Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [^{123}I]FP-CIT SPECT images

Francisco P. M. Oliveira[1,2], Diogo Borges Faria[3,4], Durval C. Costa[2,3], Miguel Castelo-Branco[1] and João Manuel R. S. Tavares[4]

[1]Institute for Nuclear Sciences Applied to Health (ICNAS), and Institute for Biomedical Imaging and Life Sciences (IBILI), Faculty of Medicine, University of Coimbra, Portugal

[2]Champalimaud Centre for the Unknown, Champalimaud Foundation, Lisbon, Portugal

[3]HPP Medicina Molecular, SA., Porto, Portugal

[4]Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

Corresponding author:
Francisco P. M. Oliveira
Champalimaud Centre for the Unknown, Fundação Champalimaud
Av. Brasília s/n, Doca de Pedrouços
1400-038 Lisboa, Portugal
Email: francisco.oliveira@fundacaochampalimaud.pt

ORCID of the authors:
Francisco P. M. Oliveira: 0000-0001-9468-8894
Diogo Borges Faria: 0000-0003-4188-6330
Durval C. Costa: 0000-0001-8039-4924
Miguel Castelo-Branco: 0000-0003-4364-6373

João Manuel R. S. Tavares: 0000-0001-7603-6526

Acknowledgments

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on this study, visit the www.ppmi-info.org website. PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Avid, Biogen Idec, Bristol-Myers Squibb, Covance, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, UCB.

Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [^{123}I]FP-CIT SPECT images

Abstract

Purpose This work aimed to assess the potential of a set of features extracted from [^{123}I]FP-CIT SPECT brain images to be used in the computer-aided "*in vivo*" confirmation of dopaminergic degeneration and therefore to assist clinical decision to diagnose Parkinson's disease.

Methods Seven features were computed from each brain hemisphere: five standard features related to uptake ratios on the striatum and two features related to the estimated volume and length of striatal region with normal uptake. The features were tested on a dataset of 652 [^{123}I]FP-CIT SPECT brain images from the Parkinson's Progression Markers Initiative. The discrimination capacities of each feature individually and groups of features were assessed using three different machine learning techniques: support vector machines (SVM), *k*-nearest neighbours and logistic regression.

Results Cross-validation results based on SVM have shown that, individually, the features that generated the highest accuracies were the length of the striatal region (96.5%), the putaminal binding potential (95.4%) and the striatal binding potential (93.9%) with no statistically significant differences among them. The highest classification accuracy was obtained using all features simultaneously (accuracy 97.9%, sensitivity 98%, and specificity 97.6%). Generally, slightly better results were obtained using the SVM with no statistically significant difference to the other classifiers for most of the features.

Conclusions The length of the striatal region uptake is clinically useful and highly valuable to confirm dopaminergic degeneration "*in vivo*" as an aid to the diagnosis of Parkinson's disease. It compares fairly well to the standard uptake ratio-based features, reaching, at least, similar accuracies and is easier to obtain automatically. Thus, we propose its day to day clinical use, jointly with the uptake ratio-based features, in the computer-aided diagnosis of dopaminergic degeneration in Parkinson's disease.

Keywords DaTSCAN^{TM}; uptake ratios; dimensional-based features; classification; PPMI database.

Introduction

Parkinson's disease (PD) is a progressive neurological disorder characterized by tremor, rigidity, and bradykinesia associated with progressive dopaminergic neuronal loss projecting into the striatum originated from the substantia nigra and other brain structures [1].

The progressive degeneration of nigrostriatal dopaminergic neurons can be assessed by using radioligands in imaging-based approaches. Dopamine transporters (DAT) loss in the presynaptic terminals of the nigrostriatal pathway is a key feature of PD. Imaging of the DAT with [^{123}I]FP-CIT (commercially available as DaTSCAN^{TM}) and Single Photon Emission Tomography (SPECT) is the most widely used Nuclear Medicine technique for routine confirmation of dopaminergic degeneration during the assessment of patients with suspected degenerative Parkinsonism [2]. With this technique, the reduction of dopamine transporters can be observed in patients with PD, Multisystem Atrophy, Progressive

Supranuclear Palsy, Corticobasal Degeneration and also in patients with dementia with Lewy bodies [3-4]. The reduction of [^{123}I]FP-CIT specific uptake in the putamen confirms PD and excludes the diagnosis of other diseases running with no dopaminergic degeneration, such as essential tremor or drug-induced Parkinsonism [5]. Frequently, results are based on qualitative visual evaluation, although several quantification methods have been developed and proposed for more objective assessments [6].

Striatal uptake ratios computed from the [^{123}I]FP-CIT SPECT brain images have been used for image classification [7-14] often using each ratio individually. The use of machine learning techniques can improve the automated classification accuracy, since discriminative features can be simultaneously used to build a more robust multidimensional classification model than the models based on a single feature.

The main goal of this work was to evaluate the potential of several features (individually and in groups) extracted from [^{123}I]FP-CIT SPECT brain images to detect and confirm dopaminergic degeneration and therefore, to assist a clinical decision for the diagnosis of PD. Hence, three machine learning based classifiers were used: support vector machines (SVM), *k*-nearest neighbour (*k*-NN) and logistic regression (LR). Seven features automatically extracted from the images were considered: five standard features related to uptake ratios in the striatum, and two features related to the length and volume of the striatal region with normal uptake.

The dimensions of the striatal uptake region, especially the length from the caudate head (most anterior) to the most posterior putamen contours, are key factors in the visual assessment of the [^{123}I]FP-CIT SPECT brain images made by physicians. However, as far as the authors know, no work has been reported using these dimensional SPECT-DAT biomarkers and their automatic extraction from the images for classification. Thus, the

development of an effective and robust computational algorithm to extract these dimensional features was another goal of this work.

Material and methods

Dataset

Data for this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data) available on August 10, 2015. The dataset contained all 652 pre-processed [^{123}I]FP-CIT SPECT brain images acquired at the Screening stage, for the groups: control female (73 images), control male (136 images), PD female (157 images) and PD male (286 images). Overall, the healthy control (HC) subjects' age was 61.8±11.3 years old, and the PD subjects' age was 61.7±9.7 years old. No statistically significant difference was found between the age of the HC and PD patients (Student *t*-test, $p = 0.320$).

PPMI is a longitudinal, multicentre study to assess the progression of clinical features, imaging and biologic markers in PD patients and HC. All the PD subjects were at an early stage of the disease, Hoehn and Yahr stage I or II at baseline, with a PD diagnosis obtained less than or equal to 2 years at Screening. PD subjects had a confirmation from imaging core that screening dopamine transporter SPECT scan is consistent with dopamine transporter deficit. All subjects were assessed at the baseline and every three to six months thereafter (further details of the inclusion and exclusion criteria can be found at the Study Protocol at http://www.ppmi-info.org/study-design/research-documents-and-sops/).

