Document downloaded from:

http://hdl.handle.net/10251/105558

This paper must be cited as:

Aguado-Sarrió, E.; Prats-Montalbán, JM.; Sanz-Requena, R.; Garcia-Marti, G.; Marti-Bonmati, L.; Ferrer, A. (2017). Biomarker comparison and selection for prostate cancer detection in Dynamic Contrast Enhanced-Magnetic Resonance Imaging (DCE-MRI). Chemometrics and Intelligent Laboratory Systems. 165:38-45. doi:10.1016/j.chemolab.2017.04.003

The final publication is available at

https://doi.org/10.1016/j.chemolab.2017.04.003

Copyright Elsevier

Additional Information

Biomarker comparison and selection for prostate cancer detection in Dynamic Contrast Enhanced-Magnetic Resonance Imaging (DCE-MRI)

Aguado-Sarrió, E.[1], Prats-Montalbán J.M.[1], Sanz-Requena, R.[2], Garcia-Martí, G.[2,3], and Martí-Bonmatí, L.[2] Ferrer, A.[1]

[1]Grupo de Ingeniería Estadística Multivariante (GIEM), Universidad Politécnica de Valencia (UPV), Valencia, Spain

[2]Área de Radiología Cuantitativa y Biomarcadores de Imagen, Servicio de Radiología, Hospital Quironsalud Valencia, Valencia, Spain

[3]CIBER-SAM, Instituto de Salud Carlos III, Madrid, Spain

Abstract

In this work, the capability of imaging biomarkers obtained from multivariate curve resolution-alternating least squares (MCR-ALS), in combination with those obtained from first and second-generation pharmacokinetic models, have been studied for improving prostate cancer tumor depiction using partial least squares-discriminant analysis (PLS-DA). The main goal of this work is to improve tissue classification properties selecting the best biomarkers in terms of prediction. A wrapped double cross-validation method has been applied for the variable selection process. Using the best PLS-DA model, prostate tissues can be classified obtaining 13.4% of false negatives and 7.4% of false positives. Using MCR-ALS biomarkers yields the best models in terms of parsimony and classification performance.

1. Introduction

Angiogenesis and neovascularization are biological processes associated to tissues with increased oxygen and nutrient demand. These processes seldom occur in healthy subjects, but they are strongly present in pathological conditions such as tumors. The formation of these new and tortuous vessels produces an increase in the blood perfusion, which can be studied with dynamic contrast-enhanced Magnetic Resonance

imaging (DCE-MRI) [1]. In DCE-MRI studies, an exogenous contrast media is administered intravenously and diffuses from the capillary network into the extravascular extracellular space (EES) and return, establishing a dynamic relationship between the image signal intensity changes and the amount of contrast media that passes and diffuses into a certain tissue. The capability to analyse tumor angiogenesis in a quantitative and reproducible way from DCE-MR images has important applications to depict and gradate tumors, and also to evaluate the therapeutic response early after treatment onset [2, 3]. To obtain quantitative measurements, it is necessary to fit and characterize the intensity versus time curves associated to each voxel of the image. Out of the different approaches proposed to achieve this characterization, mathematical pharmacokinetic models have become the most popular way due to their ability to provide clinically-oriented biomarkers in tumor analysis. Nevertheless, new biomarkers obtained from multivariate curve resolution (MCR) models have recently been also proposed [4-6].

The aim of this work is to make a model comparison between different imaging perfusion related biomarkers, selecting the more relevant ones in terms of tumor prediction in order to reduce the false negatives and false positives rates (the negatives and positives corresponds to healthy and tumoral defined tissue respectively), thus improving the tissue classification. The paper is organised as follows. In Section 2 the type of images used and the pharmacokinetic and MCR models analysed are introduced. Section 3 presents the results of the statistical comparison performed with classification methods such as PLS-DA and discusses the pros and cons of using the different types of biomarkers studied. Finally, Section 4 provides the conclusions.

2. Material and Methods

The database consists of 30 histologically-confirmed cases of peripheral prostate tumors. DCE-MRI sequences are acquired in all cases, ensuring full prostate coverage (12 slices) by using an in-plane resolution of 192 × 192 voxels, each one measuring 1.5625 × 1.5625 × 4 mm^3, and 47 time points (acquisition time 5 minutes). Data are arranged in a 3D matrix (192 × 192 × 47) for each slice. Also, reference tumor and control normal region of interest (ROIs) have been manually segmented for the peripheral zone of the prostate, considering biopsy location and image findings, using a

structured reporting scheme for evaluating prostate cancer known as Prostate Imaging – Reporting and Data System (PI-RADS) version 2 [7]. This scale ranges from 1 to 5, based on increased tumor aggressiveness, and allows defining two different types of ROIs:

- DL: Dominant Lesion, related to carcinogen tissue at the peripheral prostate zone (PI-RADS $\geq$ 4).
- HP: Healthy Peripheral, related to healthy tissue at the peripheral prostate zone (PI-RADS=1).

