Improved Parkinsonism diagnosis using a partial least squares based approach

F. Segovia, a) J. M. Górriz, J. Ramírez, and I. Álvarez
Department of Signal Theory, Networking and Communications, University of Granada, Granada 18071, Spain

J. M. Jiménez-Hoyuela and S. J. Ortega
Department of Nuclear Medicine, Hospital Universitario Virgen de la Victoria, Málaga 29010, Spain

(Received 23 January 2012; revised 16 May 2012; accepted for publication 6 June 2012; published 27 June 2012)

Purpose: An accurate and early diagnosis of Parkinsonian syndrome (PS) is nowadays a challenge. This syndrome includes several pathologies with similar symptoms (Parkinson’s disease, multisystem atrophy, progressive supranuclear palsy, corticobasal degeneration and others) which make the diagnosis more difficult. 123I-ioflupane allows to obtain in vivo images of the brain that can be used to assist the PS diagnosis and provides a way to improve its accuracy.

Methods: In this paper, we show a novel method to automatically classify 123I-ioflupane images into two groups: controls or PS. The proposed methodology analyzes separately each hemisphere of the brain by means of a novel approach based on partial least squares (PLS) and support vector machine.

Results: A database with 189 123I-ioflupane images (94 controls and 95 pathological images) was used for evaluation purposes. The application of the proposed method based on PLS yields high accuracy rates up to 94.7% with sensitivity = 93.7% and specificity = 95.7%, outperforming previous approaches based on singular value decomposition, which are used as a reference.

Conclusions: The use of advanced techniques based on classical signal analysis and their application to each hemisphere of the brain separately improves the (assisted) diagnosis of PS. © 2012 American Association of Physicists in Medicine. [http://dx.doi.org/10.1118/1.4730289]

Key words: automatic classification, 123I-ioflupane, partial least squares, support vector machines, Parkinson’s disease

I. INTRODUCTION

Parkinsonian syndrome (PS), a.k.a. Parkinsonism, is a neurological disorder characterized by the presence of hypokinesia associated with rest tremor and/or rigidity and/or postural instability. The most common cause of Parkinsonism is the Parkinson’s disease (PD), a neurodegenerative disease that originates due to the progressive loss of dopaminergic neurons of the nigrostriatal pathway (a neural pathway that connects the substantia nigra with the striatum). As a result of this cell loss, there is a substantial decrease in the dopamine content of the striatum, and a corresponding loss of dopamine transporters. Nowadays, PD is the second most frequent neurodegenerative disease. About 1%–2% of people over 65 years suffer the disease and about 20%–24% of them are incorrectly diagnosed.

123I-ioflupane (better known by its tradename DaTSCAN) is a neuroimaging radiopharmaceutical drug, widely used to assist the diagnosis of Parkinsonism. After intravenous injection, this drug binds to the dopamine transporters in the striatum. Then, using a single-photon emission computed tomography (SPECT) camera, it is possible to obtain the distribution of the radiopharmaceutical in the brain and visualize the loss of dopamine transporters. The SPECT maps obtained that way are traditionally examined by experienced clinicians to assist the PS diagnosis. For this purpose, clinicians often use proprietary software to delimit regions of interest (ROIs) and quantify the radiopharmaceutical uptake. This procedure is subjective and prone to error since it requires gross changes in the transporter density throughout the ROI to allow the differentiation between controls and pathological images. As such, it may not be sensitive to changes in the pattern of distribution that can characterize the progression of the disease.

Machine learning is a scientific discipline concerned with the design and development of algorithms that allows computers to recognize complex patterns and make intelligent decisions based on data. In combination with image analysis procedures, it allows to develop computer-aided diagnosis (CAD) systems for several neurodegenerative diseases, such as PS. These systems not only process and analyze image data but also can determine if an image belongs to the class of normal images (healthy subjects) or pathological images (patients), performing that way an automatic diagnosis.

Recently, a method that combines principal components analysis (PCA) and several classifiers in order to automatically diagnose PS has been proposed. The authors use the single value decomposition (SVD) to reduce the input space and improve the classifiers performance. This approach yields high accuracy rates and outperforms other widely used commercial systems, such as Brass and QuantiSPECT.