Since different imaging systems were used, the imaging acquisition protocol was standardized at each clinical site, and customized study phantoms were used to enable an

ongoing assessment of each camera [15]. SPECT scans lasted for 30 to 45 minutes and started 3.5 to 4.5 hours post injection of an activity between 110 to 185 MBq. The raw projection data was transferred to the imaging core lab and then iteratively reconstructed, using a hybrid ordered-subsets expectation maximization (OSEM) algorithm in a HERMES workstation (Hermes Medical Solutions, Sweden). The reconstructed images were then transferred to the PMOD software (PMOD Technologies, Zurich, Switzerland) for subsequent processing, and for attenuation correction. The final pre-processed images were saved in a DICOM format using 91×109×91 cubic voxels with 2 mm width.

Automated quantification

The quantification process developed can be divided into the following main sequential steps (Figure 1): template-based image registration [16], i.e. alignment of the image under study with a pre-built 3D [^{123}I]FP-CIT SPECT brain template image; definition of the regions of interest (ROIs); computation of the striatal uptake ratios; building of the voxel-wise binding potential image and assessment of the dimensions of the striatal uptake region; and finally, assembly of the feature vectors to be used in the classification.

Image registration

We used a registration algorithm validated in a previous study [17]. This algorithm is based on the iterative minimization of the sum of squared differences between the image under study and a [^{123}I]FP-CIT SPECT brain images template, using a rigid geometric transformation. The optimizer chosen is the Powell's method with the Brent's method available in the Insight Toolkit (ITK) software library [18]. Just for image registration step, the SPECT images of the patients are intensity normalized by the 99th percentile. The template has the orientation of the MNI space.

The registered images are resampled with a width of 2.2 mm, which is approximately half the voxel width normally used in routine clinical reconstruction of [^{123}I]FP-CIT SPECT images.

Definition of the volumetric ROI

The automated computation of the uptake ratios is based on the counts inside the volumetric ROIs placed over the registered image under study. Seven ROIs are defined and used: a large ROI over the whole nonspecific brain region that is used as the reference region, and three ROIs on each hemisphere of the brain: one over the putamen, one over the caudate, and one large ROI containing the entire striatum and surrounding area (Figures 2 and 3).

The definition of the reference ROI and the large ROI over the striatum are based on the work of Tossici-Bolt *et al* [19]. However, in the present work, all procedures are performed in 3D; while the original work was performed in a 2D image built from adding a set of transaxial slices that contained the striatum.

The large ROIs (one on each hemisphere) over the striatum are 44 mm thick (inferior-superior plane), 44 mm wide (left-right plane), 62 mm long (posterior-anterior plane), with the shape shown in Figure 2. These ROIs were defined to be large enough to take into account the partial volume effects and inter-subject shape variations.

In the dataset used, there were some situations when the ROIs over the striatum were not totally inside the external contours of the brain. In these cases, the voxels outside the brain were excluded. This occurred with subjects who had very small heads and when the original image did not contain the whole brain due to image acquisition flaws. Although these images were more complex to be analysed successfully, they were not excluded

from this study as the striatal region was intact and also because this situation can occur in routine clinical DaTSCAN exams.

The reference ROI is found following these steps: first, the registered image under study is smoothed with a Gaussian filter (16 mm full width at half maximum -FWHM); then it is binarized using a threshold level of 25% of the maximum level after smoothing; afterwards, a morphological erosion operation is carried out to shrink the binary ROI by approximately 16 mm on the top and around the head, in order to exclude the periphery of the head; finally, the volumes corresponding to the large ROIs over the striatum and the non-brain slices on the bottom of the head are removed (Figure 2). The width of the smoothing kernel, the threshold level, and the erosion operation were established experimentally and based on the authors' knowledge, being some of these parameters similar to the ones used in previously published work [19]. The goal of these operations is to guarantee that only brain voxels are included in the reference ROI, as much as it is possible and/or feasible. Using this methodology, the reference ROI shrinks or enlarges accordingly to the content of the image. For instance, if the occipital cortex is partially cut, only the portion identified in the image will be included in the reference ROI.

The ROIs of the caudate and putamen have predefined shapes (Figure 3) and are always placed in the same set of 9 mm thick slices in the middle of the striatum. In the transaxial plane, the caudate ROI is approximately an ellipse with the axes of 19 mm and 22 mm, and the putamen ROI has the same maximum width as the caudate but a length of 26 mm. The putamen ROI contains the putamen and globus pallidus. The shape of these ROIs and their inferior-superior position in the brain were established based on the SPECT template. Their placement and orientation on the axial plane can be adjusted accordingly with each image under study based on an optimization algorithm that chooses the placement and orientation to maximize total counts within the ROIs. To guarantee the

robustness of the ROI placements, even for images with very low specific uptake in the striatum, only small rotations relatively to the template are allowed. The ROIs defined for one of the sides move independently of the ROIs defined for the other side. For each side, the caudate and putamen ROIs move together.

Computation of the uptake ratios

The specific binding ratio (SBR) represents the ratio between the counts concentration in the striatum due to the specific binding only and the count concentration in the reference region due to the free and nonspecific binding [19]. For each side of the striatum, it is given by:

$$SBR = \frac{V_{VX}}{V_s} \times \frac{Ct_{ROI} - R \times V_{ROI}}{R},$$

where V_s is the volume of the striatum, Ct_{ROI} is the total count in the large striatum ROI, R is the mean count per voxel within the reference ROI, V_{ROI} is the volume of the large striatum ROI in numbers of voxels, V_{VX} is the volume of a single voxel. Since it is impossible to assess the volume of the striatum for each person based on the [^{123}I]FP-CIT SPECT brain image, a standard value of $V_s = 11.2$ ml was used as in Tossici-Bolt *et al* [19].

The developed algorithm automatically computes the binding potential (BP_{ND}) at three specific regions: the caudate binding potential (CBP), the putamen binding potential (PBP), and the striatal binding potential (SBP). Following Innis *et al* [20], BP_{ND} refers to the ratio at equilibrium of specifically bound radioligands to that of nondisplaceable radioligands in tissue. Assuming a two-tissue compartment model: $BP_{ND} = (V_T - V_{ND})/V_{ND}$, where V_T is the radioligands volume of distribution of the target region, and V_{ND} is the radioligands volume of distribution of the non-displaceable uptake.

V_{ND} is the sum of the volume of distribution of the free and non-specifically bound ligand. Since V_T and V_{ND} are directly proportional to the counts per voxel of the target and reference regions, respectively, then $CBP = (C - R)/R$, $PBP = (P - R)/R$, and $SBP = (S - R)/R$, where C is the mean count per voxel in the caudate ROI, P is the mean count per voxel in the putamen ROI, S is the mean count per voxel in the striatum (here caudate and putamen ROIs), and R is the mean count per voxel within the reference region. The putamen to caudate ratio (PCR) is also computed and is given by $PCR = (PBP)/(CBP)$.

Dimensional analysis of the striatal uptake region

As stated in the previous section, the binding potential can be defined at the voxel level as $BP_{ND}(x,y,z) = (I(x,y,z) - R)/R$, where $I(x,y,z)$ is the count associated to the voxel with coordinates (x,y,z), and R is the mean count per voxel within the reference ROI.