These ROIs are manually defined by radiologists, considering the PI-RADS score and the biopsy result, and are used as the gold reference for tissue classification.

It is assumed that the behaviour of the healthy regions on the peripheral zone of the gland has no significant differences between cancer and healthy patients [8]. Therefore, it could be used safely as the healthy tissue reference.

All patients gave consent for using their medical images, which were anonymized before post-processing.

2.1. Pharmacokinetic models calculation

In radiology, pharmacokinetic models try to characterize the absorption, distribution and excretion dynamics followed by some tissue when injecting a contrast agent. In this work, compartmental models have been used to describe tissue dynamics, considering the intravascular and the extracellular extravascular (EES) spaces as main compartments. Four different models are considered, divided into two groups according to its complexity. On the one hand, “classical” or “first-generation” models, considering the Tofts model, and on the other hand “second-generation” models, 2CXM (2-compartment exchange model), AATH (adiabatic approximation to tissue homogeneity) and DP (distributed parameter) [9].

Tofts et al. [9] firstly introduced a one-compartment model as a generalization of the Kety model. The mass balance equation can be defined as follows.

$$\frac{dC_t(t)}{dt} = K^{trans} \cdot C_{AIF}(t) - k_{ep} \cdot C_t(t) \qquad \text{EES mass balance (eq.1)}$$

Where K^{trans} (min^{-1}) is the volume transfer constant; k_{ep} (min^{-1}) is the washout constant; C_{AIF} (mg/ml) is the contrast concentration of the arterial input function (AIF), which describes the contrast agent input to the tissue of interest; and C_t (mg/ml) is the contrast concentration in the EES.

This model assumes that the effect of the vascular tracer can be ignored. K^{trans} represents the total contrast transference from the plasma space to the EES, whereas k_{ep} is associated with the washout of the contrast from the EES to the plasma. In this case, the tracer transport is modelled through the EES compartment with normalized volume ($v_e=K^{trans}/k_{ep}$). This model is known as "first-generation" pharmacokinetic model and have been widely applied in oncology for perfusion analysis till the present days. However, the development and evolution of MRI hardware has provided an improvement in image quality and pixel resolution that exposes the limitations of the classical models. This allows developing new approaches designed to overcome the limitations of 'classical' models to obtain additional and more accurate information about the tissue. The most important models are the 2CXM, AATH and the DP model [9]. These new approaches are known as "second-generation" models and a scheme is shown in Figure 1.

[INSERT FIG. 1 ABOUT HERE]

The main advantage of second-generation models is the possibility of measuring the plasma flow (Fp) separately from the capillary permeability-surface area product (PS) rather than a single parameter (K^{trans}), whose physiological meaning represents a combination of Fp and PS. The separation of both phenomena allows a better understanding of perfusion behaviours in tumors.

Second-generation models are bicompartmental. In this way, the EES (v_e) and plasma space (v_p) can be used undivided or be further divided into infinitesimal subcompartments according to the model scheme (Figure 1). The notation is defined as follows: v_p (ml/ml) is the normalized volume of the plasma space; v_e (ml/ml) is the normalized volume of the EES; C_p (mg/ml) is the contrast concentration in the plasma space; C_e (mg/ml) is the contrast concentration in the EES; F_p (min^{-1}) is the plasma flow;

PS (min^{-1}) is the permeability-surface area product; E (%) is the extraction fraction; and L (mm) is the distance between the first and last subspace.

In the same way as the Tofts model, mass balances can be applied to the different compartments for tissue modelling, depending on the corresponding complexity. 2CXM is the simplest of the second-generation models, where both spaces are undivided.

$$v_p \frac{dC_p}{dt}(t) = F_p \cdot C_{AIF}(t) - F_p \cdot C_p(t) + PS \cdot C_e(t) - PS \cdot C_p(t)$$ Plasma space balance (eq.2)

$$v_e \frac{dC_e}{dt}(t) = PS \cdot C_p(t) - PS \cdot C_e(t)$$ EES balance (eq.3)

The next one is the AATH model, where the plasma space is divided in infinitesimal subcompartments and it is assumed that the permeability contrast transference is produced at the end of the vessels (last plasmatic compartment).

$$\frac{v_p}{L}\frac{\partial C_p}{\partial t}(x,t) = -F_p \frac{\partial C_p}{\partial x}(x,t)$$ Plasma subspace balance (eq. 4)

$$v_e \frac{dC_e}{dt}(t) = E \cdot F_p \cdot C_p(L,t) - E \cdot F_p \cdot C_e(t)$$ EES balance (eq. 5)

Finally, in the DP model both compartments are divided into infinitesimal spaces and the contrast transference due to permeability (PS) is assumed constant along the vessel.

