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

Maria do Carmo Lopes
IPOCFG, EPE, Coimbra
I3N, University of Aveiro
Aveiro, Portugal

Brígida Ferreira
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 minimize the damages in 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 technique of radiation treatment that allows the achievement of a high degree of conformity between the area to be treated and the dose absorbed by healthy tissues. Nevertheless, it is still not possible to eliminate completely the potential treatments' side-effects. In this retrospective study we use the clinical data from patients with head-and-neck cancer treated at the Portuguese Institute of Oncology of Coimbra and explore the possibility of classifying new and untreated patients according to the probability of 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 "at risk for xerostomia at 12 months".**

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

I. INTRODUCTION

Cancerous cells are characterized by 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 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 by the radiation oncologist [1]. With the patient immobilized on the treatment table, in the same position he/she was when the CT scan was 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]. Inside the gantry there is a multileaf collimator (MLC), which is composed by a number of movable leaves that can block part of the radiation beam and controlling the intensity of the radiation beam.

During treatment planning the goal is to irradiate homogeneously the target volume while trying to minimize the probability of inducing complications. For radiation therapy of head-and-neck cancer patients, one of the most frequent long term 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. It is a side effect of the exposure of salivary glands to radiation.

In this study the response of patients with head-and-neck cancer treated at the Portuguese Institute of Oncology of Coimbra (IPOCFG) is used. Our aim is to be able to predict

whether a given patient subject to a given IMRT treatment will or will not experience xerostomia 12 months after radiation therapy. 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 "at risk for xerostomia at 12 months". The retrospective clinical and treatment data from treated patients are used to train the predictor, which is then used to estimate if new patients will or will not have xerostomia after 12 months of radiation treatments. 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 in 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. DATASET

RESPONSE is the electronic health information system used at IPOCFG to store patient response to radiation therapy [3]. It contains the data of head-and-neck cancer patients including several patients' features and medical registers such as patient's and tumor characteristics, treatment details and patient's response to radiation therapy recorded during the follow-up medical consultations.

The aim of this work is to be able to find a model that may accurately predict whether future patients will or will not have xerostomia 12 months after the start of radiation therapy. This means that the only information that should be used in the model is the one 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.

Fifteen attributes were indicated by the medical team as probably having an important expected association with xerostomia. They are:

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

(2) treatments applied before or concomitantly with the radiotherapy: surgery (namely, if the patient was submitted to surgery or not), chemotherapy agents (if performed) and type of radiotherapy;

(3) attributes related to the radiation treatment, more precisely, the treatment technique;

(4) the overall planned treatment time (more precisely, the total number of days from the first session of radiotherapy to the last one as planned in the beginning of the treatment);

(5) the planned mean dose on the primary tumor (physical and converted for a fractionation of 2 Grays (Gy));

(6) the severity of xerostomia prior to radiation therapy;

(7) the planned mean dose in Gy (physical and corrected for a fractionation of 2 Gy) 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).

It is possible to know the 15 chosen attributes for each new arriving patient, since all these attributes are related to information that is available at the beginning of the treatment. We also need to have information available regarding the dependent variable that we can describe as "*xerostomia 12 months after the beginning of IMRT treatments*". Even considering only a subset of 15 attributes, there are some missing values in the database that limit the number of total patients that can be used. Although there are many techniques described in the literature that propose ways of dealing with missing values (see [4], for instance), taking into account the type of attributes that we are working with, we felt that it was better not to consider registers with missing values. Nevertheless, we intend to circumvent the limitation of missing values in the future by using the valuable expertise of the radiation oncologist.

We have a set of 68 patients with complete registers for the 15 independent variables and for the dependent variable. However, a total of 19 individuals interrupted the radiation therapy sometime during the treatment. For patients that interrupt the radiation therapy, delivered doses may not correspond to the planned. Thus, in order to simply the model at this point, these patients were excluded from the analysis. The existence or not of xerostomia 12 months after the start of the treatment will be related to the treatment that was delivered and not to the one that was planned. As it is not possible to know, at the start of the treatment, whether it is going to be delivered exactly as planned or not, and we can only work with attributes that are known at the beginning of the treatment, we decided to discard all patients with treatments that did not obey to the original plan. Otherwise, we could be introducing biases in the predictions of treatment plan's outcomes for new patients. In the present study, we have thus worked with a set of 49 patients. Nineteen out of these 49 patients did not present xerostomia after 12 months (belonging to class "0"), and thirty presented xerostomia (belonging to class "1").