In this work, we propose an improved approach that uses partial least squares (PLS) instead of SVD. PLS and SVD are similar techniques since both provide a data decomposition using the variance, however, the former takes into account the information about the underlying classes (labels) during the decomposition. In addition, we perform the classification by means of a support vector machine (SVM)
classifier\(^8\) rather than the Bayes’ rule. Another major difference is that we analyze and classify the two hemispheres of the brain separately and then an image is classified as PS whether one of its hemispheres is classified as such. That way two goals are achieved: on the one hand, we improve the detection of the disease even when it only affects a given hemisphere (patients often develop symptoms on one side of the brain before developing bilateral disorder\(^3\)) and, on the other hand, we reduce by half the input space for classification operating on the ROIs which are relevant in medical practice.

This paper is organized as follows. Section II describes the background of the key techniques used in the paper. Section III discusses the materials and methods employed in this work, i.e., the database of 123I-ioflupane images used for evaluation purposes and the proposed methodology based on PLS and SVM. The application of the latter methods to DaTSCAN images is shown in Sec. IV, performing a fair comparison with the results obtained in Ref. 7 which are used as a baseline. Results are discussed in Sec. V and finally, conclusions are shown in Sec. VI.

II. BACKGROUND

II.A. Partial least squares

PLS (Ref. 13) is a statistical method for modeling relations between sets of observed variables by means of latent variables. The underlying assumption of all PLS methods is that the observed data are generated by a system or process which is driven by a small number of latent (not directly observed or measured) variables.

Mathematically, PLS is a linear algorithm for modeling the relation between two data sets \(X \subset \mathbb{R}^N\) and \(Y \subset \mathbb{R}^M\). After observing \(n\) data samples from each block of variables, PLS decomposes the \(n \times N\) matrix of zero-mean variables, \(X\), and the \(n \times M\) matrix of zero-mean variables, \(Y\), into the form

\[
X = TP^T + F, \quad Y = UQ^T + F, \quad (1)
\]

where \(N\) and \(M\) are, respectively, the number of features and the number of properties of the observed variables. \(T\) and \(U\) are \(n \times p\) matrices of the \(p\) extracted score vectors (also known as components or latent vectors), the \(N \times p\) matrix \(P\), and the \(M \times p\) matrix \(Q\) are the matrices of loadings and the \(n \times N\) matrix \(E\) and the \(n \times M\) matrix \(F\) are the matrices of residuals (or error matrices). The \(x\)-scores in \(T\) are linear combinations of the \(x\)-variables and can be considered as good summaries of the \(x\)-variables. Similarly, the \(y\)-scores in \(U\) are linear combinations of the \(y\)-variables and can be considered as good summaries of them.

The model structures of PLS and PCA are essentially the same. Both methods transform the data into a set of a few intermediate linear latent variables (components) and then, these new variables are used for regression, dimension reduction, etc. The main difference between PLS and PCA is that the former creates orthogonal weight vectors by maximizing the covariance between elements in \(X\) and \(Y\). Thus, when PLS is used for dimension reduction, it not only considers the variance of the samples (as PCA) but also considers the class labels. Fisher discriminant analysis (FDA) is, in this way, similar to PLS. However, FDA has the limitation that after dimensionality reduction, there are only \(c - 1\) meaningful latent variables, where \(c\) is the number of classes being considered. Additionally, when the number of features exceeds the number of samples, the covariance estimates when applying PCA do not have full rank and the weight vectors cannot be extracted.

II.B. Support vector machine

Support vector machine\(^8\) separates a set of binary labeled training data by means of a hyperplane (called maximal margin hyperplane) that is maximally distant from the two classes. The objective is to build a function \(f : \mathbb{R}^N \rightarrow \pm 1\) using training data, that is, \(N\)-dimensional patterns \(x_i\) and class labels \(y_i\) so that \(f\) will correctly classify new examples \((x, y)\) \((x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i) \in \mathbb{R}^N \times \pm 1\).

(2)

Linear discriminant functions define decision hyperplanes in a multidimensional space

\[
g(x) = w^T x + w_0 = 0,
\]

where \(w\) is the weight vector that is orthogonal to the decision hyperplane and \(w_0\) is the threshold. The optimization task consists of finding the unknown parameters \(w\) and \(w_0\) that define the decision hyperplane. When no linear separation of the training data is possible, SVM can work effectively in combination with kernel techniques so that the hyperplane defining the SVM corresponds to a nonlinear decision boundary in the input space. A kernel function is defined as

\[
K(x_i, x_j) = \varphi(x_i) \varphi(x_j).
\]

(4)

The use of kernel functions avoids directly working in the high dimensional feature space, thus the training algorithm only depends on the data through dot products in Euclidean space, i.e., on terms of the form \(\varphi(x_i)\varphi(x_j)\).