$$\frac{v_p}{L}\frac{\partial C_p}{\partial t}(x,t) = -F_p \frac{\partial C_p}{\partial x}(x,t) + \frac{PS}{L}C_e(x,t) - \frac{PS}{L}C_p(x,t)$$ Plasma subspace balance (eq. 6)

$$\frac{v_e}{L}\frac{\partial C_e}{\partial t}(x,t) = \frac{PS}{L}C_p(x,t) - \frac{PS}{L}C_e(x,t)$$ EES subspace balance (eq. 7)

These systems of partial differential equations have been already solved and the contrast concentration C(t) can be expressed as a convolution product of the C_{AIF} and the R(t) function, which is the solution of the mass balances. These contrast concentrations are directly related with the DCE-MRI signal intensity. In this work, C(t) has been obtained from I(t) using a direct conversion based on relaxivity and field strength, being I(t) the signal intensity of each pixel at each time point.

$$C(t) = (C_{AIF} * R)(t)$$ (eq. 8)

In this expression, R(t) only depends on the first (K^{trans}, k_{ep}, v_e) or second (Fp, PS, ve, vp) generation perfusion biomarkers at each time instant (t). In order to obtain these biomarkers, the models need as input the reference arterial input function (C_{AIF}), which has been calculated in this work using a PCA model [10], selecting the voxels related to the arterial dynamic pattern [11]. Once the C_{AIF} is calculated for each patient individually, the perfusion sequence is analysed voxel-by-voxel applying the different pharmacokinetic approximations.

The biomarkers are calculated using non-linear optimization algorithms. For this purpose, the values of these biomarkers are evaluated in order to minimize, for each model at each voxel, the difference between the contrast measured concentration and the convolution product function: (C_{AIF} * R) (t). Note that non-linear optimization can only provide local optimums. In order to obtain reliable results, the optimization method defines different starting points and selects the best result minimizing the Residual Sum of Squares (RSS) evaluated as the sum of the squared differences at each time point. Using this technique, the probability of obtaining the global optimum is higher as the number of starting points increases, testing a relatively high number of "starting points" in the variable space.

Following the previous method, 3 biomarkers are obtained in the case of the Tofts model (K^{trans}, k_{ep}, v_e) and 4 when considering second-generation models (F_p, PS, v_e, v_p). It must be strengthened, however, that these pharmacokinetic models assume some *a priori* knowledge about the dynamics followed by the tissues. Nevertheless, it is possible that, when the tissue starts producing new vessels and tissue structures in a chaotic way, the dynamics do not behave as expected. Therefore, this work proposes to use, not only these biomarkers, but also the RSS as a complementary biomarker, which would provide information about how well the voxel is fitted by the assumed model. Figure 2 shows the biomarker images obtained from an example of the AATH model calculated for a specific slice of a tumoral affected patient.

[INSERT FIG. 2 ABOUT HERE]

2.2. MCR-ALS models calculation

One characteristic of the pharmacokinetic models is the lack of *a priori* knowledge about the tissue vascular environment, which leads to a series of assumptions conditioning the use of different approaches [12]. Accordingly, and depending on the tissue dynamics patterns, the pharmacokinetic models may provide biased measurements, which may not properly reflect the true physiology of the tissue [13]. Therefore, *a priori* knowledge about the tissue dynamics might help interpreting the information provided by the pharmacokinetic parameters.

One possible way to look for physiological meaningful dynamics is by applying multivariate statistical projection models to the DCE-MRI data. When dealing with images, the application of these models is known as Multivariate Image Analysis (MIA) [14, 15], usually based on the principal component analysis (PCA) [10]. Its application to oncology [16] allows extracting the sources of variation from a relevant number of time-sequenced images from different individuals, providing new statistical models that help explaining the perfusion differences between healthy tissues and tumors.

Nevertheless, a relevant drawback of the application of PCA in DCE-MR image analysis is that the estimated dynamics patterns are orthogonal by design. The orthogonality of the principal components is a limitation to model different perfusion behaviors that are not necessarily orthogonal. In order to overcome these drawbacks, it is possible to use more flexible models as Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) [17, 18], which do not impose this restriction. MCR is preferred to PCA because of its ability to provide physiologically more interpretable behaviors by imposing *a priori* knowledge on the model. MCR-ALS is an iterative method that performs a bilinear decomposition of an **S** matrix by means of an alternating least squares optimization algorithm,

$$\mathbf{S} = \mathbf{C}\mathbf{D}^{\mathbf{T}} + \mathbf{E} \qquad \text{(eq. 9)}$$

Where **S** contains the signal intensity registered for each voxel in rows; $\mathbf{D}^{\mathbf{T}}$ is a matrix containing in its rows each of the dynamic habits modelled (figure 4); **C** gathers in its rows the relative contribution of each modelled for each voxel of the image; and **E** is a residual matrix [4-6].