Complications' severity is ranked from 0 to 5, where 0 means no complication and 5 death from toxicity [5]. In the present study we are only interested in predicting if the patient has or has not complications, but not the degree of the complication. Therefore, all severity degrees comprised between 1 and 2 (maximum severity obtained at this institution) were grouped and thus we only consider two severity classes, 0 and 1, where 0 means that xerostomia was

not present 12 months after the start of the treatment, and 1 otherwise.

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 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) $y \in \{0,1,\ldots\}$. The objective is to construct a decision or classification rule (also known as predictor, classifier or model) 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 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) [6]. 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 [7]. The logistic regression classifier, also known as logit model, is used to predict a new 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 [7,8] and is given by the following formula:

$$f(x) = \frac{1}{1+e^{-S}} \in [0,1], \text{ in which}$$

$$S = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots$$

being β_0, β_1, β_2, … the parameters estimated by the model based on the attributes of the training set and $x=(X_1, X_2, \ldots)$ the attribute values for the new observation. Logistic regression measures 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 estimates the probability of belonging to a particular category [8]. 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 responses

```
1: p ← matrix(# observations,1)
2: for i in L:
3:         L_train ← L\L[i,]
4:         L_test ← L[i,]
5:         LogRegModel ←
           glm(L_12~.,family="binomial"(link="logit"),data= L_train))
6:         p[i] ← 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 [9]. 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 [10]: 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 negative instances ("0") classified as positives ("1") by the model;
- false negatives (FN): number of positive instances classified as negatives by the model;
- true positives (TP): number of positive instances classified as positives by the model;
- true negatives (TN): number of negative instances classified as negatives by the model.

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 "1" would be equal to 61.2%, since 30 patients out of 49 will present xerostomia. This accuracy is better than the one obtained by 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 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 [10]. 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) [10].

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 [9,11], 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 [9-11]. 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 close to the diagonal line that links the points (0,0) and (1,1), and thus produces an AUC of approximately 0.5 [10]. 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 [9]. The script behind the computation of the AUC is shown in algorithm 2.

Algorithm 2: ROC and AUC (adapted from [10])

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 15 variables described in section 2, always considering a total of 49 patients. The best results were attained when considering the attributes: age, gender, surgery (Yes/No), type of chemotherapy, type of radiotherapy, overall planned treatment time, treatment technique, planned mean dose converted to 2Gy on the primary tumour, severity of xerostomia at baseline and the planned 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.

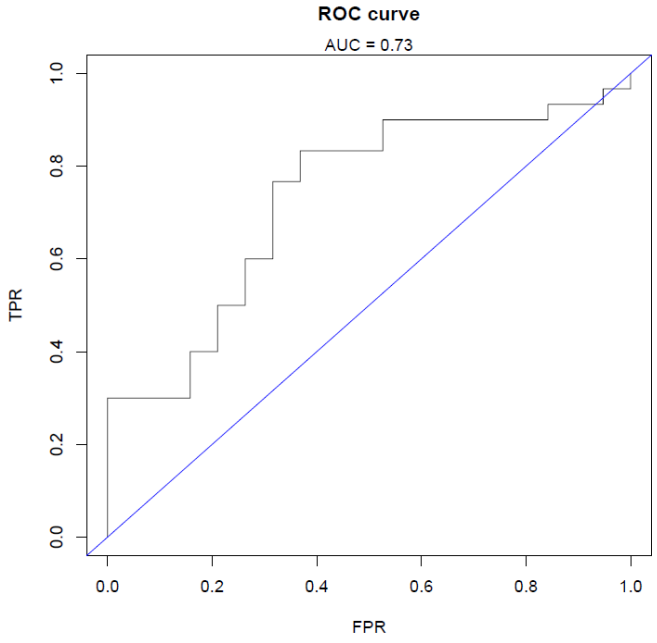

Fig. 1. ROC curve generated by logistic regression predictor when applied to our dataset by a LOOCV technique. The AUC produced is 0.73. The diagonal line produces an AUC of 0.5.