III. MATERIALS AND METHODS

III.A. Image database

A database consisting of 189 SPECT images from 189 subjects (94 controls and 95 PS patients) was used in order to evaluate the proposed methodology. The images were acquired by the “Virgen de la Victoria” hospital (Málaga, Spain) from January 2003 until December 2008 (see Table I for demographic details). All of them are from patients that

<table>
<thead>
<tr>
<th></th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>Controls</td>
<td>94</td>
<td>49 45</td>
</tr>
<tr>
<td>Patients</td>
<td>95</td>
<td>54 41</td>
</tr>
<tr>
<td>(\mu)</td>
<td>69.26</td>
<td>68.29</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>10.16</td>
<td>9.62</td>
</tr>
<tr>
<td>Range</td>
<td>33–89</td>
<td>30–87</td>
</tr>
</tbody>
</table>

Table I. Demographic details of the SPECT images used in this work. \(\mu\) and \(\sigma\) stand for the average and the standard deviation, respectively.
attended to the hospital redirected by primary care services and who had symptoms that could be considered as Parkinsonism symptoms. Patients on treatment with drugs which have an effect, known or suspected, by a direct competitive mechanism at the level of dopaminergic transporters were excluded. The SPECT data acquisition was performed after a period of between 3 and 4 h after the intravenous injection of 185 MBq (5 mCi) of $^{123}$I-ioflupane, with prior thyroid blocking by means of the Lugol’s solution. The gamma camera used was a Millennium model from General Electric, equipped with a dual head and general purpose collimator. A 360° circular orbit was made around the cranium, at 3° intervals, 60 images per detector with a duration of 35 s per interval, 128 × 128 matrix. Image reconstruction was carried out using filtered backprojection algorithms without attenuation correction, application of a Hanning filter (frequency 0.7) and images were obtained with transaxial cuts.

All the images were spatially normalized using the SPM software and yielding a $73 \times 73 \times 45$ three-dimensional functional activity map for each patient. This software performs the normalization in two steps. The first one assumes a general affine model with 12 parameters and a cost function, $f$, that measures the difference between a given image and a template

$$f = \sum_{k} (i(Mx_k) - t(x_k))^2,$$

where $i$ is the initial image, $Mx_k$ is the affine transformation for each voxel $x_k = \{x_1, x_2, x_3\}$, and $t$ is the template. The second step consists of several nonlinear deformations defined by a linear combination of three-dimensional discrete cosine transform (DCT) basis functions. This spatial normalization ensures that the same position in the volume coordinate system in different images corresponds to the same anatomical position. Moreover, the intensity of the images was normalized by equaling the background (average intensity of voxels which intensity is between 15% and 30% of maximum intensity) of each image and the template, and scaling the remaining voxels. The template used for spatial and intensity normalization was computed as follows:

- The $n - 1$ controls of the database were spatially normalized using the remaining control (randomly selected) as template.
- After spatial normalization, the average of the $n$ controls was computed.
- The average image was made symmetric and used as template.

Finally, the images were reduced in order to improve the computational complexity of the system. The reduction was performed by averaging blocks of $2 \times 2 \times 2$ voxels, thus, the final images were a size of $37 \times 37 \times 23$ voxels.

Once the images have been properly normalized, they were visually labeled by three nuclear medicine specialists from the hospital using only the information contained in the images, without any other medical information. The criteria followed, reached by consensus among the three specialists, were: a study was considered to be normal when bilateral, symmetrical uptake appeared in caudate and putamen nuclei, and abnormal when there were areas of significant reduced uptake in any of the striatal structures. Quantitative evaluation was not regarded for image assessment. It is worth noting that most of the patients are borderline cases.

III.B. Automatic classification based on PLS and SVM

The $^{123}$I-ioflupane radiopharmaceutical provides brain images with high activation in the striatum, a region of high interest for the diagnosis of PS. Figure 1 shows 4 axial slices of two $^{123}$I-ioflupane-SPECT images, one of them from a healthy subject and the other one from a patient. Observe that the most of the activity is gathered in the striatum. However, the images contain a lot of information (a large number of voxels) that is not relevant for the diagnosis of the disorder. The main feature of the proposed method is to automatically extract the two high-intensity regions and analyze them separately. Thus, the first stage consists of separating the two regions by dividing the axial slices into two halves. After that, two volumes are obtained, each one containing one hemisphere of the brain. Then, a binary mask is applied to each volume in order to select only the high-intensity voxels.
of the striatum area. This mask is computed as follows:

\[ m_i = \begin{cases} 
1 & \text{if } c_i \geq \frac{1}{2} \max c_i, \\
0 & \text{otherwise}, 
\end{cases} \]

where \( m_i, i = 1 \ldots n \) is the value (0 or 1) at position \( i \) of the mask, \( c_i, i = 1 \ldots n \) is the intensity of the voxel at position \( i \) of an intermediate image, \( c \), and \( \max c \) is the highest intensity of \( c \). The image \( c \) is computed by squaring and making symmetric (regarding to the axial axis) the average of all controls of the database. Applying this mask allows to select the voxels whose intensity is high (compared with the maximum intensity) in healthy subjects. In practice, this is equivalent to select the voxels of the striatum. Subsequently, the PLS decomposition is performed to obtain the score vectors that will be used as features by the statistical classifier. Therefore, a given image is independently represented and processed by two score vectors, one of them models the high-intensity region of the left hemisphere of the brain and the other one models the high-intensity region of the right hemisphere.

Finally, a SVM classifier is used to estimate the underlying class of the image (normal or PS). The classification is carried out for the two score vectors separately, obtaining two estimations for the class of each image. Both estimations are logically combined as follows: if at least one is positive (PS), the image is labeled as normal. Figure 2 illustrates the whole procedure.

IV. EXPERIMENTS AND RESULTS

In order to evaluate the proposed method, we have used the image database described in Sec. III.A to calculate several measures. Then, these measures have been compared with the ones obtained by a previous approach.

The implementation of the experiments consists of two procedures. First, the striatum’s voxels of each hemisphere are selected for all the images. This procedure involves the division of each image into two halves (the two hemispheres) and the selection of the high-intensity area of each hemisphere by means of a binary mask as it is described in Sec. III.B (pseudocode is shown in algorithm 1). Second, the accuracy estimation is performed following a leave-one-out (LOO) cross-validation (CV) strategy: the classifier is trained as many times as the size of the database and in each iteration an image is used for test and the remaining ones for training. The global accuracy is then calculated as the average of the accuracy achieved in each iteration. The PLS score vectors for the test images are computed as follows:

\[ T = VW, \]

where \( T \) is the score vector for a given image, \( V \) are the selected voxels for that image rearranged in a vector form, and \( W \) is a weight matrix obtained from the PLS decomposition of the remaining images of the database (see the Appendix for further details). Computing the score vectors that way avoids that the label of the test images are taken into account for the calculation.

Since each hemisphere is used separately, a given image is represented by two score vectors, modeling the two high-intensity regions of the image. The first scores of the two vectors contain the most part of the covariance between images and labels and can be used to distinguish between pathological images and controls.

Once the scores vectors are extracted for both, test image and training set, the classification is performed by means of
**Algorithm II.** Feature extraction and classification.

**Input:** The labels, \( Y \), and the volumes for the high intensity area of the left and the right hemispheres, i.e. \( DB^*_L \) and \( DB^*_R \).

**Output:** Accuracy, sensitivity and specificity rates

```plaintext
foreach \( I_L, I_R \in DB^*_L \) and \( DB^*_R \) respectively do
  foreach \( k \in [L, R] \) do
    \( X_k = \) select voxels from all but \( I_k \) images in \( DB^*_k \);
    \( Y_k = \) select labels for volumes in \( X_k \) from \( Y \);
    \( [T_k, P_k, W_k, \ldots] = \text{SIMPLS}(X_k, Y_k); \)
    \( V_k = \) rearrange \( I_k \) into a vector form;
    \( T_k = V_k W_k^*; \)
    \( SVM_k = \) train a SVM classifier using \( T_k \) and \( Y_k \);
    \( class_k = \) predict class for \( T_k \) using \( SVM_k \);
  end
  \( class_i = class_{I_L} \) OR \( class_{I_R}; \)
  if \( class_i \neq Y_i \) then
    \( \text{errors} = \text{errors} + 1; \)
  end
  update accuracy, sensitivity and specificity rates;
end
```

a SVM classifier. As it is shown in algorithm 2, each hemisphere of the brain is classified individually obtaining two predictions per image. If one of them is positive (classified as pathological), the whole image will be labeled as such. Figure 3 shows the accuracy, sensitivity, and specificity obtained as a function of the number of scores which are used as input features, i.e., as a function of the feature vector dimension. The SVM classifier uses a radial basis function (RBF) as kernel with parameters \( \sigma \) and \( C \) equal to 5 and 1, respectively. The highest accuracy rates are obtained using few components (five or less) and then they remain quite stable and around 94%. This stability is numerically estimated by means of the mean and the standard deviation shown in Table II. The results are also compared with the ones obtained by means of the baseline approach,\(^7\) which to our knowledge, is the best so far.