As commented in Section 2, for each slice of the studied tissue in each patient, DCE-MR images are arranged as a 3D data volume constituted by 2D images obtained at each time point t. The **S** matrix is constructed by reshaping the slices one-by-one (i.e. 47 images per slice) in a bidimensional matrix where the rows are the voxels and the columns are the time points. Thus, in this case the **S** dimensions are $((n_1 \times n_2) \text{ x } 47)$ where $n_1 \times n_2$ is the image resolution. Then, the global **S** matrix (per patient) is built stacking all the slices column-wise. A schema about the dimensions of the matrices in equation 9 is shown in figure 3.

[INSERT FIG. 3 ABOUT HERE]

By refolding the **C** matrix into the original spatial dimensions, new biomarker images are obtained (figure 5), which allow locating those voxels more related to each one of the corresponding modelled dynamic behaviors. In this case, equal length normalization is applied for $\mathbf{D}^{\mathbf{T}}$ matrix during the MCR-ALS iterative process in order to obtain concentration **C** that can be directly compared between them.

The process of obtaining the number of components and the initial estimation of $\mathbf{D}^{\mathbf{T}}$ matrix is described in previous works [4,6]. According to these works, two components are considered in MCR models regarding to the dynamic patterns related to perfusion studies that can be expected using the *a priori* knowledge about the process. (i.e. normal tissue (type NT) and vascular tissue (type VT) (figure 4). Also, a third component is obtained because of the appearance of an artefact introduced by the equipment defined as CMA [4,6]]. However, this component is not included on this work because it is a non-informative parameter for tumor classification. The NT and VT are the dynamic behaviors that we should expect when studying perfusion. They correspond to how a non-affected and vascularized tissue should behave respectively. When MCR is applied on different patients we found few variations among them, but they are essentially the same dynamic behaviors in terms of interpretation. Moreover, also in this case, it is possible to include the RSS of the MCR model as an additional potential biomarker measuring the lack of fit in each voxel location (Figure 5).

[INSERT FIG. 4 ABOUT HERE]

[INSERT FIG. 5 ABOUT HERE]

MCR-ALS is based on an iterative process that can provide infinite solutions for the same data matrix causing a problem known as ambiguity in the solution. This problem can be solved by imposing constraints commonly related to prior knowledge about the problem faced, so that it is possible to obtain easier-to-interpret solutions, which also tend to be unique when constraints introduced under the hypothesized assumptions are sensible.

For this, two additional constraints were imposed:

- Non-negativity on the pixel concentration **C** values, because the intensity in a pixel has to be nonnegative.
- Non-negativity on the dynamics profiles $\mathbf{D}^{\mathbf{T}}$.

Besides, equal length normalization is applied to $\mathbf{D}^{\mathbf{T}}$ matrix during the MCR-ALS iterative process in order to obtain concentrations **C** that can be directly compared between them.

The ambiguity in the solution problem has been checked in previous works through MCR-bands [4]. In this paper the tuned bands were very close to the proposed solution and the dynamic habits found can be considered unambiguous.

2.3. Variable selection using double cross-validation (2CV) with PLS-DA

In this work, the capability of imaging biomarkers obtained from multivariate statistic methods in combination with biomarkers from second-generation pharmacokinetic models have been studied and compared for improving prostate tissue classification using statistical classification methods based on latent structures, such as partial least squares-discriminant analysis (PLS-DA) [19, 20].

Firstly, the input matrices (**X**, **Y**) for the PLS-DA model have been built using the biomarker images obtained from the different approximations, i.e. pharmacokinetic or MCR models. The **X** matrix is constructed by stacking all the selected voxels in rows with the value of the different considered biomarkers in columns. The selected voxels are the ones defined by the radiologists as DL (lesion) or HP (healthy) ROI’s for each

patient. These ROI's are logical images associated to one specific slice of the prostate. Thus, the X matrix is constructed by stacking all the voxels assigned at the corresponding local image for all the patients. For the same voxels, a 2-column **Y** matrix is defined with two dummy variables (0-1). The first column defined the "DL" variable (value 1 if the voxel belonged to the "DL" region and value 0 if otherwise). The second column defined the "HP" variable and is built complementary to the first one. From this model, the class showing higher value is assigned to the corresponding voxel. Using the PLS-DA model predictions, the assigned category of the voxels has been compared with the original classification. Therefore, if the voxel belongs to HP, it can be evaluated as a TN (True Negative) or FP (False Positive) depending on the prediction, and as a TP (True Positive) or FN (False Negative) if it is DL. Then, two different parameters, precision and recall, are calculated from the percentage of TP, FN and FP:

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}} \quad \text{recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad \text{(eq .10)}$$

Both indexes are combined in a new performance index, f-score, chosen to evaluate the classification model performance. The f-score [19] is defined as the weighted harmonic mean of these two parameters:

$$\text{f-score} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision+recall}} \qquad \text{(eq. 11)}$$

This parameter determines the goodness of prediction for a classification model. It ranges between 0 and 1, and takes the maximum when the precision and recall are one (the number of FP and FN are zero). The closer the f-score is to one, the better the model is in terms of prediction.

