OPEN ACCESS
International Journal of Neural Systems, Vol. 33, No. 8 (2023) 2350041 (20 pages)

DOI: 10.1142/S0129065723500417

World Scientific
www.worldscientific.com

Nonlinear Weighting Ensemble Learning Model to Diagnose Parkinson's Disease Using Multimodal Data

D. Castillo-Barnes*,‡, F. J. Martinez-Murcia†, C. Jimenez-Mesa†, J. E. Arco*, D. Salas-Gonzalez†, J. Ramírez† and J. M. Górriz†

*Department of Communications Engineering, University of Malaga
Blvr. Louis Pasteur 35 29004, Malaga, Spain

†Department of Signal Theory, Telematics and Communications
University of Granada
Periodista Daniel Saucedo Aranda S/N 18071, Granada, Spain

‡diegoc@uma.es

Accepted 8 May 2023
Published Online 21 July 2023

Parkinson's Disease (PD) is the second most prevalent neurodegenerative disorder among adults. Although its triggers are still not clear, they may be due to a combination of different types of biomarkers measured through medical imaging, metabolomics, proteomics or genetics, among others. In this context, we have proposed a Computer-Aided Diagnosis (CAD) system that combines structural and functional imaging data from subjects in Parkinson's Progression Markers Initiative dataset by means of an Ensemble Learning methodology trained to identify and penalize input sources with low classification rates and/or high-variability. This proposal improves results published in recent years and provides an accurate solution not only from the point of view of image preprocessing (including a comparison between different intensity preservation techniques), but also in terms of dimensionality reduction methods (Isomap). In addition, we have also introduced a bagging classification schema for scenarios with unbalanced data.

As shown by our results, the CAD proposal is able to detect PD with 96.48% of balanced accuracy, and opens up the possibility of combining any number of input data sources relevant for PD.

Keywords: Ensemble learning; neuroimaging; Parkinson's disease; MRI; SPECT; computer-aided-diagnosis; machine learning; image processing.

1. Introduction

Parkinson's Disease (PD) is a degenerative condition of the brain that causes motor symptoms such as rigidity and bradykinesia and a wide variety of non-motor complications including limitations in speech, lack of emotion expressiveness and cognitive impairment.[1,2] Although PD is the most prevalent movement disorder among adults, there are other movement disorders such as the Multiple System Atrophy (MSA), the Progressive Supranuclear Palsy (PSP) or the Corticobasal Degeneration (CD). Since the symptoms of these pathologies are very similar to those of PD, the accuracy of its diagnosis is sometimes limited especially in early stages.[3–5]

In clinical practice, PD is diagnosed through the use of motor tests (such as the UPDRS) and the

‡Corresponding author.

visual inspection of functional imaging to determine the distribution of dopaminergic transporters in the brain.[6] Though some recent works have claimed that PD can be diagnosed only recurring to structural scans, the most reliable diagnostic tool remains to be the FP-CIT SPECT imaging modality based on the use of the $I^{[123]}$-Ioflupane radioligand.[7]

Nowadays, there is a growing interest on research combining structural and functional scans.[8,9] Nevertheless, these proposals present some challenges. For example, if we want to compare two or more acquisitions, it will be necessary to compare the same anatomical positions between them. To solve this, we have to spatially normalize the brain scans by means of: (1) rigid transformations (including translations, rotations, enlargements and cropping); or (2) the combination of rigid and nonlinear transformations which might alter the morphology of some anatomical regions to get an accurate match between voxels. Though it would be preferably to only use affine transformations, specially in scenarios analyzing the morphological distribution of intensities in the brain,[10] when the equipment used to obtain the brain scans is not the same (different cameras, calibration conditions, etc.) or when we want to compare different subjects voxel-by-voxel (or even the same subject but at different time instants), then we have to rely on the nonlinear approach. But there is an additional issue to be taken into account. After the warping process by a nonlinear registration, scans from patients with PD exhibit larger changes in their shape.[11] This alters the intensity preservation during the spatial normalization and increases (artificially) the interclass separation between control subjects and patients with PD.[12] Therefore, if we want to combine brain images from different modalities, it would be interesting to study also the effect of intensity preservation on each imaging modality and determine which approach might be more convenient when differentiating patients with potential PD.

Once these problems have been solved, it remains to be seen how to perform the data analysis.[13] In the last years, the introduction of automated Computer-Aided Diagnosis (CAD) systems based on Machine Learning (ML) makes it possible to alleviate the lack of trained specialists and allows researchers to find new ways in early detection of dementia.[14,15] Moreover, these kind of tools allow us to identify subtle patterns that help us to subtype better the disease and understand its pathogenesis.[7,16,17] In the current literature about PD, most of the published works refer to the study of brain imaging of an specific modality (mainly FP-CIT SPECT).[18] However, from a scientific point of view, it would be more interesting to combine as much information from different data sources including (among others) conventional medical imaging from both functional and structural modalities, blood markers, genetic testing, metabolomics and proteomics. In fact, it is expected that their combination leads us to find out what the triggers of PD are and how they act.[17,19–21]

Ensemble learning refers to a methodology that attempts to improve predictive performance by combining predictions from multiple single models that are learning from distinct data subsets.[22,23] In neuroimaging, the most popular ensemble learning method for imaging classification is based on Majority Voting (MV).[24–35] Though this approach helps to prevent potential overfitting while reducing the variance of input data, if there are many input sources but they are not all reliable, this kind of solutions tends to underfit.[36] In this sense, there is a growing interest in finding new proposals able to solve some of their limitations.[22] For example, Haque *et al.*[37] proposed the use of a genetic algorithm-based search method to find the optimum combination of classifiers to develop a heterogeneous ensemble. In Liu *et al.,*[38] the addition of fully-connected layers (with sigmoid activation functions) models the contribution of several input data sources. Authors in Rondina *et al.*[39] made use of Multiple Kernel Learning to combine a set of ROIs from MRI, FDG PET and rCBF SPECT for Alzheimer's. The boosting schema described in Nanni *et al.* (2018)[40] learns from a partitioned feature space during the training phase and assigns different classifiers to features subsets in test samples to evaluate their labels.

