MATRIX FOR A THRESHOLD EQUAL TO 0.5

		Predicted Values	
		0	**1**
True Values	**0**	8	2
	1	1	5

TABLE IV. CONFUSION MATRIX FOR THE RAMDOM CLASSIFIER

		Predicted Values	
		0	**1**
True Values	**0**	6	4
	1	4	2

In summary, the logistic regression model revealed undoubtedly a high discriminative ability in the context of predicting xerostomia problem 12 months after the beginning of radiation therapy.

V. CONCLUSIONS AND FUTURE WORK

In the present article we describe a methodology capable of accurately predicting the existence of xerostomia, most well-known as dry mouth sensation, for head-and-neck cancer patients, 12 months after starting the radiation therapy treatment. The obtained results revealed a good performance of the logistic regression classifier, showing that the application of this predictive model to estimate the class for new patients will lead to robust results. The small size of the available database is the main weakness of this study. This problem will probably fade in the future, since the database is continuously being updated and the medical professionals that have to fill in the information are increasingly awaken for the importance of rigorous and systematic data registrations.

Being able to predict treatment induced complications in the long-run at early stages of radiation therapy treatments has, as major advantage, the possibility of adjusting the treatment plan such that the probability of such complications are as low as possible.

We are currently exploring this database further, trying to apply data mining algorithms not only to the short term and long term predictions of treatment induced complications but also tumor response. The obtained results can, in the future, be integrated in treatment planning optimization procedures.

ACKNOWLEDGMENT

This work was supported by FEDER, COMPETE, iCIS (CENTRO-07-ST24-FEDER-002003), and also Portuguese Foundation for Science and Technology under project grants PTDC/EIA-CCO/121450/2010, PEst-OE/EEI/UI308/2014.

REFERENCES

[1] A. Holder and B. Salter, "A tutorial on radiation oncology and optimization", H. Greenber (Eds.), Emerging Methodologies and Applications in Operations Research, Kluwer Academic Press, Boston, USA, 2004.

[2] T. Bortfeld, "IMRT: a review and preview", Physics in Medicine and Biology, vol. 51, 2006, R363-R379.

[3] H. Rocha, J. Dias, B. Ferreira, M. Lopes, "Beam angle optimization for intensity-modulated radiation therapy using a guided pattern search method", Physics in Medicine and Biology, vol. 58, 2013, pp. 2939-2953.

[4] H. Romeijn, R. Ahuja, J. Dempsey, A. Kumar, J. Li, "A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning", Physics in Medicine and Biology, vol. 48, 2003, pp. 3521-3542.

[5] H. Rocha, J. Dias, B. Ferreira, M. Lopes, "Discretization of optimal beamlet intensities in IMRT: a binary integer programming approach", Mathematical and Computer Modeling, vol. 55, 2012, pp. 1969-1980.

[6] E. Schreibmann, M. Lahanas, L. Xing, D. Baltas, "Mutiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy", Physics in Medicine and Biology, vol. 49, 2004, pp. 747-770.

[7] M. Ergott, C. Güller, H. W. Harmacher and L. Shao, “Mathematical optimization in intensity modulated radiation therapy”, A Quarterly Journal of Operations Research, vol. 6, no. 3, 2008, pp. 199-262.

[8] J. Grzymala-Busse and M. Hu, “A comparison of several approaches to missing attribute values in data mining”, in Lecture Notes in Artificial Inteligence, W. Ziarko and Y. Yao, Editors. 2001, Springer. p. 378-385.

[9] Y. Yang, “Consistency of cross validation for comparing regression procedures”, The Annals of Statistics, vol. 35, no. 6, 2007, pp. 2450-2473.

[10] C. Bishop, “Pattern recognition and machine learning”, Springer, 2006.

[11] G. James, D. Witten, T. Hastie, R. Tibshirani, “An introduction to statistical Learning with applications in R”, Springer, 2013.

[12] J. Haley and B. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve”, Radiology, vol. 143, 1982, pp. 29-36.

[13] T. Fawcett, “ROC graphs: notes and practical considerations for data mining researchers”, Technical report hpl-2003-4, HP Laboratories, Palo Alto, CA, USA, January 2003.

[14] A. Bradley, “The use of the area under the ROC curve in the evaluation of machine learning algorithms”, Pattern Recognition, vol. 30, no. 7, 1997, pp. 1145-1159