The variable selection method proposed in this work (Figure 6) is a wrapped double cross-validation (2CV) with variable selection, showing high similarities with other 2CV methods [21, 22]. This methodology is applied in order to determine which parameters supply relevant information for classification. In this work, the variable selection method allows us to remove all non-informative variables improving the classification performance (misclassifications rate based on the f-score).

The method consists of dividing the voxels from the cases of the data set (30 patients in this work) in three randomized groups of cases (i.e. 10 patients), defined as training, validation and test. All the voxels from each case have their own class identification (DL or HP) and are located in their corresponding group. This way, the voxels from a specific case are always included in the same group in order to avoid any type of overfitting. Moreover, the number of voxels in each category has been balanced as much as possible in order to avoid any bias of the model.

The method performs as follows: starting from a number of latent variables (NLV) equal to one, the training set is used for PLS-DA model building, using all the variables (biomarkers) of the considered pharmacokinetic or MCR model. Then, projecting the validation set onto the model fitted with the training set, an initial f-score (0) is calculated evaluating the performance in the model classification.

The f-score (0) is stored and then, the values of the $\mathbf{B}_{\mathbf{PLS}}$ coefficients for each variable or biomarker are compared with their "null" distribution obtained after breaking the correlation structure between **X** and **Y** of the training set. This breakage process consists of randomizing the order of the **Y** matrix rows keeping the same **X** and building a PLS-DA model to obtain the "null" model coefficients. This is internally repeated 500 times in order to obtain the null distribution [21, 22] of these coefficients. This way, the variable (biomarker) is removed from the **X** matrix if the coefficient of a certain variable is not statistically significant. It is considered statistically significant if the real coefficient is out of the central 95% range for the random null distribution values (i.e. α=0.05).

Once all the non-statistically significant variables are removed, a new PLS-DA model is built with the remaining significant variables from the training set **X** matrix, obtaining a new value of f-score (1) after projecting again the same validation set (only using the remaining significant variables) onto the new PLS-DA model. If the new f-score (1) is higher than f-score (0), the model is improved and the new value of f-score (1) is updated. In this case, the iterative process continues with a new variable selection comparing the new $\mathbf{B}_{\mathbf{PLS}}$ coefficients after breaking again the correlation structure between **X** and **Y**. However, if f-score (1) is lower than f-score (0), the best model is the

one considered in the previous step. This iterative process is repeated until the f-score (n) is lower than the one obtained in the previous step, f-score (n-1); keeping this "best" model with its associated f-score (n-1) and its own variable selection. From the best model selected, a final external set (test) is projected onto this model obtaining the "final f-score". This value is stored for further comparisons.

Afterwards, the NLV is increased in one repeating the same process explained for one NLV. This way, at the end we obtain *N*-1 (where *N* is the initial number of variables considered in the **X** matrix) improved models (with its own variable selection) and with their associated value of NLV and final f-score. After this, the final f-score of these *N*-1 models are compared and the highest final f-score determines the best model with the best variable selection. Note that if different models provide the same value of final f-score, the most parsimonious model is preferred for simplicity.

Once this process is completed, the initial groups (training, validation and set) are permuted (P) in the three possible different combinations (i.e. training to validation, validation to test and test to training) ensuring that every case belonged, at least, one time to each group. Then the iterative process is repeated again with the new groups, obtaining, at the end, three different improved models. Finally, the initial group randomization is repeated 500 times obtaining (500 x 3 = 1500) different data organizations. This scheme is shown in Figure 6.

[INSERT FIG. 6 ABOUT HERE]

At the end of the process, the value of the final f-score, the variable selection (biomarkers that remain in the model), the percentage of TN, TP, FN, FP and the related NLV for this "best" model obtained for each distribution of the groups (1500) are stored.

Once the process is finished, the variable relative inclusion rate (percentage of times a variable is included from a pharmacokinetic, MCR or combined model) is evaluated as an additional indicator.

In order to show summarized and consistent results, the analysis is simplified selecting the highest 5% final f-score models. Also, as an additional constraint for this "5%

selection”, the percentage of FN is limited to 20%, considering a higher percentage excessive for tumor detection.

All calculations are performed in Matlab R2014b (The Mathworks Inc., Natick, MA, USA).

3. Results and Discussion

The pharmacokinetic and MCR models have been evaluated, individually and in combination, in order to improve the results with different types of biomarkers. This way, the number of columns of the **X** matrix may vary depending on the number of biomarkers of the selected models. Table 1 shows the results:

[INSERT TABLE 1 ABOUT HERE]

Three one-way ANOVAs have been proposed to study the statistical effect of the model on f-score, %FP and %FN by means of the Least Significant Difference (LSD) Intervals (Figures 7, 8 and 9), respectively.