In this context, the work presented here aims to combine the contributions of several previous works to develop a CAD model for PD assessment using an ensemble learning methodology to evaluate subjects with potential PD. This includes, among others, the classification schema based on a Weighted Majority

Voting (WMV) schema with a nonlinear windowing function that strongly penalizes the contributions of non-reliable input data sources presented in Castillo–Barnes *et al.*[41]; its improvement by penalizing data sources with high variability[36]; the replacement of traditional windowing functions by Merino's curve[36]; the inclusion of a reference method for dimensionality reduction of brain scans using a nonlinear decomposition[42]; and finally, the evaluation of different intensity preservation techniques during the spatial normalization of input brain scans.[11] Indeed, though many other new features have also been included (like the bagging of classifiers to alleviate the strong imbalance of our reference database), our proposal can be easily adapted to other analysis scenarios for the evaluation of psychiatric and neurological disorders.[43–45]

2. Materials & Methods

2.1. *Parkinson's progression markers initiative dataset*

Data used in the preparation of this work were obtained from the Parkinson's Progression Markers Initiative (PPMI) database.[a] For up-to-date information on the study, please visit https://www.ppmi-info.org/.

PPMI is a public-private partnership funded by the Fox Foundation for Parkinson's Research and funding partners (including a consortium of industry players, nonprofit organizations and private individuals) listed at https://www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors. The Institutional Review Boards approved the PPMI program of each participating site and all the participants have given their written informed consent to participate in the program.

For this work, we have used 337 subjects from PPMI dataset including: Healthy Control (HC) subjects and patients with de novo Parkinson's Disease (PD). Demographics about these two groups are summarized in Table 1.

As one of the main objectives of this work is to exploit information from both structural and functional brain imaging to evaluate patients with potential Parkinson's Disease, we have evaluated subjects from PPMI dataset whose structural (MRI-T1) and functional (FP-CIT SPECT) imaging acquisitions have been taken at the same time. Unfortunately, only nine subjects were evaluated for both imaging modalities at the same day. In this context, it seems reasonable to us that the progression of neurodegeneration is negligible if the time difference between acquisitions is small even if it is not the same day. Thus, we have decided to extend this margin to 21 days to obtain a sample size that, although being small, will be enough to perform our classification experiments (34 samples per class).

Table 1. Demographics.

	Male		Female		All	
	#	Age	#	Age	#	Age
HC	25	62.9 ± 11.9	9	60.6 ± 8.2	34	62.3 ± 11.0
PD	196	63.1 ± 9.9	107	61.2 ± 8.3	303	62.4 ± 9.4

Although many other types of markers (genetic, proteomic, etc.) could have been included, due to the lack of data available in the PPMI repository for all the input samples, the methodology developed for this work had to be only evaluated using the aforementioned imaging modalities.

2.2. *Image preprocessing*

In order to compare the same anatomical voxels from MRI-T1 and FP-CIT SPECT imaging modalities, we have to perform the spatial normalization of each input brain scan to a common reference space.[12] The standard procedure for this consists of (1) applying the spatial co-registration of all input FP-CIT SPECT scans to a native space defined by their respective MRI imaging; and (2) if we want to perform voxel-by-voxel analysis, computing the spatial normalization of each MRI scan to a common reference space (template) and applying this deformation to its corresponding FP-CIT SPECT image.

A common tool to perform brain imaging spatial transformations is the *Statistical Parametric Mapping* (SPM) framework (available at: https://www.fil.ion.ucl.ac.uk/spm/). For its latest version, it is recommended that spatial normalization of any functional scan makes use of a segmented version of

[a] https://www.ppmi-info.org/access-data-specimens/download-data.

its respective MRI imaging.[46,47] However, as both structural and functional brain scans from the PPMI dataset are rarely co-registered, this procedure has to be carried out by minimizing the mean squared difference between a reference template and a warped version of each input scan. In addition to this, since many FP-CIT SPECT images in the PPMI repository were cropped or included artifacts, some co-registrations failed as the model did not converge or the resulting image was totally erratic.

As input FP-CIT SPECT scans would be more likely to be similar to each other than to a structural template, we tried to perform this registration to a reference template using the same imaging modality.[12] Thus, following the procedures described by Salas *et al.*,[48] a functional template is generated and then co-registered to the same position as the structural Tissue Probabilistic Map (TPM) included in SPM12 for MRI-T1 scans.[49] Once generated, final co-registrations between both functional and structural brain scans are calculated by means of their spatial registration to our new functional template (in case of FP-CIT SPECT scans) or straightforwardly to the TPM template (MRI).

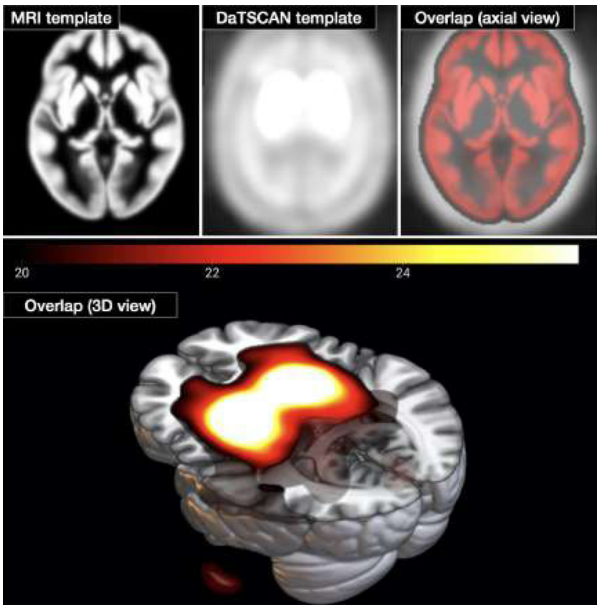

Fig. 1. Top row: axial views of an intermediate slice showing the structural template (left), the functional template (center) and their overlap (right). Bottom row: 3D montage showing the overlap between the structural and functional templates.

MRI-T1 scans from PPMI dataset have been segmented using SPM12. However, since only Grey Matter (GM) and White Matter (WM) tissues might be relevant for PD diagnosis,[8,50–52] any other tissue generated is discarded.[53,54] During this process, SPM12 has to redistribute the intensities of some brain regions to compensate for the deformation needed to fit each sample to our reference space. As explained in SPM12 guidelines,[47] this can be performed in two ways: (1) by weighting the average of the signal under an smoothed kernel (*intensity preservation of the concentration*); or (2) by preserving the total amount of signal for each region so intensities from those areas which are most expanded during the warping are reduced (*intensity preservation of the amount*). Figure 2 shows the differences between these two alternatives when applying some spatial deformations to the intensity map on the left.