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 0.73. 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	1	0
1	2	0
1	3	0
1	4	0
1	5	0
1	6	0
1	7	0
1	8	0
0.9999	9	0
0.9982	9	1
0.9957	9	2
0.9869	9	3
0.9865	10	3
0.9801	11	3
0.9435	12	3
0.9075	12	4
0.8979	13	4
0.8883	14	4
0.8686	15	4
0.8466	15	5
0.8198	16	5
0.8180	17	5
0.7741	18	5
0.7254	18	6
0.7017	19	6
0.6311	20	6
0.5408	21	6
0.5120	22	6
0.4981	23	6
0.4891	23	7
0.4502	24	7
0.4033	25	7
0.3267	25	8
0.3025	25	9
0.2655	25	10
0.1958	26	10
0.1957	27	10
0.1725	27	11
0.1325	27	12
0.1236	27	13
0.1101	27	14
0.0576	27	15
0.0363	27	16
0.0109	28	16
0.0023	28	17
0.0020	28	18
0.0016	29	18
7.96e-04	29	19
1.41e-09	30	19

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, which depend on the threshold value. 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 0.35, we are able of correctly predicting the outcome for 37 patients in a total of 49 (see the confusion matrix depicted on Table II). This value produces an accuracy of 76%. In the case of considering the most commonly used threshold, 0.5, we correctly estimate the output for 35 samples among the total of 49, obtaining an accuracy of 71% (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 51%, since the probability of a sample belonging to class "1" is 0.61 and the number of elements in this class is 30 from a total of 49 (see table IV). In fact, the random classifier is the one which produces poorer results, as expected.

TABLE II. CONFUSION MATRIX FOR A THRESHOLD EQUAL TO 0.35.

		Predicted Values	
		0	**1**
True Values	**0**	12	7
	1	5	25

TABLE III. CONFUSION MATRIX FOR A THRESHOLD EQUAL TO 0.5

		Predicted Values	
		0	**1**
True Values	**0**	13	6
	1	8	22

TABLE IV. CONFUSION MATRIX FOR THE RAMDOM CLASSIFIER

		Predicted Values	
		0	**1**
True Values	**0**	7	12
	1	12	18

The logistic regression model revealed thus 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 xerostomia, most well-known as dry mouth sensation, for head-and-neck cancer patients, 12 months after starting the radiation therapy. The obtained results revealed a good performance of the logistic regression classifier, showing the ability of the model to estimate the class for new patients. The small size of the available sample 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. Considering the number of total patients that we are considering, it is possible that the number of variables that we are considering is too large. This is a preliminary approach, and in future work we will apply several variable screening methods in order to select the most relevant features to the goal of the study. A compromise between the number of patients and the number of attributes used is of great importance to guarantee a good performance of the model and to avoid overfitting.

Being able to predict treatment induced complications in the long-term prior to radiation therapy has, as major advantage, the possibility of adjusting the treatment plan such that the probability of such complications are minimized.

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 to 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] B.C. Ferreira, L. Khouri, M.C. Lopes and H. Ferreira, "RESPONSE, an electronic health patient information software for Radiation Therapy", IFMBE Proceedings Series, 2014 (accepted for publication).

[4] 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.

[5] J.D. Cox, J. Stetz and T.F. Pajak, "Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC)". International Journal of Radiation Oncology - Biology - Physics, vol. 31, no. 5, 1995, pp. 1341-1346.

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

[7] C. Bishop, "Pattern recognition and machine learning", Springer, 2006.

[8] G. James, D. Witten, T. Hastie and R. Tibshirani, "An introduction to statistical Learning with applications in R", Springer, 2013.

[9] 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.

[10] 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.

[11] 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