[INSERT FIG. 7 ABOUT HERE]

[INSERT FIG. 8 ABOUT HERE]

[INSERT FIG. 9 ABOUT HERE]

As can be seen in the figures, statistical significant differences are observed between the proposed models for the mean f-score and %FP (Figures 7 and 8). However, no statistically significant differences appear between FN rates, as can be seen in Figure 9. Therefore, the ANOVA analyses show that the differences between the f-score are due to the differences in the %FP. Regarding the f-score and %FP, Toft’s model (mean f-score of 0.73 and mean %FP of 20.7) performs significantly worse than two of the second-generation models (AATH and 2CXM) because their LSD intervals do not overlap obtaining, a mean f-score of 0.755 and a mean %FP of 15.4 . Furthermore, second-generation 2CXM or AATH models are preferred over DP models due to the excessive computational time required to obtain the DP biomarkers, much higher than

the other two second-generation models (6 times longer per voxel). Nevertheless, MCR alone or in combination with the best pharmacokinetic models obtained the best results for prediction, significantly better than any other types of model. This way, the mean f-score raised up to 0.7857 with 0.8145 as its maximum value, heavily reducing the rate of false positives (from 15.4% to 7.4%), which corresponds with 0.74 of sensitivity and 0.85 of specificity. These results are better than other studies including only first generation pharmacokinetic models [8, 23-26]. Additionally, no statistically significant differences were observed between the MCR and combined models. This suggests that using just MCR can be considered the best option for improving the f-score, since it is the most parsimonious model (i.e. adding the second-generation biomarkers do not improve statistically the classification results).

Regarding the variable inclusion rate in the combined models, the MCR-ALS and the pharmacokinetic RSS variables have been included in more than 90% of the times, a much higher percentage in comparison with the 70% inclusion rate of the pharmacokinetic biomarkers. This result suggests that voxels that do not fit the models properly (high RSS) should be considered as potential locations for abnormal vascular behaviors. They could be related to the presence of a tumor, and could be used as a surrogate indicator in order to locate lesions (i.e. a new biomarker). This finding needs further validation. Additionaly, the RSS parameter can be used as an indicator of how reliable the provided pharmacokinetic biomarkers are. This way, voxels that are not well fitted by the model could be re-estimated by using imputation methods [27].

This study has some limitations. The classification index (f-score) parameter was selected because it is a balanced combination of the parameters that need to be minimized for improving the classification goodness (rates of FN and FP), but other indexes can be used instead (e.g. AUROC, area under the receiver operating characteristic curve, a common quality index in medicine). Both parameters are used in the field [28-30] and provide equivalent information. However, f-score is preferred in this work due to its simplicity and independence from the number of TN (the comparison of both figures-of-merit is out of the scope of this work). Also, the method proposed in this work is a combination of 2CV with variable selection, and could be tested by comparing the results with other variable selection methods (e.g. VIP's, LASSO or selectivity ratio) [31-33]. Therefore, this work can be complemented making

a comparison between different methods and indexes for selecting the one that provides the best results.

4. Conclusions

Out of the pharmacokinetic models proposed, AATH has showed up as the best one, whereas the DP model is discarded due to the high computational time, and 2CXM is less consistent in comparison with AATH. These second-generation models have performed better for tumor detection than classical Tofts models, as demonstrated by the higher values of f-score. Also, RSS has risen as a potential biomarker in terms of inclusion rate.

Nevertheless, the use of imaging biomarkers from MCR-ALS methods has provided better results than the DCE-MRI pharmacokinetic approximations for voxelwise classification. The MCR dynamic behaviours (NT and VT) and the RSS were most of the times statistically significant, being the best biomarkers for lesion detection and tissue.

5. Acknowledgements

This research has been supported by "Generalitat Valenciana (Conselleria d'Educació, Investigació, Cultura I Esport)" under the project AICO/2016/061.

6. References

[1] Collins DJ, Padhani AR. Dynamic magnetic resonance imaging of tumor perfusion. *IEEE Eng. Med. Biol. Mag*. 2004; 23: 65-83.

[2] Jackson ASN, Reinsberg SA, Sohaib SA, Charles-Edwards EM, Jhavar S, Christmas TJ, Thompson AC, Bailey MJ, Corbishley CM, Fisher C, Leach MO, Dearnaley DP. Dynamic contrast-enhanced MRI for prostate cancer localization. *Br. J. Radiol*. 2009; 82: 148–156.

[3] Leach MO, Brindle KM, Evelhoch JL, Grif ths JR, Horsman MR, Jackson A, Jayson GC, Judson IR, Knopp MV, Maxwell RJ, McIntyre D, Padhani AR, Price P, Rathbone R, Rustin GJ, Tofts PS, Tozer GM, Vennart W, Waterton JC, Williams SR, Workman P. The assessment of antiangiogenic and antivascular therapies in early-stage clinical trials using magnetic resonance imaging: issues and recommendations. *Br. J. Cancer* 2005; 92: 1599–1610.