Preprocessing of FP-CIT SPECT modality requires some additional steps. First, we had to identify and discard samples with errors including cropped images, very noisy samples, acquisitions with artifacts, etc.[55] At this point, when using SPM12 to perform the spatial registration of functional brain scans, we realized that due to differences

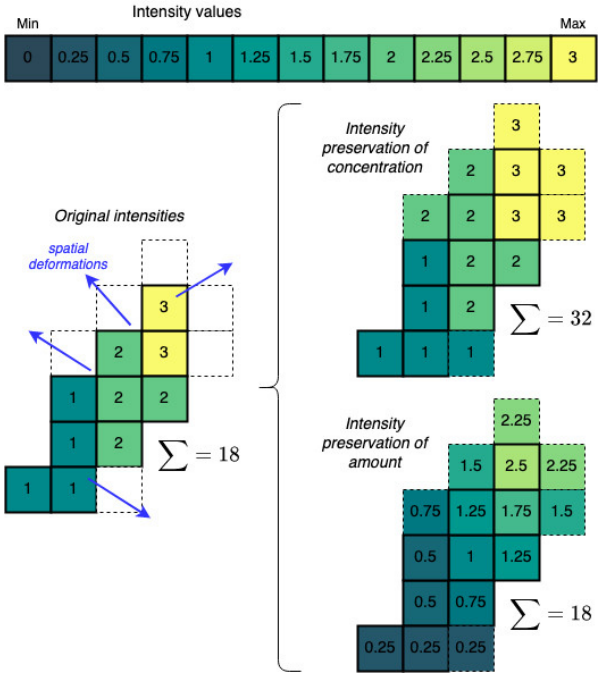

Fig. 2. Example showing the effects of the two intensity preservation methods included in SPM12 when applying some spatial deformations to the intensity map on the left.

in intensity ranges (in some cases by several orders of magnitude), the fit between the samples and our reference template failed many times. To solve this problem, we proposed to standardize all input scans based on the procedure described by expression (1). In this, terms $\text{Template}^{\text{MIN,MAX}}$ refer to the minimum or maximum intensity values of our reference template, whereas $\text{Subject}_i^{\text{MIN,MAX}}$ represent the minimum or maximum intensity values from the ith input sample. The transformation helps us not only to improve the nonlinear spatial normalization and reduce the negative effects due to artifacts and background noise.

$$\Lambda = \frac{\text{Template}^{\text{MAX}} - \text{Template}^{\text{MIN}}}{\text{Subject}_i^{\text{MAX}} - \text{Subject}_i^{\text{MIN}}}.$$

$$\text{Subject}_i = \Lambda(\text{Subject}_i - \text{Subject}_i^{\text{MIN}}) + \text{Template}^{\text{MIN}}. \tag{1}$$

Once applied the spatial normalization, by means of either intensity preservation of the concentration or intensity preservation of the amount, the next step will be to apply an intensity normalization approach. To this end, and following the recommendations given in Brahim *et al.*,[56] where the authors compared several intensity normalization methods during the classification of FP-CIT SPECT images from HC subjects and patients with PD, we decided to include in our pipeline the intensity normalization procedure using alpha-stable distributions.[57,58] This method assumes that the intensity values of the FP-CIT SPECT acquisitions follow α-stable distributions with parameters α (stability), β (skewness), γ (scale) and δ (location). Together, $\{\alpha_i, \beta_i, \gamma_i, \delta_i\}$, these four parameters determine the shape of the intensity distributions of FP-CIT SPECT scans. Thus, if we apply the linear transformation in expression (2), the original intensities from the ith sample, X_i, will be transformed into a new intensity map, Y_i, whose range is comparable to any other transformed sample.

$$Y_i = \frac{\sum_{i=1}^{N} \gamma_i}{\gamma} X_i + \frac{\sum_{i=1}^{N} \delta_i}{\delta} - \frac{\sum_{i=1}^{N} \gamma_i}{\gamma} \mu_i. \tag{2}$$

Figure 3 includes an arrangement showing the intensity distribution of several FP-CIT SPECT scans randomly selected from PPMI dataset before/after their normalization. As we can observe, intensity ranges after normalization are comparable, and it is easier to distinguish between samples with a profuse loss of dopaminergic transporters especially at striatum region.

2.3. *Ensemble learning*

Our ensemble learning proposal is based on a weighting of the predictions obtained by means of simple classifiers focused on particular data sources and/or data subsets.[22]

2.3.1. *Majority voting*

The simplest form of weighted ensemble is Majority Voting (MV) where the final decision on instance x depends only on how many classifiers decided that x belongs to the jth class (c_j). Applying (3), the predicted class of the ensemble for sample x is computed from the labels predicted by the kth classifier,

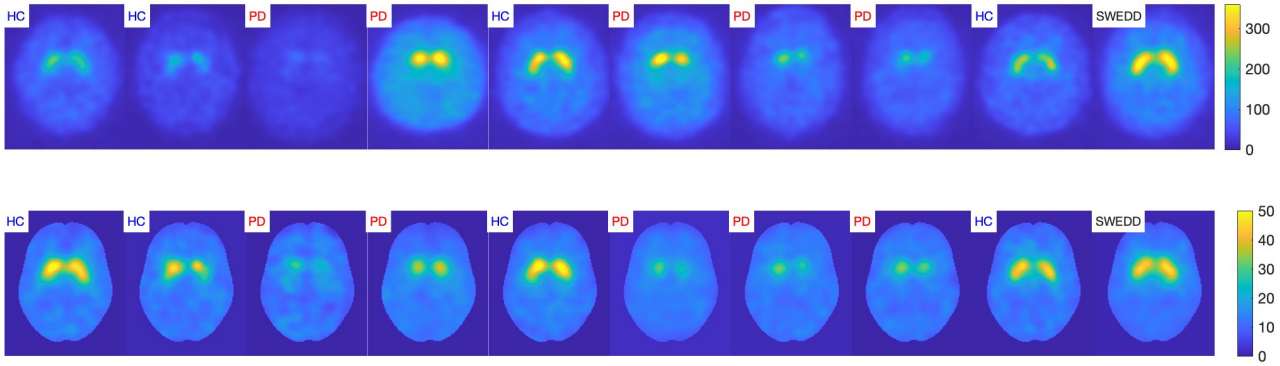

Fig. 3. Comparison of eight randomly selected FP-CIT SPECT brain scans from the PPMI repository before (top row) and after (bottom row) their preprocessing.