The BP_{ND} images are automatically built based on the registered images. As such, they can easily be used for comparison: intra-subject, inter-subjects, or a subject against the mean image of a reference population. Figure 4 shows an example of a BP_{ND} image built.

The functional segmentation, i.e. extraction, of the region with uptake is performed on the BP_{ND} image, and is based on a threshold, or cut-off level, i.e. only voxels with a BP_{ND} higher than the cut-off level defined are included in the segmented region. After the initial cut-off, the binary image obtained is cleaned of scattered voxels by selecting only the larger group of connected voxels in each hemisphere. The total volume of the two sides of the segmented striatal region and the corresponding lengths are automatically calculated from the segmented image based on the orientation established for the template image (Figure 5). When there are no voxels with BP_{ND} greater than the cut-off level defined, the dimensional features are set to zero. It should be noted that the goal of the

implemented image segmentation process is not to extract the anatomical striatal region, as we are only carrying out a functional segmentation; on the contrary, we expect the extracted region to be significantly different from the anatomical region in individuals with PD.

The optimal cut-off level is chosen based on the classification accuracy, i.e. the cut-off value that originates the best discrimination between HC and PD patients using the volume and length dimensional measures is chosen.

Validation of the methodology

The methodology was validated in two main aspects: the placement of the ROI and the classification accuracy using cross validation. The placement of the ROI was totally automatic, and visually examined by an expert just for quality control. To assess the classification accuracy, the well known leave-one-out cross validation (LOOCV) paradigm was used and three machine learning techniques were considered: linear SVM, LR and *k*-NN.

Classification

In this work, the classification problem is to decide if a [^{123}I]FP-CIT brain image is from a subject with dopaminergic degeneration, i.e., PD or from a HC. The decision is based on a set of data for which the classification is known (training dataset).

As the designation suggests, LOOCV method involves using the features from a single subject from the original dataset for validation, i.e. to be classified, and the remaining subjects as the training data. This procedure is repeated until all subjects are used once as validation data. Thus, for the dataset used, 652 training datasets are built (each one with 651 subjects). Finally, comparing the gold standard diagnosis and the classification of

each subject using the LOOCV, the accuracy, sensitivity, and specificity of the classification model are estimated.

The classification accuracy for the image dataset was assessed using each feature individually and three groups of features: a) uptake ratio-based features (SBR, CBP, PBP and PCR), b) dimensional-based features (volume and length) and c) the entire set of features altogether, i.e., a) plus b). In the uptake ratio-based group and all features group, the SBP was not included since it is a linear combination of the CBP and PBP features. With these feature-specific classification models, we assessed the potential of the features under study for classification, and thus gained more insights into the problem of [^{123}I]FP-CIT brain image classification.

Two values of each feature are used for classification: one from the left side and the other from the right side of the brain. Thus, in practice, the dimension of each feature vector is twice the number of features used.

Statistical analysis

Statistical comparison among the mean values of the uptake ratios and dimensional features were assessed using *t*-tests. Statistical comparison among the accuracies obtained by the features individually and by the groups of features was done using the Cochran's Q test followed by the pairwise *post hoc* Dunn's test with Bonferroni correction for multiple comparisons (p_{adj}). The statistical comparison was conducted using SPSS 20.0 (SPSS Inc., Chicago, USA), and a significance level of 0.05 was established.

Implementation

The computational algorithm was fully implemented in C++ programming language and compiled using the compiler Microsoft Visual Studio 2008. The executable program built

by the compiler can run in any computer with the Windows operating system, without the need for any extra software package or hardware. We tested the program on a notebook PC with an Intel I7-2670QM microprocessor, 8 GB of RAM, and running Microsoft Windows 7. Techniques of image processing and visualization were partially implemented using the following free open source toolkits: The CImg Library, Insight Toolkit (ITK) [18] and Visualization Toolkit (VTK) [21], which allow compile the code to run on Linux or MAC operating systems. The tasks of learning and classification were partially implemented using the LIBSVM [22] and MATLAB R2014a (The MathWorks, USA).

Results

Robustness assessment

The robustness of the methodology was assessed directly by visual inspection of the image registration and placement of the ROI. In terms of the image registration, in approximately 95% of the subjects, the image registration was adequate using the default parameters. For the other 5%, it was necessary a small change in the intensity normalization parameters. All ROI were adequately placed.

Estimation of the optimal cut-off level for the functional segmentation

To test which cut-off level is the best for the functional segmentation based on its ability to discriminate between HC and PD, we tested the BP cut-offs 1.0, 1.2, 1.4, 1.6, 1.8, and 2.0. We have used only cut-off levels between 1.0 and 2.0 because cut-off levels inferior to 1.0 tend to include too many voxels that do not belong to the striatum, and cut-off levels superior to 2.0 tend to exclude too many voxels that belong to the striatum. As

shown in figure 6, the two best cut-off levels for the length are 1.4 and 1.6 without significant difference between them. The cut-off 1.6 gave rise to segmentation results in HC subjects visually more similar to the manual segmentations obtained by the experts and it is closer to the mean PBP in HC (Table 1), thus this value was chosen.

Comparison of the uptake ratios and dimensional measures between groups

The mean values for each feature computed from the entire dataset are presented in Table 1. For all features there is a statistically significant difference between the values of the HC and the PD groups (*t*-test, $p < 0.001$). Healthy women have shown statistically significantly higher mean CBP, PBP, SBP and volume than healthy men (*t*-test: $p = 0.002$, $p = 0.003$, $p = 0.008$, respectively).

Classification accuracy

The classification accuracies, sensitivities and specificities obtained using each feature individually and the defined groups of features are shown in Table 2. The accuracies obtained using the three machine learning techniques are very similar and not significantly different among the machine learning techniques for most of the cases. However, for the feature SBR the accuracy obtained using the *k*-NN is significantly inferior than the accuracy obtained using the SVM ($p_{adj} = 0.042$) or Logistic Regression ($p_{adj} = 0.001$); for the length the accuracy obtained using the *k*-NN is significantly inferior to the accuracy obtained using the SVM ($p_{adj} = 0.043$); and for the set of all features the SVM originated an accuracy significantly superior to the accuracy obtained using the Logistic Regression ($p_{adj} = 0.043$).

For all three classifiers, there was a statistically significant difference in the classification accuracies obtained from individual features (Cochran's Q test, $p < 0.001$). The accuracy

obtained using the length was statistically significantly higher than using the SBR, CBP, PCR and volume (Table 3). Although the length has also achieved accuracies higher than the PBP and SBP (Table 2), the differences are not statistically significant (Table 3). The lower accuracies were obtained using the SBR and CBP, with no statistically significant difference between them ($p > 0.091$, $p_{adj} = 1.000$).

Taking into account the defined groups of features, using the SVM classifier the accuracy obtained using the uptake ratio-based features is significantly inferior than the accuracies obtained using the dimensional features ($p = 0.005$, $p_{adj} = 0.014$) or all features ($p < 0.001$, $p_{adj} = 0.001$). There was not a statistically significant difference among these three groups of features using the *k*-NN (Cochran's Q test, $p = 0.202$) and LR (Cochran's Q test, $p = 0.078$) classifiers.