[4] Prats-Montalbán JM, Sanz-Requena R, Martí-Bonmatí L, Ferrer A. Prostate functional magnetic resonance image analysis using multivariate curve resolution methods. *J. Chemometrics*. 2014: 28:672-680.

[5] Aguado-Sarrió, E, Prats-Montalbán JM, Sanz-Requena R, Alberich-Bayarri A, Martí-Bonmatí L, Ferrer A. Prostate Diffusion Weighted-Magnetic Resonance Image analysis using Multivariate Curve Resolution Methods. *Chemometrics and Intelligent Laboratory Systems*, 140, 43-48, 2015.

[6] Prats-Montalbán JM, Aguado-Sarrió E, Ferrer A. (2016) Multivariate Curve Resolution for Magnetic Resonance Image analysis: applications in prostate cancer biomarkers development. In *"Resolving Spectral Mixtures, with application from ultrafast spectroscopy to super-resolution imaging"*, Data Handling in Science and Technology 30, 519–550, Elsevier, Amsterdam.

[7] Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, Margolis D, Schnall MD, Shtern F, Tempany CM, Thoeny HC, Verma S. PI-RADS prostate imaging – reporting and data system: 2015, version 2. *Eur Urol* 2016: 69:16-40.

[8] Sanz-Requena R, Martí-Bonmatí L, Pérez-Martínez R, García-Martí G. Dynamic contrast-enhanced case-control analysis in 3T MRI of prostate cancer can help to characterize tumor aggressiveness. *Eur J Radiol* 2016:85:2119-2126.

[9] Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, Larsson HB, Lee TY, Mayr NA, Parker GJ, Port RE, Taylor J, Weisskoff RM. Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J. Magn. Reson. Imaging* 1999; 10: 223–232.

[10] Jackson JE. A User’s guide to Principal Components. Ed. Wiley: New York, 1991.

[11] Sanz-Requena R, Prats-Montalbán JM, Martí-Bonmatí L, Alberich-Bayarri A, García-Martí G, Pérez R, Ferrer A Automatic individual arterial input functions calculated from PCA outperform manual and population-averaged approaches for the pharmacokinetic modeling of DCE-MR images. *J. Magn. Reson. Imaging*. 2015: 42: 477-487.

[12] Sourbron SP, Buckley DL. Tracer Kinetic modelling in MRI: estimating perfusion and capillary permeability. *Phys. Med. Biol*. 2012; 57:R1-R33.

[13] Lüdemann L, Prochnow D, Rohl ng T, Franiel T, Warmuth C, Taupitz M, Rehbein H, Beyersdorff D. Simultaneous quantification of perfusion and permeability in the prostate using dynamic contrast-enhanced magnetic resonance imaging with an inversion-prepared dual-constrast sequence. *Ann. Biomed. Eng*. 2009; 37: 749–762.

[14] Geladi P, Grahn H. Multivariate Image Analysis. Wiley: Chichester, England, 1996.

[15] Prats-Montalbán JM, Ferrer A, de Juan A. Multivariate image analysis: a review with applications. *Chemom. Intell. Lab. Syst*. 2011; 107: 1–23.

[16] Bruwer MJ, MacGregor JF, Noseworthy D. Dynamic contrast-enhanced MRI diagnosis in oncology via principal component analysis. *J. Chemometrics.* 2008; 22: 708–716.

[17] Tauler R, Smilde AK, Kowalski BR. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemometrics*. 1995; 9: 31–58.

[18] Tauler R. Multivariate curve resolution applied to second order data. Chemom. Intell. Lab. Syst 1995; 30: 133–146.

[19] Sjöström M, Wold S, Söderström B. (1986) PLS Discriminant Plots, Proceedings of PARC in Practice, Amsterdam, June 19-21, 1985. Elsevier Science Publishers B.V., North-Holland.

[20] Geladi P, Kowalski BR. Partial Least-Squares Regression: A Tutorial, *Analytica Chimica Acta.* 1986:185: 1-17.

[21] Westerhuis J, Hoefsloot H, Smit S, Vis D, Smilde A, van Velzen E, van Duijnhoven J, van Dorsten F. Assesment of PLSDA cross validation. *Metabolomics*. 2008: 4:81-89.

[22] Kuligowski J, Pérez-Guaita D, Escobar J, de la Guardia M, Vento M, Ferrer A, Quintás G. Evaluation of the effect of chance correlations on variable selection using partial least squares-discriminant analysis. *Talanta*. 2013: 116:835-840.

[23] Jackson AS, Reinsberg SA, Sohaib SA, Charles-Edwards EM, Jhavar S, Christmas TJ, Thompson AC, Bailey MJ, Corbishley CM, Fisher C, Leach MO, Dearnaley DP. Dynamic contrast-enhanced MRI for prostate cancer localization. *Br J Radiol* 2009;82:148-156.