The kernel function used along the SVM classification and its parameters were empirically chosen. For this purpose, we computed the accuracy of several possibilities for parameters \( C \) and \( \sigma \) by following a leave-one-out strategy. The mean and the maximum accuracies achieved with each pair are gathered in Fig. 4.

### Table II. Mean, standard deviation, and maximum of the 20 measures obtained with the proposed methodology by varying from 1 to 20 the size of the feature vectors. These rates are compared with the ones obtained in Ref. 7.

<table>
<thead>
<tr>
<th></th>
<th>Proposed approach</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>Mean: 93.39% Std: 1.01 Max: 94.71%</td>
<td>Mean: 84.71% Std: 9.30 Max: 89.42%</td>
</tr>
<tr>
<td><strong>Sensitivity</strong></td>
<td>Mean: 93.16% Std: 1.00 Max: 94.74%</td>
<td>Mean: 86.37% Std: 7.04 Max: 91.58%</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>Mean: 93.62% Std: 2.18 Max: 95.74%</td>
<td>Mean: 83.03% Std: 11.84 Max: 88.30%</td>
</tr>
</tbody>
</table>

The main differences regarding the methodology proposed in this work are as follows:

- **Preprocessing:** Spatial normalization is carried out in the same way. Both methods use the SPM software, which perform a linear affine transformation followed by non-linear warping using basis functions. However, we perform the intensity normalization by equaling the average value of the background of all images instead of using a volume centered on the occipital cortex to scale the intensity of the voxels. In addition, we do not need to...
apply a mask to remove voxels outside the brain since we select only the voxels of the striatum (in practice, the voxels with an intensity above the 50% of maximum intensity in controls).

- **Handling of the nonbilateral disease:** In some patients, the disorder affects only one side of the brain before progressing to bilateral disease. While in the previous work the authors face the problem realigning the images to cause the hemispheres with the maximum uptake to be on the same side, we try to detect these cases and diagnose them as pathological by analyzing and classifying the hemispheres separately.

- **Feature extraction:** The main difference between PCA (Ref. 7) and the proposed PLS is that the latter creates orthogonal weight vectors by maximizing the covariance between the images and their labels. Thus, PLS not only considers the variance of the samples but also considers the class labels. Using PLS, we get feature vectors with more “information” about how to distinguish between controls and PS images. In a sense, PLS is a supervised approach and PCA is unsupervised.

- **Classification:** We have chosen a SVM classifier which usually reports highest accuracy rates and generalization ability in Refs. 19–21 which are usually affected by the “small sample size” problem. The hyperplanes of decision defined by SVM are shown in Fig. 5.

There are other widely used commercial systems for similar purpose, however, they have a different philosophy focused on measuring the intensity/activity of a set of previously defined regions and not on the classification. The main advantage of the methodology presented in this work is that it is fully automatic. When the system is trained, it is able to provide a diagnosis from only a $^{123}$I-ioflupane SPECT image, i.e., no prior knowledge about the disease and its ROIs is required. However, it is worth noting that the system only reproduces the knowledge contained in the labeled images used for training. The results reported are a measure about how similar are
the diagnosis given by a CAD system and the given one by a clinician team. Thus, if the training set is large enough and assuming the labels are correct (gold standard), a CAD system for PS based on the proposed methodology provides a diagnosis as reliable as the one provided by a clinician team but including the robustness of the SVM based machine learning.

All the classification results shown in this work were achieved following a LOO CV scheme. CV is an effective method for estimating the actual risk of the classifier, however, the error estimated that way could be biased if the steps of the algorithm, such as classifier parameter tuning, feature selection or classifier type selection are left out during each CV loop. In this sense, this principle is satisfied ensuring that: (i) the data used to test the classifier are not part of the data used to train it and (ii) the inclusion of the feature selection in a nested CV loop as shown in Sec. IV.

V.A. Receiver operating characteristic (ROC) analysis

A ROC curve is a plot of the trade-off achieved between sensitivity and specificity for a classification procedure. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The optimal solution is located in the upper left corner and corresponds to a sensitivity and specificity of 100%. Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the procedure. A value to measure this accuracy is provided by the area under the curve (AUC).