All images that appeared misclassified by the best classification model were examined and, based on the value of the features extracted, all the classifications obtained appeared to be, in fact, correct. However, since the PPMI classification was considered to be the gold standard, those classifications were regarded as incorrect classifications.

Computer processing time

In the notebook computer used, the mean total time necessary for the whole computational process to classify an image was about 10 seconds. This time includes all tasks, from image reading to classification.

Discussion

For all features under evaluation, a statistically significant difference was found between the mean values of the HC and PD groups (*t*-test, $p < 0.001$). Healthy women have shown

statistically significant higher mean CBP, PBP and SBP than healthy men. This is in agreement with the results found by Nobili *et al* [23] based on a database of the European Normal Control Database of DaTSCAN (ENC-DAT) study containing 139 [^{123}I]FP-CIT SPECT scans of HC [24]. Likewise, the mean functional segmented volume in healthy women was also significantly higher than in healthy men.

All features revealed good to very good discriminative power (minimum accuracy obtained was about 86%). The features that individually gave the best accuracy were the length and the PBP. It is not surprising that these two features gave rise to high classification accuracies, since it is known that, compared to HC subjects, there is a reduction of the [^{123}I]FP-CIT uptake in the putamen of PD subjects, with a geometrical pattern of putamen caudal loss, leading to predictable decreases in length of the striatal region with a normal uptake. For all three classifiers, the length always originated the best accuracy, however the difference was not statistically significant when compared with the accuracies obtained using the PBP and SBP features.

Although, at first glance, the length seems to be more vulnerable to artefacts and partial volume effects than the uptake ratio-based features, the solution that was adopted to compute its value proved to be sufficiently robust to be used in the computer assisted diagnosis of Parkinson's disease. On the other hand, the length is easier to assess than the uptake ratios, as the PBP, for instance. Additionally, the uptake ratios are highly dependent on the delineation of the striatal ROIs.

The volume of the striatal region with normal uptake originated high classification accuracies and better than the uptake ratio-based SBR and CBP, but not as good as the length and the PBP.

It was a surprise that the group consisting of only the two dimensional-based features (length and volume) has achieved higher accuracies than the group of all uptake ratio-

base features, although this difference was statistically significant only when the SVM classifier was used. This result shows the high potential of the dimensional-based features to be used in this kind of classification problems.

Using all features simultaneously, a LOOCV classification accuracy of 97.9% (98.0% sensitivity, 97.6% specificity) was achieved using the SVM classifier. This result is remarkable since it was obtained from a large dataset of 652 images, and is highly competitive when compared with the ones reported in the literature. For instance, Koch *et al* [12] reported an accuracy close to 92% on a dataset of 155 images, and Tossici-Bolt *et al* [19] achieved an accuracy of 95% (97% sensitivity and 92% specificity) on a dataset of 55 images. It should be noted that, in both these latter cases, the accuracy was not validated via cross-validation. Using cross-validation and SVM based classifiers, Prashanth *et al* [14] found an accuracy of approximately 96% (96.6% sensitivity and 95.0% specificity) on a PPMI dataset comprising well established pre-computed uptake ratios of 548 subjects. Using SVM and voxel-as-feature (VAF) approach, Illán *et al* [25] claimed an accuracy of approximately 91% (89% sensitivity and 93% specificity) on a dataset of 208 images, and Oliveira and Castelo-Branco [26] found an accuracy of 98% in a dataset from PPMI very similar to the one used here.

In the developmental stage of this study, the gender and age of the subjects were also tested as additional features, but the classification accuracy remained unchanged. This result was not expected, since in healthy subjects of both genders there is a reduction of uptake in the caudate and putamen nucleus with age; besides women have a higher mean uptake than men (Table 1) [23]. May be we are facing some kind of ceiling effect, since the accuracies obtained are very close to 100%.

For most of the features, there is no statistically significant differences among the accuracies obtained using the three machine learning techniques. This shows that the high

accuracy achieved is a consequence of the discrimination power of the features and not due to some particular aspect of the machine learning technique used.

A small number of PD subjects were classified as HC, since the values of all their features were close, and sometimes higher than the mean values of the HC subjects. This was not totally unexpected since the classification made at the PPMI core was based on several clinical criteria plus the dopamine transporter images, while for our work the classification was based just on the dopamine transporter images results.

In this work, a rigid geometric transformation was adopted in the image registration process. Although this option may be seen as a limitation, since the corrections for the size and shape of the subjects' head were not addressed, it guarantees a more robust image registration process than if a non-rigid geometric transformation were to be used. This is because this kind of images has very low signal-to-noise ratio, low resolution and little anatomical information. Besides, the common standard Chang attenuation correction, which depends on the placement of an ellipse around the head on the transaxial slices, may significantly influence the apparent size and shape of the head, and consequently the image registration if non-rigid geometric transformations are used. Non-rigid registration is, therefore, not recommended whenever Chang attenuation correction is used. Whenever available, anatomical images (preferably MRI) of the patients' head should always be used for non-rigid registration algorithms.

We used the whole brain (excluding a large ROI around the striatum) as reference region, as it was done in Tossici-Bolt et al [19], because it includes much higher number of voxels than the occipital or cerebellum. Thus it is less sensitivity to noise or artifacts. Besides, no significant changes in the uptake between healthy controls and patients are expected. In addition the specific uptake in the whole brain is very small comparatively to the striatum. When using the occipital region as reference, normal subjects showed higher

coefficient of variation (standard deviation / mean) for all indices in comparison with whole brain as reference. The cerebellum was not considered due to greater number of subjects with data acquisition that did not include entirely this region.

The choice for the linear SVM, *k*-NN and LR classifiers was based on their popularity and good ability for many common classification problems. However, our intention was not to compare the classifiers, but the features used in the classification process.

Naturally, an exhaustive search for other configurations of the parameters used in the extraction of the features or with the classifiers could lead to slightly different results. However, we strongly believe that our findings are robust to changes in the parameters.

For Nuclear Medicine clinical sites where an anatomical image of the patient's head is available, for instance using a SPECT/CT, a more accurate delineation of the striatal nucleus can be done based on the anatomical image, and consequently improve the accuracy on the assessment of the binding potential. Also, in these cases, we recommend to normalize the functional length and volume features extracted from the [^{123}I]FP-CIT by the length and volume measured in the anatomical images.

To the best of our knowledge, this algorithm is the more complete quantitative evaluation of striatal dopaminergic features so far available. The use of the new dimensional-based features in combination with the standard features based on uptake ratios may enable nuclear medicine experts to improve the accuracy of their diagnostic imaging skills. Other main clinically useful applications of this algorithm are: (a) improvement in inter- and intra-rater diagnostic (pre-synaptic dopaminergic degeneration) agreement; (b) implementation of multivariate quantitative analysis for easier integration into more modern diagnostic automatic criteria; (c) higher discriminative power with some difficult cases; (d) useful as a training tool; (e) quantification of disease evolution from early poorly symptomatic to more symptomatic cases; (f) assessment of efficacy of new

treatments that may delay the disease evolution; (g) quality control in busy nuclear medicine services, since a disagreement between the physician visual evaluation and the machine learning output imply more careful re-evaluation of the entire study.