$y_k^{\text{pred}}(x)$, restricted to the function $g(y, c)$ in (4).

$$y_{\text{MV}}^{\text{pred}}(x) = \underset{c_j \in \text{domain(y)}}{\text{argmax}} \left(\sum_k g(y_k^{\text{pred}}(x), c_j) \right), \quad (3)$$

$$g(y, c) = \begin{cases} 1, & y = c, \\ 0, & y \neq c0. \end{cases} \quad (4)$$

2.3.2. *Weighted majority voting*

Though MV achieves great classification results while reducing potential overfitting, these solutions tend to underfit when combining a large amount of input data sources that do not generalize well.[41] One way to alleviate this problem is to penalize the contribution of weak classifications through a Weighted Majority Voting (WMV) approach including a factor $w_k \in [0, 1]$ that modulates the contribution of the kth input data source to our final decision.[59] Since this term usually represents the result of a classification metric (e.g. accuracy or balanced accuracy), its value is determined using an internal Cross-Validation (CV) loop fitted to the training data.

$$y_{\text{WMV}}^{\text{pred}}(x) = \underset{c_j \in \text{domain(y)}}{\text{argmax}} \left(\sum_k w_k \cdot g(y_k^{\text{pred}}(x), c_j) \right). \quad (5)$$

2.3.3. *Nonlinear windowing*

Although to a lesser extent, if we try to classify a large amount of input data sources that do not generalize well, then proposals based on WMV will also be affected by the underfit. To mitigate this problem, some proposals have emerged over the last few years. In Castillo–Barnes *et al.*,[41] authors introduced two restrictions. First, setting to zero the contribution (vote) of input sources with rates lower than 50%. Second, since simple learning models trained on data resulting in classification rates slightly above 50% should not have a large impact on the final ensemble decision, any of the functions $\Gamma(w_k) \in [0, 1]$ in (6) modulate the contribution of each input source nonlinearly. The effect of this term on the final ensemble is to drastically reduce WVM dependence on sources with poor but $> 50\%$ classification rates.

$$\Gamma(w_k) = \begin{cases} \text{Quadratic} & 4w_k^2 - 4w_k + 1, \\ \text{Exponential} & e^{0.9624 \cdot w_k} - 1.6180. \end{cases} \quad (6)$$

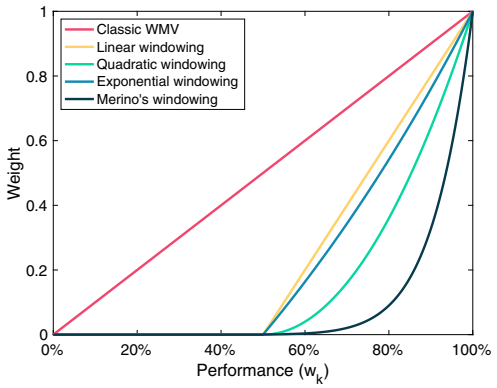

Fig. 4. Windowing functions for the WMV.

Later, the windowing function in (7) (also called *Merino's Curve*) was introduced in Castillo–Barnes *et al.*,[36] As shown in Fig. 4 this proposal is more robust to small variations in final classification.

$$\Gamma(w_k) = \frac{w_k}{0.3015\sqrt{1 + \frac{10}{w_k^{20}}}}. \quad (7)$$

Regardless of the chosen windowing function, output predictions using a WMV with NonLinear Windowing (NLW) model remains as follows in (8):

$$y_{\text{NLW}}^{\text{pred}}(x) = \underset{c_j \in \text{domain(y)}}{\text{argmax}} \left(\sum_k \Gamma(w_k) \cdot g(y_k^{\text{pred}}(x), c_j) \right). \quad (8)$$

2.3.4. *Penalizing the contribution of input data sources with high-variability*

In this work, we have introduced a novel method to penalize the contribution of input data sources with a high-variability. Though this technique is similar to the one presented in Castillo–Barnes *et al.*,[36] where contribution from each source was divided by its variability, here we have opted for the $\Lambda(w_k)$ term in (9) as a way to reduce the impact of sources with low classification rates but reduced variability. In this expression, $\mathbf{W}_k^{\text{Internal}}$ refers to a vector containing all the classification metrics computed within the fold of the kth internal CV loop[b].

$$\Lambda(w_k) = \left| \frac{\log_{10}(\text{std}(\mathbf{W}_k^{\text{Internal}}))}{10} \right|. \quad (9)$$

As mentioned by a reviewer, if the standard deviation of any input tends to zero, $\Lambda(w_k)$ term will tend

[b] $\mathbf{W}_k^{\text{Internal}} = [w_1, w_2, \ldots, w_{10}]$ when using a 10-fold CV schema.

to inf. To avoid that input sources with very small standard deviation results but poor classification rates will not result in a huge impact on final decision, it has been proposed to establish a minimum standard deviation value of $\pm 1\%$ to limit that $\Lambda(w_k)$ term will not become larger.

If we combine the expressions (8) and (9), the ensemble learning methodology proposed in this work will result as follows:

$$y^{\text{pred}}_{\text{NLW}}(x) = \underset{c_j \in \text{domain(y)}}{\text{argmax}} \times \left(\sum_k \Lambda(w_k)\Gamma(w_k) \cdot g(y^{\text{pred}}_k(x), c_j) \right). \tag{10}$$

2.4. *Classification schema*

Figure 5 depicts the overall diagram of our classification model based on the ensemble learning methodology explained above.

Starting with the preprocessing, we have balanced our input dataset to prevent classifiers from overestimating any of the input classes.[60] However, as shown in Table 1, as the number of HC subjects is clearly lower than the number of patients with PD, it would be interesting to repeat our classification experiments several times (bagging of classifications) but subsampling randomly the most populated class. After several repetitions, results obtained can be averaged obtaining a more generalizable model that also depends on many other samples that would have been discarded otherwise.[22,34]

The classification procedure uses as many external CV loops as the number of input data sources. Before attempting to predict the labels of test samples (x) using the kth input data source, it is always recommended to first discard input features with zero-variance and then standardize the remaining ones.