Figure 6 shows a ROC curve for the proposed PLS and SVM based approach and compares it with the ones for the two intermediate classification processes (one per hemisphere) and with the one obtained if the brain is treated as a whole (without division into hemispheres). In addition, a curve for the baseline approach is shown. The two curves for intermediate one-hemisphere classification achieve similar accuracy, which suggests that the two hemispheres are equally appropriate to diagnose the disease. When the two hemispheres are treated as a whole, the results are better, however, they are not as good as possible because the cases with nonbilateral disease are not correctly diagnosed. Finally, the proposed method, described in Sec. III.B, also takes into account the information from both hemispheres but separately...
and achieves a more accurate ROC curve that significantly improves the baseline.

V.B. PLS-brains

Similar to the concept of eigenfaces,27 it is possible to define the concept of eigenbrains or, more precisely, PLS-brains (please see Fig. 7). As it has been shown in Sec. II.A, the PLS algorithm decomposes the data in two matrices known as scores and loadings [see Eq. (1)]. The latter would be viewed as elementary brain images and the former would represent the quantity of loadings used for building a specific image. Since the decomposition is performed so that the majority of the covariance is gathered in the first PLS-brains, the first scores are enough for classification. In addition, a given image may be reconstructed from only the first scores and loadings getting a filtered image (without low covariance information). Figure 8 shows two images from the database (one control on the left and one PS image on the right) and the same images reconstructed from their two first scores and loadings. Note the difference in the shape of the high-intensity areas in the striatum. This image uses only two scores and loadings for reconstruction because this number of components allows for reconstruction because this number of components allows.

FIG. 8. Central slices of a control image (left column) and a PS image (right) compared with same slices reconstructed after PLS decomposition but using only the two first scores and loadings.

VI. CONCLUSIONS

In this work, an improved approach to develop a fully automatic CAD system for Parkinsonism based on 123I-ioflupane SPECT images has been presented. The proposed methodology independently analyzes each hemisphere of the brain with the purpose of properly diagnosing early stages of the disease when only one side of the brain is affected.17,18 First, the brain volumes were masked to extract high-intensity voxels, usually located in the striatum area. Then, a reduced score vector was obtained per volume using a PLS algorithm. Finally, the state (normal or pathological) of a given image was determined by means of a SVM classifier.

In order to evaluate this approach, a database consisting of 189 123I-ioflupane SPECT images was used. The images have been visually labeled by experienced clinicians as control (94 images) and PS (95 images). The experiments performed yielded an accuracy rate over 94%, outperforming other previous approaches.

ACKNOWLEDGMENTS

This work was partly supported by the Ministerio de Ciencia e Innovación (Spain) (MICINN) under the TEC2008-02113 project and the Consejería de Innovación, Ciencia y Empresa (Junta de Andalucía, Spain) under the Excellence Projects P07-TIC-02566, P09-TIC-4530, and P11-TIC-7103.

APPENDIX: SIMPLS ALGORITHM

The SIMPLS algorithm28 was proposed by Sijmen de Jong in 1993 as an alternative to the NIPALS algorithm for PLS. The main difference between the initial algorithm, NIPALS, and SIMPLS is the kind of deflation: In the latter, no deflation of the centered data matrices X and Y is made, but the deflation is carried out for the cross-product matrix \( S = X^T Y \) between the\( x \)-data and \( y \)-data.29

Algorithm III. SIMPLS (Ref. 29)

Input: \( X \) and \( Y \)
Output: \( W \) and \( T \)
\[
S_0 = X^T Y
\]
for \( j = 1 \) to \( n \) do
    if \( j = 1 \) then \( S_j = S_0 \);
    else \( S_j = S_{j-1} - P_{j-1}(P_{j-1}^T P_{j-1})^{-1} P_{j-1}^T S_{j-1} \);
    Compute \( w_j \) as the first singular vector of \( S_j \);
    \[
w_j = \frac{w_j}{(w_j^T w_j)};
\]
    \( t_j = X w_j \);
    \( p_j = X^T t_j \);
    \( P_j = [p_1, p_2, \ldots, p_{j-1}] \);
    Use \( w_j \) as the \( j \)-th column of \( W \);
    Use \( t_j \) as the \( j \)-th column of \( T \);
end

Medical Physics, Vol. 39, No. 7, July 2012