The dataset used in our study contains only patients diagnosed with idiopathic Parkinson's disease and healthy controls. Thus, the results cannot be extrapolated to other pre-synaptic dopaminergic neuronal loss diseases without further evaluation. Although the diagnosis of idiopathic Parkinson's disease has been made using the most rigorous clinical criteria, the lack of post mortem diagnosis confirmation is, no doubt, a limitation of this study.

In conclusion, when the semi-quantitative ratios and the dimensional-based features are combined there is a statistically significant increase in the classification accuracy in comparison with methods using just the semi-quantitative ratios (SVM classifier, Dunn's test with Bonferroni correction $p_{adj} < 0.001$). The high accuracy obtained using all features under evaluation simultaneously, was not a consequence of the classifier used, but an outcome of the features selected. In addition, generalization can be expected because the results were obtained using a considerably large dataset with a fully automated computational algorithm. The results showed that the length of the striatal uptake region revealed clinical added value, since the accuracy obtained was slightly higher than the best accuracy achieved by the standard uptake ratio-based features (PBP and SBP). In addition, it is easier to assess. Thus, we believe that it should be implemented in the computer-aided confirmation "*in vivo*" of dopaminergic degeneration with [^{123}I]FP-CIT SPECT investigations.

Compliance with ethical standards

Funding This work was partially financially supported by the public projects with the references: CENTRO-07-ST24-FEDER-00205 - "From molecules to man: novel diagnostic imaging tools in neurological and psychiatric disorders"; PTDC/BBB-BMD/3088/2012 - financially supported by Fundação para a Ciência e a Tecnologia (FCT) in Portugal; and NORTE-01-0145-FEDER-000022 - SciTech - Science and Technology for Competitive and Sustainable Industries, co-financed by "Programa Operacional Regional do Norte" (NORTE2020), through "Fundo Europeu de Desenvolvimento Regional" (FEDER).

Conflict of interest None.

Ethical approval We declare that all human data studies have been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

References

1. Tolosa E, Wenning G, Poewe W. The diagnosis of Parkinson's disease. Lancet Neurol 2006;5:75-86. doi:10.1016/S1474-4422(05)70285-4.
2. Varrone A, Halldin C. New developments of dopaminergic imaging in Parkinson's disease. Q J Nucl Med Mol Imaging 2012;56:68-82.
3. O'Brien JT, Colloby S, Fenwick J, Williams ED, Firbank M, Burn D, et al. Dopamine transporter loss visualized with FP-CIT SPECT in the differential diagnosis of dementia with Lewy bodies. Arch Neurol 2004;61:919-25. doi:10.1001/archneur.61.6.919.
4. Kasanuki K, Iseki E, Ota K, Kondo D, Ichimiya Y, Sato K, et al. ^{123}I-FP-CIT SPECT findings and its clinical relevance in prodromal dementia with Lewy bodies. Eur J Nucl Med Mol Imaging 2017;44:358-65. doi:10.1007/s00259-016-3466-6.

5. Cuberas-Borrós G, Lorenzo-Bosquet C, Aguadé-Bruix S, Hernández-Vara J, Pifarré-Montaner P, Miquel F, et al. Quantitative evaluation of striatal I-123-FP-CIT uptake in essential tremor and parkinsonism. Clin Nucl Med 2011;36:991-6. doi:10.1097/RLU.0b013e3182291a7b.
6. Badiavas K, Molyvda E, Iakovou I, Tsolaki M, Psarrakos K, Karatzas N. SPECT imaging evaluation in movement disorders: far beyond visual assessment. Eur J Nucl Med Mol Imaging 2011;38:764-73. doi:10.1007/s00259-010-1664-1.
7. Mirzaei S, Zakavi R, Rodrigues M, Schwarzgruber T, Brücke T, Bakala J, et al. Fully automated 3D basal ganglia activity measurement in dopamine transporter scintigraphy (Spectalyzer). Ann Nucl Med 2010;24:295-300. doi:10.1007/s12149-010-0353-2.
8. Habraken JBA, Booij J, Slomka P, Sokole EB, Royen EAv. Quantification and visualization of defects of the functional dopaminergic system using an automatic algorithm. J Nucl Med 1999;40:1091-7.
9. Jensen PS, Ziebell M, Skouboe G, Khalid U, Nijs Rd, Thomsen G, et al. Validation of a method for accurate and highly reproducible quantification of brain dopamine transporter SPECT studies. J Nucl Med Technol 2011;39:271-8. doi:10.2967/jnmt.111.090324.
10. Calvini P, Rodriguez G, Inguglia F, Mignone A, Guerra UP, Nobili F. The basal ganglia matching tools package for striatal uptake semi-quantification: description and validation. Eur J Nucl Med Mol Imaging 2007;34:1240-53. doi:10.1007/s00259-006-0357-2.
11. Zubal IG, Early M, Yuan O, Jennings D, Marek K, Seibyl JP. Optimized, automated striatal uptake analysis applied to SPECT brain scans of Parkinson's disease patients. J Nucl Med 2007;48:857-64. doi:10.2967/jnumed.106.037432.
12. Koch W, Radau PE, Hamann C, Tatsch K. Clinical testing of an optimized software solution for an automated, observer-independent evaluation of dopamine transporter SPECT studies. J Nucl Med 2005;46:1109-18.
13. Morton RJ, Guy MJ, Clauss R, Hinton PJ, Marshall CA, Clarke EA. Comparison of different methods of DatSCAN quantification. Nucl Med Commun 2005;26:1139-46.
14. Prashanth R, Roy SD, Mandal PK, Ghosh S. Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. Expert Syst Appl 2014;41:3333-42. doi:10.1016/j.eswa.2013.11.031.
15. Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol 2011;95:629-35. doi:10.1016/j.pneurobio.2011.09.005.
16. Oliveira FPM, Tavares JMRS. Medical image registration: a review. Comput Method Biomech Biomed Eng 2014;17:73-93. doi:10.1080/10255842.2012.670855.
17. Oliveira FPM, Faria DB, Costa DC, Tavares JMRS. A robust computational solution for automated quantification of a specific binding ratio based on [^{123}I]FP-CIT SPECT images. Q J Nucl Med Mol Imaging 2014;58:74-84.
18. Ibáñez L, Schroeder W, Ng L, Cates J, Consortium IS. The ITK Software Guide. In. Kitware, Inc. 1999. http://www.itk.org/.
19. Tossici-Bolt L, Hoffmann SMA, Kemp PM, Mehta RL, Fleming JS. Quantification of [^{123}I]FP-CIT SPECT brain images: an accurate technique for measurement of the specific binding ratio. Eur J Nucl Med Mol Imaging 2006;33:1491-9. doi:10.1007/s00259-006-0155-x.
20. Innis RB, Cunningham VJ, Delforge J, Fujita M, Gjedde A, Gunn RN, et al. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. J Cerebr Blood F Metab 2007;27:1533-9. doi:10.1038/sj.jcbfm.9600493.