After preprocessing, each brain scan has a size of $121 \times 145 \times 121$ voxels (more than two million of input features). Since this number exceeds the

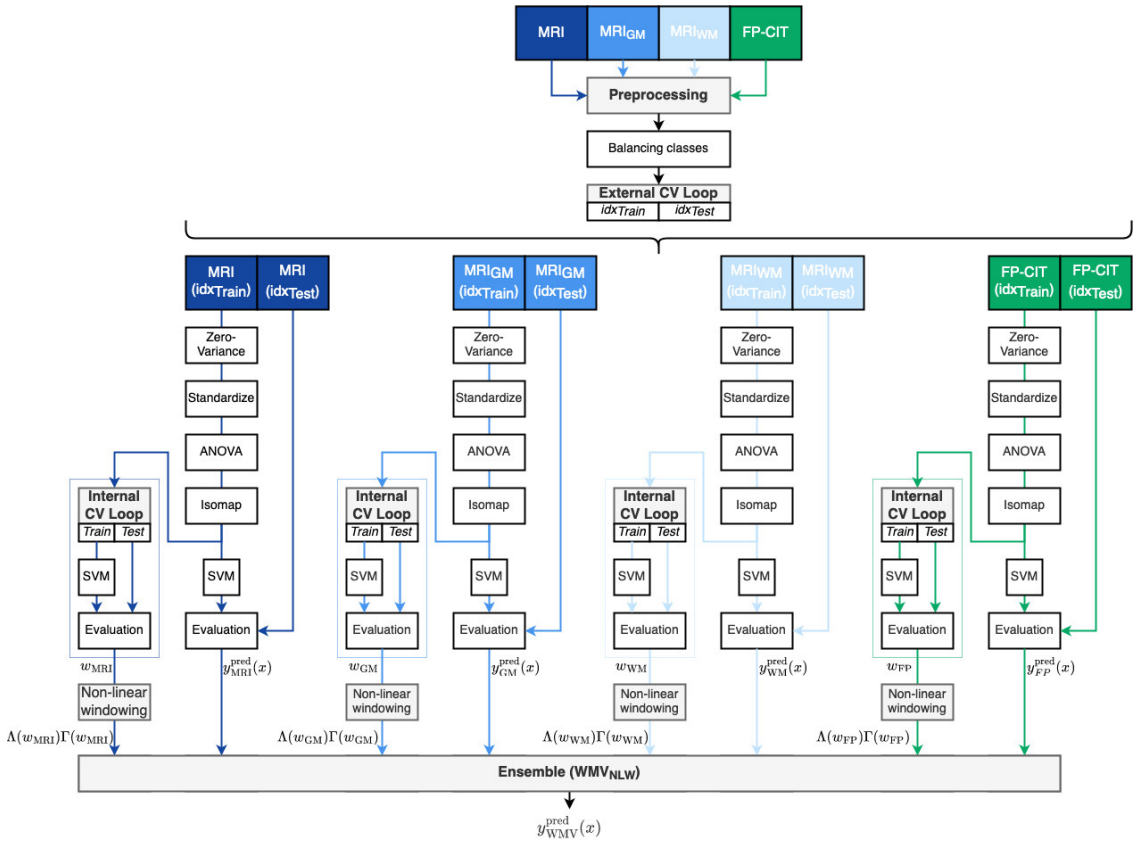

Fig. 5. General schema of the ensemble learning methodology.

number of input samples by six orders of magnitude, it is imperative to implement some kind of feature selection and/or dimensionality reduction method. For that, we have made use of the Analysis of Variance (ANOVA) to select only those features with $p_{\text{value}} \leq 0.001$.[61,62] Although feature selection by ANOVA considerably reduces the number of input features, it may not be sufficient. Therefore, and following the recommendations given in Simon–Rodriguez *et al.*,[42] we have reduced the dimensionality of training subsets by means of a nonlinear decomposition using the Isometric Feature Mapping (Isomap) algorithm.[63] Although Isomap has some disadvantages (such as being sensitive to noise and sample size), in comparison to alternatives like the Principal Component Analysis (PCA), it is capable of discovering nonlinear degrees of freedom that underlie complex patterns.[64] Furthermore, Isomap does not define an explained variance ratio term as PCA. Instead, Tenenbaum *et al.*[63] defined a term called *residual variance* that should not be compared with PCA, so there is no method to estimate how many components (number of coordinates for the manifold) select. To solve this, we propose using the same number of components as would be used in PCA (as many as needed to explain the 90% of the variance) for two reasons: (1) experimentally, either this value (or a slightly lower one) always achieved the same or even better classification results than PCA; and (2) it is easier to compare both proposals. In any case, if the number of components calculated is greater than the number of input samples, its value is limited to our input sample size.

Regarding to the classifiers, for this work, we have made use of Support-Vector-Machines (SVM) with Linear Kernel,[65] As shown in Fig. 5 each input data type makes use of two independent classifiers. On one hand, we have a classifier within each internal cross-validation loop to estimate the weights of each input data type (w_k). And on the other hand, another classifier in the external validation loop to obtain our predicted test data labels, $y_k^{\text{pred}}(x)$. Note that, though more complex algorithms could also be used (including the use of nonlinear Kernels such as the RBF Kernel), since Isomap is able to reduce the complexity of our training data, it was not expected that a further nonlinear decomposition will be necessary. Besides, linear classifiers are usually faster and easier to train.

Table 2. Example showing how our ensemble learning proposal decides whether a sample should belong to Class_0 ($y_k^{\text{pred}}(x) = 0$) or Class_1 ($y_k^{\text{pred}}(x) = 1$).

	Source_1	Source_2	Source_3	Source_4
$y_k^{\text{pred}}(x)$	1	0	0	1
w_k	0.9	0.7	0.8	0.75
$\Lambda(w_k)\Gamma(w_k)$	0.3272	0.0207	0.0900	0.0443
$y_{\text{NLW}}^{\text{pred}}(x) = (10)$	$0.2608 \geq 0 \rightarrow y_{\text{NLW}}^{\text{pred}}(x) = 1$			

Once the classifiers have been trained, it only remains to combine both predictions within the external CV loop and weights obtained from the internal CV loop. For that, we can apply (10) to obtain the final decision of any new (unknown) sample given to our CAD system model and evaluate its robustness.

Table 2 exemplifies how our ensemble learning method works. Suppose we have a test sample that is being evaluated. Once we calculate the final weights after applying the nonlinear windowing, $\Lambda(w_k)\Gamma(w_k)$, a final decision is given as the sum of these weights by means of (10).

2.5. *Experimental design*

To evaluate our ensemble learning methodology we have conducted two types of experiments. First, from a clinical application point of view, we have combined the learning of the four brain imaging modalities mentioned in Sec. 2.2. Then, to confirm that the classification results obtained were not dependent on the input data, we have tested our model using a large synthetic dataset. The following subsections give more details on both experiments.