[24] Franiel T, Lüdemann L, Rudolph B, Lutterbeck E, Hamm B, Beyersdorff D. Differentiation of prostate cancer from normal prostate tissue: role of hotspots in pharmacokinetic MRI and histologic evaluation. *AJR Am J Roentgenol* 2010;194:675-681.

[25] Vos EK, Litjens GJ, Kobus T, Hambrock T, Hulsbergen-van de Kaa CA, Barentsz JO, Huisman HJ, Scheenen TW. Assessment of prostate cancer aggressiveness using dynamic contrast-enhanced MRI at 3T. *Eur Urol* 2013;64:448-455.

[26] Durmus T, Vollnberg B, Schwenke C, Kilic E, Huppertz A, Taupitz M, Franiel T. Dynamic contrast-enhanced MRI of the prostate: comparison of gadobutrol and Gd-DTPA. *Fortschr Röntgenstr* 2013;185:862-868.

[27] Folch-Fortuny A, Arteaga F, Ferrer A. (2015) PCA model building with missing data: New proposals and a comparative study. *Chemometrics and Intelligent Laboratory Systems* 146, pp. 77-88.

[28] Yetisgen-Yildiz M, Gunn M.L, Xia F, Payne T.H. A text processing pipeline to extract recommendations from radiology reports. *J. Biomed. Inform*. 2013: 46:354-362.

[29] Pham A.D, Névéol A, Lavergne T, Yasunaga D, Clément O, Meyer G, Morello R, Burgun A. Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. *BMC Bioinformatics*. 2014: 15:266.

[30] Obuchowski NA. Receiver Operating Characteristic Curves and their use in radiology. *Radiology*. 2003: 229:3-8.

[31] Favilla S, Durante C, Li Vigni M, Cocchi M. Assessing feature relevance in NPLS models by VIP. *Chemometrics and Intelligent Laboratory Systems*. 2013: 129:76-86.

[32] Tibshirani R. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B*. 1996: 58:267-288.

[33] Hastie T, Tibshirani R, Friedman J. (2009) The Elements of Statistical Learning. New York: Springer.

Table citations

Model	Max f-score	Mean f-score	Mean %FP	Mean %FN	Best Biomarkers (Inclusion rate)
TOFTS	0.7871	0.7302	20.7	14.6	RSS, v_e (80%) Others (<60%)
AATH	0.8125	0.7530	16.8	13.2	RSS (95%) Others (80%)
2CXM	0.8187	0.7556	15.4	13.6	Fp RSS (90%) Others (50-60%)
DP	0.8318	0.7455	16.1	14.7	RSS (100%) Others (<60%)
MCR	0.8145	0.7857	7.5	13.5	Type VT (95%) Type NT (100%)
MCR + 2CXM	0.8162	0.7789	9.2	13.9	MCR (>90%) 2CXM (50-70%)
MCR + AATH	0.8388	0.7804	9.6	12.6	MCR (>90%) AATH (60-70%)
MCR + DP	0.8659	0.7725	10.1	13.8	MCR (>90%) DP (60-70%) RSS DP (90%)

Table 1. Results summary for each considered model or models combinations using the 5% best models according to the final f-score. The best biomarkers were selected in terms of inclusion rate.

Figure Captions

Figure 1. Second-generation models scheme, (left to right: two-compartment exchange model, 2CXM; adiabatic approximation of tissue-homogeneity, AATH; distributed-parameter, DP) sorted in terms of complexity attending to the number of compartmental divisions.

Figure 2. AATH model biomarker images representing the value of FP, PS, v_e and v_p, calculated voxel-by-voxel. The residual sum of squares (RSS) biomarker image is included as a new biomarker type, showing the lack of fit in each voxel.

Figure 3. MCR model schema containing the dimensions of the matrices for the DCE-MRI perfusion analysis. The 3 components considered in this model are the ones related to NT (normal tissue), VT (vascularized tissue) and CMA (contrast media arrival).

Figure 4. Dynamic behaviors obtained from MCR-ALS model, representing VT (vascularized tissue) and NT (normal tissue).

Figure 5. MCR-ALS biomarker images representing the relative contribution of each dynamic behavior (vascularized tissue VT and normal tissue NT) to a specific voxel. The residual sum of squares (RSS) distribution map shows the voxels that are not well fitted by the model.

Figure 6. Variable selection process scheme, repeated 500 times to obtain the results for each proposed model. P represents the number of the three different group permutations proposed and NLV is the number of latent variables for PLS-DA.

Figure 7. LSD ANOVA intervals for the f-score mean of the different individual and combined models (α=0.05) separated by the black dotted line.

Figure 8. LSD ANOVA intervals for the %FP mean of the different individual and combined models (α=0.05) separated by the black dotted line.

Figure 9. LSD ANOVA intervals for the %FN mean of the different individual and combined models (α=0.05) separated by the black dotted line.