21. Schroeder W, Martin K, Lorensen B. The Visualization Toolkit. In. Kitware. 1999.
22. Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2:27:1-. doi:10.1145/1961189.1961199.
23. Nobili F, Naseri M, Carli FD, Asenbaum S, Booij J, Darcourt J, et al. Automatic semi-quantification of [^{123}I]FP-CIT SPECT scans in healthy volunteers using BasGan version 2: results from the ENC-DAT database. Eur J Nucl Med Mol Imaging 2013;40:565-73. doi:10.1007/s00259-012-2304-8.
24. Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, et al. European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis. Eur J Nucl Med Mol Imaging 2013;40:213-27. doi:10.1007/s00259-012-2276-8.
25. Illán IA, Górriz JM, Ramírez J, Segovia F, Jiménez-Hoyuela JM, Lozano SJO. Automatic assistance to Parkinson's disease diagnosis in DaTSCAN SPECT imaging. Med Phys 2012;39:5971-80. doi:10.1118/1.4742055.
26. Oliveira FPM, Castelo-Branco M. Computer-aided diagnosis of Parkinson's disease based on [^{123}I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines. J Neural Eng 2015;12:026008. doi:10.1088/1741-2560/12/2/026008.

FIGURE CAPTIONS

Fig. 1 Flow diagram for the quantification of a [^{123}I]FP-CIT SPECT brain image

Fig. 2 Visualization on axial, coronal and sagittal planes of the striatal large ROI and reference ROI defined on an example image

Fig. 3 The 3D ROIs defined on the caudate and putamen of an example image in the axial, coronal and sagittal planes

Fig. 4 Example of a BP_{ND} image built from a HC subject: On the left, an axial slice is visible; middle top, a coronal slice; middle bottom, a sagittal slice; and on the right, the dimensionless BP_{ND} colour scale used (only voxels with a positive BP_{ND} are shown)

Fig. 5 Example of the segmented region with normal uptake obtained from an HC subject. The arrows indicate the directions defined for measuring the length

Fig. 6 Variation of the LOOCV accuracy as a function of the segmentation cut-off level and using a SVM classifier

TABLE CAPTIONS

Table 1 Mean and standard-deviation values of each feature obtained from the four groups of subjects included in the experimental dataset

Table 2 Results of the cross-validation accuracy (Acc.), sensitivity (Sen.), specificity (Spe.) for the three classifiers tested using the LOOCV approach (values are in percentage)

Table 3 *P*-values for pairwise comparison among the accuracy obtained using the length against the accuracies obtained using the other features (p_{adj} is the p value corrected for multiple comparisons)