2.5.1. *Neuroimaging data*

The classification experiments with neuroimaging data have been performed with the following assumptions in mind:

(1) As mentioned before, PPMI dataset has a highly unbalanced distribution of samples between HC and PD classes (only 34 HC subjects versus 303 patients with PD). This presents a challenge for training our model because it may focus too heavily on the PD class and not adequately represent the HC class. To mitigate this

problem, and ensure that our model is not biased towards the larger class, we employed a technique called subsampling.[22] This method involves randomly selecting a subset of samples from the larger class to create a new, balanced dataset with an equal number of samples in each class. Thus, for each of these balanced datasets, we trained a ML model to predict whether a subject belonged to the HC or PD class. Once training is completed, we can combine the results of each model by averaging their predictions to obtain a more generalizable solution that is less dependent on the specific set of samples used for training.

(2) Neuropathologically, PD mainly affects regions responsible for dopamine production.[66,67] Based on this, we have focused the evaluation of our model on the *striatum* area and its surrounding regions including: *accumbens*, *amygdala*, *anterior cingulate cortex*, *cingulate*, *substantia nigra* and *thalamus*. The parcellation of these structures has been performed by means of the Automated Anatomical Labelling atlas v3 (AAL3) presented in.[68] As the AAL3 atlas is spatially in the same position as our MRI template, the edges of each region perfectly match the edges of our input samples. Figure 6 includes a visual representation of the same axial slice obtained from a random subject in our dataset, referred to each imaging modality.

2.5.2. *Synthetic data*

To generate the synthetic dataset, we have made use of the `make_classification` function described in Guyon *et al.*[69] and implemented in Pedregosa *et al.*[70] As explained by the authors, this tool generates a synthetic dataset for a binary or multi-class classification problem using n_{samples} input samples (examples) and n_{features} input features. Each input sample is labeled with a class label that corresponds to one of n_{classes} classes by means of an allocation procedure that assigns to each class one or more normally-distributed clusters of points, resulting is a random n-class classification problem. Along this process, the function creates $n_{\text{informative}}$ informative features (they contribute to the relationship between the input variables and the target) for each class. These informative features are generated by drawing random values from a normal distribution for each class separately, and scaling them by a random matrix. But they are not the only features that we could add to our resulting dataset. For example, we can incorporate to our model $n_{\text{redundant}}$ redundant features for each class. These features are linear combinations of the informative features, so they increase the difficulty of the classification problem in a similar way to what happens with the n_{repeated} duplicate features (i.e. exact copies of some of the informative or redundant features). In addition to the samples and features, the `make_classification` function also allows us to control the separation between the classes. For that, the function introduces a parameter, `CSep`, that specifies the distance between the means of the informative features for each class. A higher value of this parameter leads to a larger separation between the classes, making the classification problem easier. The function can also introduce noise into the dataset by randomly flipping the labels of a fraction of the examples (`flip_y` parameter) and create imbalanced datasets by means of a `weight` parameter that specifies the proportion of each class in the dataset. Table 3 summarizes the arguments given to the function when generating six subsets of synthetic data.

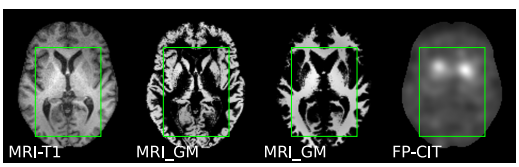

Fig. 6. Axial view of an slice obtained from a random subject in PPMI dataset. Green squares represent the minimum cube that contains the boundaries of the Regions Of Interest (ROIs).

Table 3. Arguments given to `make_classification` function to generate the synthetic datasets.

Parameters	Values
Input samples	$n_{\text{samples}} = 10000$
Balanced classes	`weight` $= (0.1, 0.9)$
Input features	$n_{\text{features (total)}} = 1000$
	$n_{\text{redundant}} = 500$
	$n_{\text{informative}} = 100$
	$n_{\text{repeated}} = 100$
Class separation	`CSep` $= [0.01, 0.05, 0.1, 0.5, 0.9, 2.5]$

3. Results

As mentioned before, the ensemble learning methodology proposed in this work has been evaluated by means of two experiments. First, a bagging of comparisons between HC subjects and patients with PD in PPMI dataset. Then, by combining the learning from several synthetic data, subsets generated *ad-hoc*.

3.1. *PPMI dataset*

For the neuroimaging data, we have randomly subsampled 25 times the samples from the PD class. This gave us as many balanced subsets (34 samples from patients with PD versus 34 samples from HC subjects) to be classified (bagging of classifications described in Sec. 2.5.1). Thus, and following the general diagram depicted in Fig. 5, we have evaluated the performance of our ensemble learning methodology taking into account the two methods for intensity preservation explained in Sec. 2.2.

3.1.1. *Classification results in the external CV loop*

Figure 7 is referred to the intensity preservation of the amount approach and it shows the classification results[c] obtained for the original MRI-T1 modality (67.00%); its segmented versions of GM (77.85%) and WM (66.53%); and the functional FP-CIT SPECT (94.31%).

Similarly, Fig. 8 depicts the performance obtained for the intensity preservation of the concentration. In this case, we had: original MRI-T1 (66.95%); GM (77.97%); WM (67.78%); and FP-CIT SPECT (96.48%).

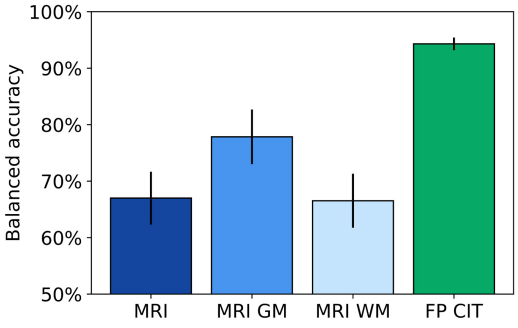

Fig. 7. Classification results obtained when comparing the HC and PD classes for the intensity preservation of the amount approach.

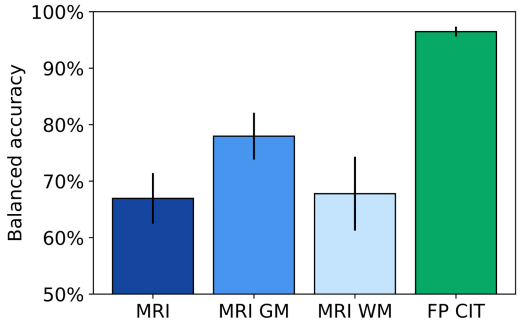

Fig. 8. Classification results obtained when comparing the HC and PD classes for the intensity preservation of the concentration approach.

3.1.2. *ROC curves*

To illustrate the trade-off between sensitivity and specificity parameters for every possible cut-off using our tests set, we have included in Figs. 9 and 10 the resulting Receiver Operating Characteristic (ROC) curves referred to each imaging modality.[71] In both cases, these curves were calculated by averaging the results obtained from all the classifications in our bagging. Apart from the curves representation, we also included the values of Area Under the Curve (AUC) parameters, which help us to quantify how

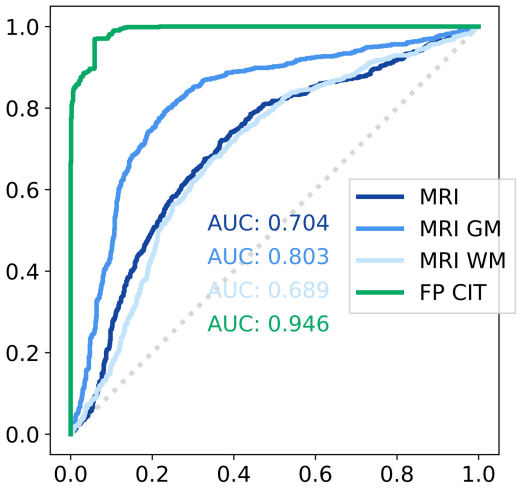

Fig. 9. ROC curves obtained for the intensity preservation of the amount.

[c]In terms of the balanced accuracy classification parameter.

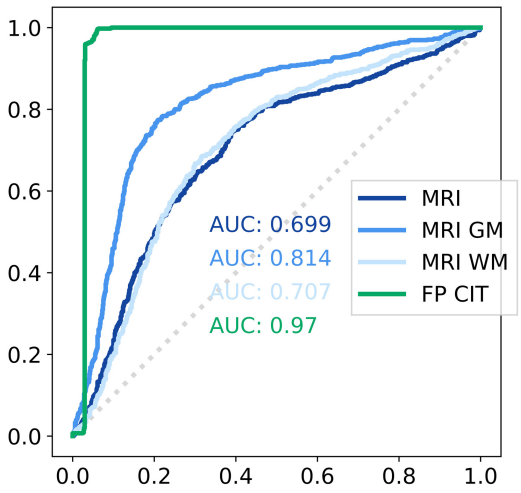

Fig. 10. ROC curves obtained for the intensity preservation of the concentration.

well our classifiers are able to distinguish between input classes.

3.1.3. *Classification results using our ensemble learning approach*

The ensemble learning methodology proposed in this work has been compared with other traditional methods described in Sec. 2.3. This also includes the MV and WMV approaches. In Figs. 11 and 12, we depicted the classification results (in terms of balanced accuracy) obtained using the three ensemble learning methodologies (MV, WMV and WMV-NLW) depending on the intensity preservation approach. In addition to these representations, we have

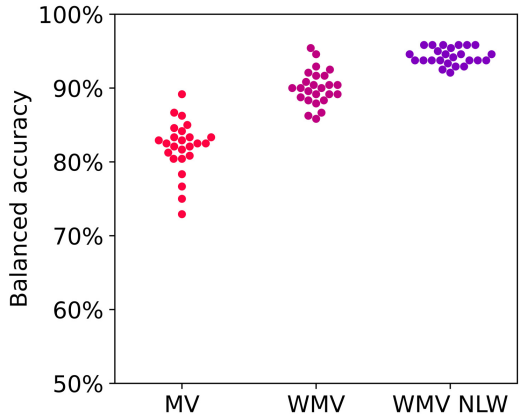

Fig. 11. Classification results for the MV, WMV and WMV-NLW proposals when adopting the intensity preservation of the amount strategy.

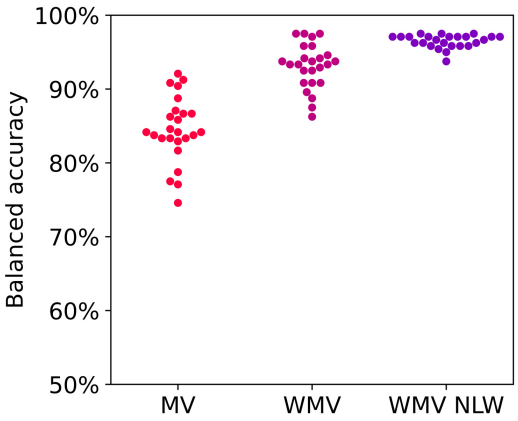

Fig. 12. Classification results for the MV, WMV and WMV-NLW proposals when adopting the intensity preservation of the concentration strategy.

also included in Table 4 more details about other classification parameters obtained when classifying the acquisitions from the PPMI dataset.

3.2. *Synthetic dataset*

Regarding to the synthetic data, we performed the same classification experiments as in neuroimaging.

3.2.1. *Classification results in the external CV loop*

Figure 13 shows the classification results (in terms of the balanced accuracy parameter) obtained within the external CV loop.

In addition to this, and trying to emulate the behavior of our classifications with neuroimaging data, we have also classified our synthetic data by means of several subsets with balanced classes. Therefore, we have depicted in Fig. 14 the distributions of our classification results using the synthetic data sources.

3.2.2. *ROC curves*

ROC curves associated to classifications in Fig. 13 were included in Fig. 15. As before, we have also annotated the results of AUC parameters for each input data source when joining all classifications from the bagging.

3.2.3. *Classification results using our ensemble learning approach*

Similar to when we were analyzing the neuroimaging data, we can find in Fig. 16 the results of

Table 4. Details about our classification results using different ensemble learning approaches when comparing HC and PD classes.

		Accuracy	Sensitivity	Specificity	Bal. acc.
Intensity preservation of amount	**MV**	81.78% ± 3.60%	71.60% ± 4.88%	92.46% ± 3.47%	82.03% ± 3.46%
	WMV	89.84% ± 2.35%	89.70% ± 3.09%	90.46% ± 3.26%	90.08% ± 2.30%
	WMV NLW	94.17% ± 1.12%	94.20% ± 0.43%	94.43% ± 2.09%	94.31% ± 1.11%
Intensity preservation of concent.	**MV**	84.40% ± 4.30%	74.16% ± 6.90%	94.86% ± 3.86%	84.51% ± 4.32%
	WMV	93.07% ± 2.87%	92.60% ± 3.90%	93.63% ± 3.88%	93.11% ± 2.97%
	WMV NLW	96.34% ± 0.90%	97.20% ± 0.57%	95.76% ± 1.82%	96.48% ± 0.87%

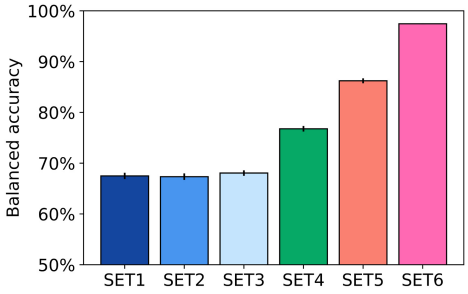

Fig. 13. Classification results obtained for the synthetic dataset.