FIGURES

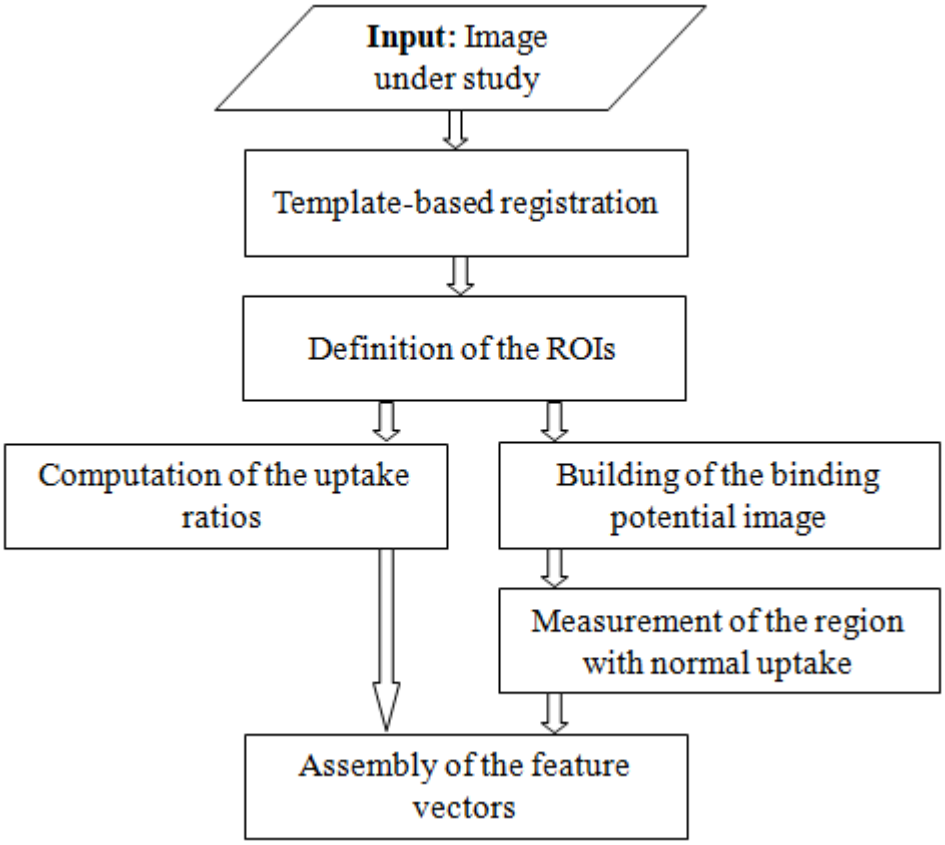

Figure 1

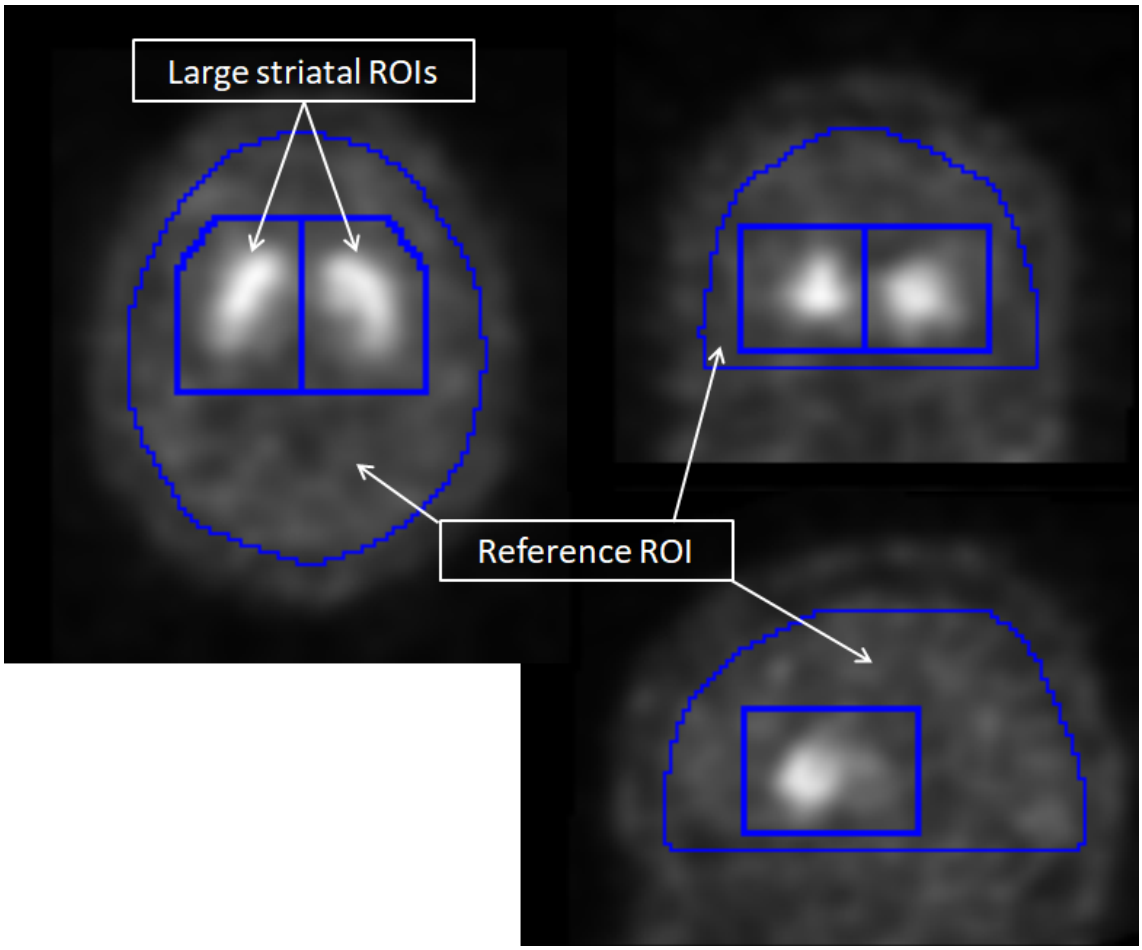

Figure 2

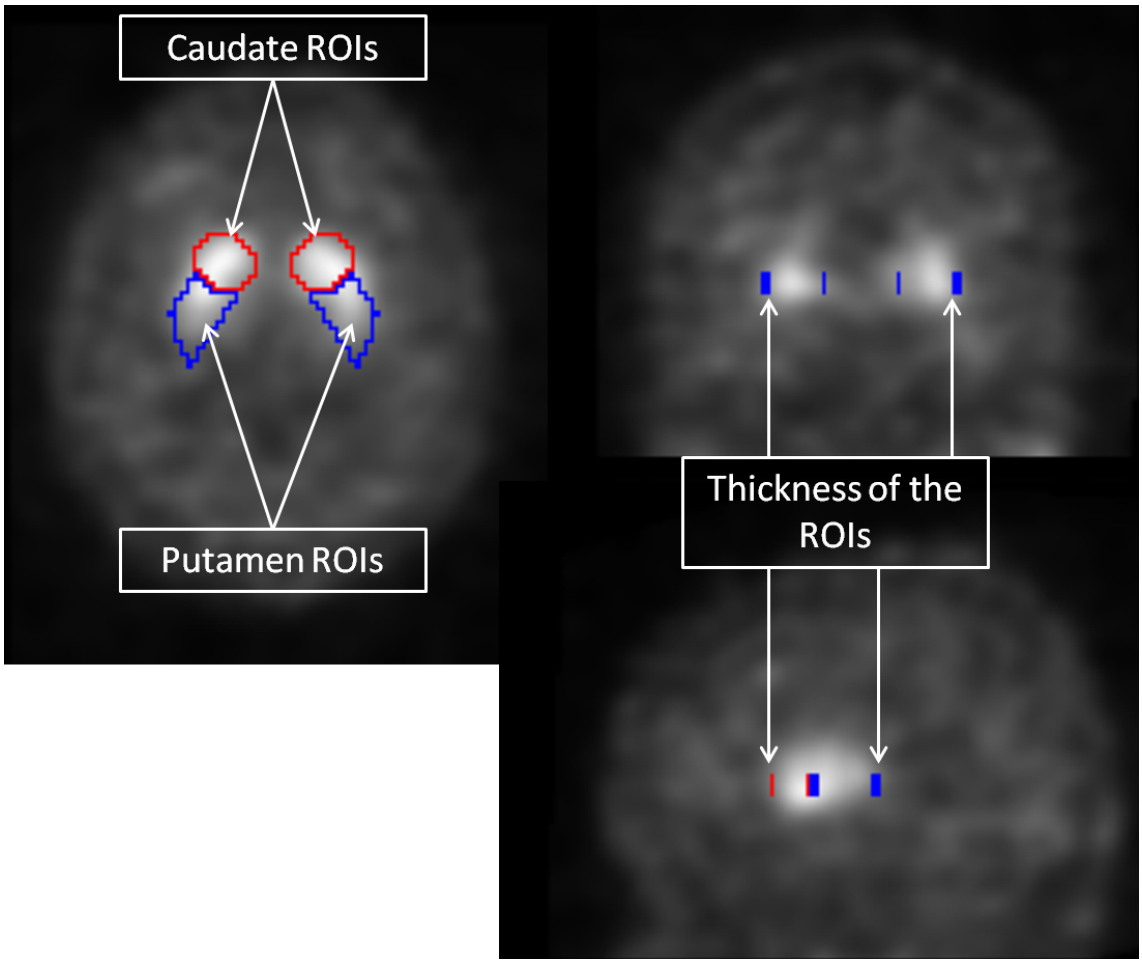

Figure 3

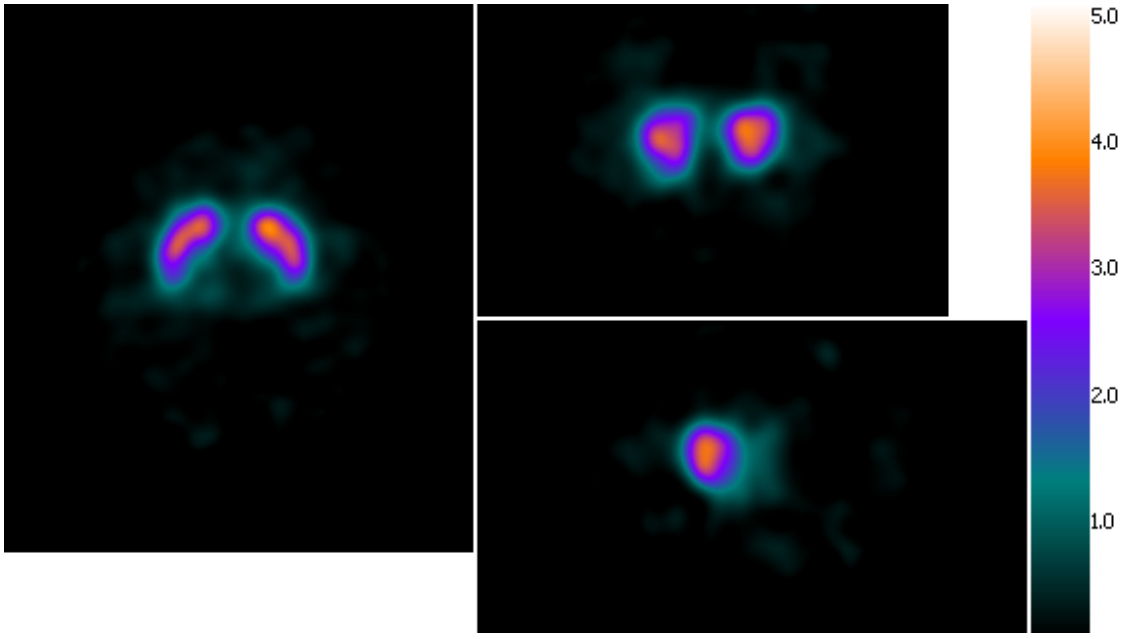

Figure 4

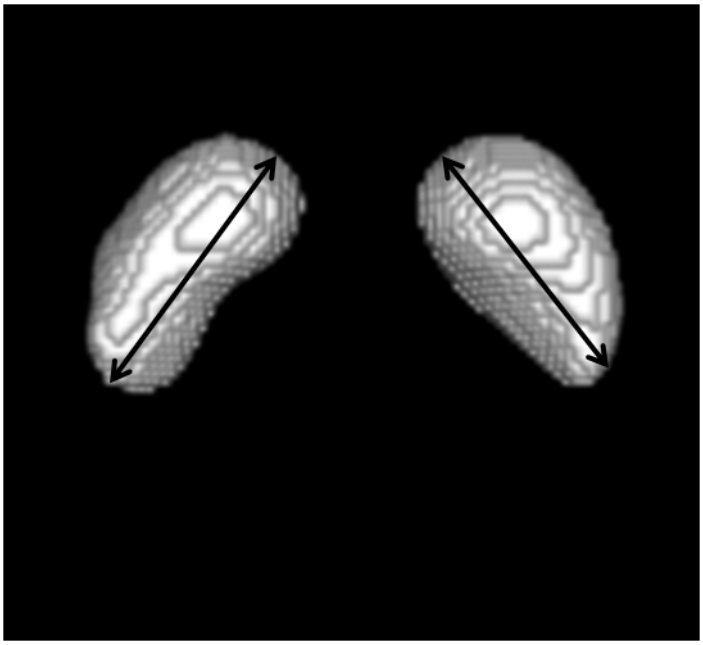

Figure 5

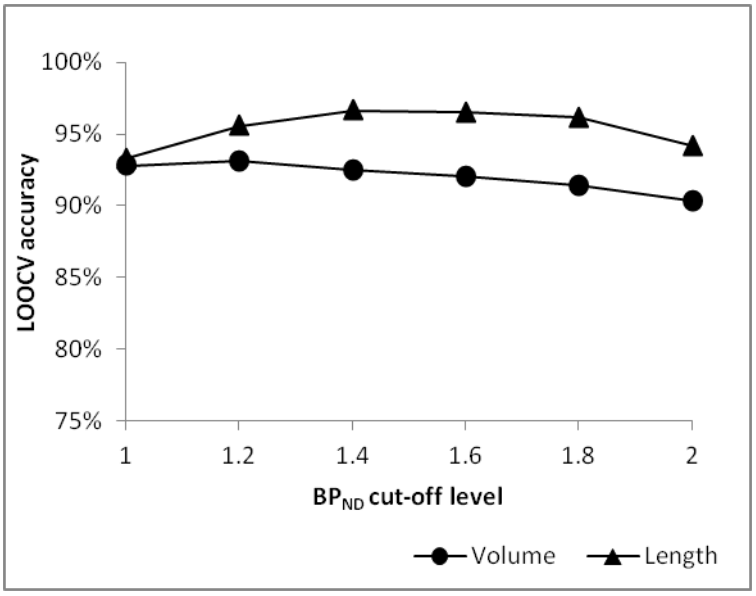

Figure 6

TABLES

Table 1

Feature	Control female	PD female	Control male	PD male
SBR	6.19±1.30	4.00±1.06	5.88±1.23	3.89±1.00
CBP	2.47±0.44	1.62±0.41	2.29±0.31	1.54±0.41
PBP	1.84±0.41	0.87±0.27	1.67±0.37	0.81±0.23
SBP	2.14±0.41	1.22±0.32	1.96±0.37	1.15±0.30
PCR	0.74±0.07	0.53±0.08	0.72±0.08	0.53±0.08
Volume [ml]	11.0±4.0	3.4±2.6	9.5±3.7	2.8±2.4
Length [mm]	36.3±4.9	17.0±7.4	34.9±5.6	15.9±7.5

Table 2

Features used	SVM classifier			KNN classifier			LR classifier		
	Acc.	**Sen.**	**Spe.**	**Acc.**	**Sen.**	**Spe.**	**Acc.**	**Sen.**	**Spe.**
SBR	88.8	92.3	81.3	87.1	91.2	78.5	89.6	93.5	81.3
CBP	88.2	89.8	84.7	86.7	88.3	83.3	87.7	91.2	80.4
PBP	95.4	96.8	92.3	95.6	96.2	94.3	95.1	97.3	90.4
SBP	93.9	95.0	91.4	94.2	95.3	91.9	94.0	95.9	90.0
PCR	92.0	93.7	88.5	91.1	92.3	88.5	91.4	93.9	86.1
Volume	92.0	94.1	87.6	92.0	93.5	89.0	92.5	95.3	86.6
Length	96.5	97.5	94.3	96.0	97.5	92.8	96.6	97.7	94.3
Uptake ratio-based	96.3	97.5	93.8	96.6	97.3	95.2	96.6	97.5	94.7
Dimensional-based	97.5	97.7	97.1	97.2	97.3	97.1	97.5	97.7	97.1
All features	97.9	98.0	97.6	97.2	97.5	96.7	96.9	97.7	95.2

Table 3

Classifiers	Features					
	SBR	**CBP**	**PBP**	**SBP**	**PCR**	**Volume**
SVM	$p < 0.001$, $p_{adj} < 0.001$	$p < 0.001$, $p_{adj} < 0.001$	$p = 0.324$, $p_{adj} = 1.0$	$p = 0.017$, $p_{adj} = 0.347$	$p < 0.001$, $p_{adj} = 0.001$	$p < 0.001$, $p_{adj} = 0.001$
k-NN	$p < 0.001$, $p_{adj} < 0.001$	$p < 0.001$, $p_{adj} < 0.001$	$p = 0.696$, $p_{adj} = 1.0$	$p = 0.118$, $p_{adj} = 1.0$	$p < 0.001$, $p_{adj} = 0.001$	$p = 0.001$, $p_{adj} = 0.015$
LR	$p < 0.001$, $p_{adj} < 0.001$	$p < 0.001$, $p_{adj} < 0.001$	$p = 0.696$, $p_{adj} = 1.0$	$p = 0.118$, $p_{adj} = 1.0$	$p < 0.001$, $p_{adj} = 0.001$	$p = 0.001$, $p_{adj} = 0.015$