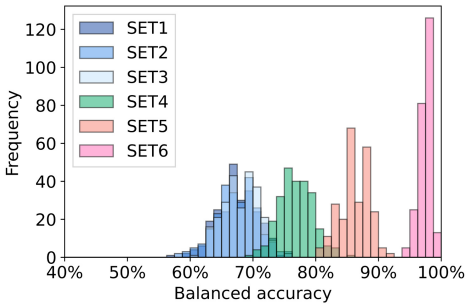

Fig. 14. Histograms of classifications obtained using the synthetic dataset (results derived from the external CV loop).

classifications by the ensemble learning approach. It is also worth mentioning here that since this representation only includes information about the balanced accuracy, more details about other classification parameters (Accuracy, Sensitivity and Specificity) were also included in Table 5.

4. Discussion

4.1. *Neuroimaging data*

MRI-T1 and FP-CIT SPECT brain scans provide us comprehensive and accurate information about structural and functional patterns in the brain associated with PD progression. For example, MRI-T1 images can be used to visualize subtle changes in the volume or shape of certain brain structures, such as the substantia nigra, that are associated with Parkinson's Disease. It is widely known that FP-CIT SPECT imaging can give us much information about regional cerebral blood flow and metabolic activity. In this context, the development of automatic CAD systems based on ML, and focused on the analysis of these two modalities, can be a valuable for clinicians and radiologists as these tools can assist in the

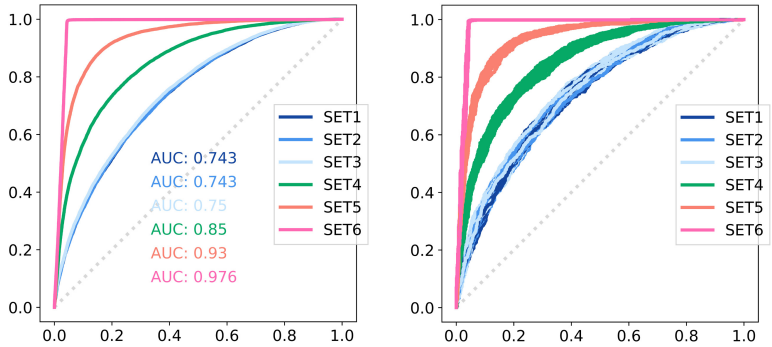

Fig. 15. Averaged ROC curves (left) and individual ROC curves (right) for the synthetic dataset.

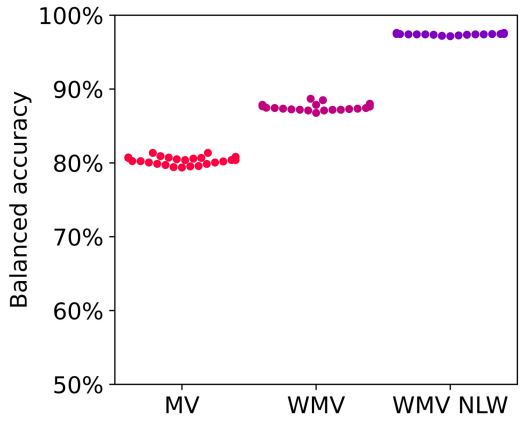

Fig. 16. Classification results for the MV, WMV and WMV-NLW proposals when classifying synthetic data.

interpretation of complex imaging data, provide quantitative measures of structural and functional changes in the brain, and improve the overall accuracy and efficiency of the diagnostic process.[72] This can ultimately lead to earlier and more accurate diagnosis, which can in turn lead to earlier interventions and improved outcomes for patients.[7,66]

Ensemble learning-based methods involve combining multiple ML models to improve overall system performance.[22] For this work, this means combining the classifications from classifiers trained with MRI-T1 and FP-CIT SPECT scans while reducing the risk of overfitting and increasing the generalizability of the system. But to reach this solution, we will first have to take into account some previous steps that allow us to take full advantage of our system's performance.

The first point is related to the spatial registration of the MRI-T1 and FP-CIT SPECT scans in PPMI dataset. As our model performs voxel-wise comparisons, an accurate match between all our input samples results mandatory.[12] In our context, this translates as the spatial registration of all brain scans should be made to a common reference space like the MNI. During this process, selecting an appropriate intensity preservation approach can help us to reduce the impact of samples with worse intensity distributions (generally from patients with PD) in brain areas related to PD.[11] As shown by our classification results in Figs. 7 and 8, the intensity preservation of the concentration approach is more suitable when analyzing voxel-by-voxel intensities as it outperforms (on average) the alternative intensity preservation method by 2.17% in terms of the balanced accuracy parameter.

Although we could have considered other methods for the intensity normalization of functional scans, some previous works such as Brahim *et al.* have suggested that the approach using α-stable distributions is most appropriate when trying to compare controls and patients with PD.[56–58] Consequently, we will not go into further detail on the choice of this method or its comparison with other alternatives.

After discussing all the results that are relevant to the preprocessing of the input data, next step is to address the impact of Isomap for dimensionality reduction to our final classification rates. In comparison to other alternatives such as PCA, Isomap got worse classification results[d] in experiments involving the segmented versions of MRI-T1 scans: −2.25% (GM) and −1.74% (WM). On the contrary, when referring to the other two imaging modalities, we obtained an averaged improvement of: +0.95% (MRI-T1) and +0.94% (FP-CIT), which is quite similar to the results published in Simon-Rodriguez *et al.*[42]

We can also perform a similar analysis to evaluate the improvement that could be achieved by using a

Table 5. Details about our classification results obtained by different ensemble learning approaches when classifying the synthetic data sources.

		Accuracy	Sensitivity	Specificity	Bal. acc.
Synthetic	MV	80.30% ± 0.53%	70.71% ± 0.93%	89.90% ± 0.52%	80.30% ± 0.53%
	WMV	87.57% ± 0.43%	85.56% ± 0.70%	89.59% ± 0.61%	87.57% ± 0.42%
	WMV NLW	97.44% ± 0.10%	95.41% ± 0.14%	99.48% ± 0.16%	97.44% ± 0.11%

[d]Given in terms of the balanced accuracy classification parameter.

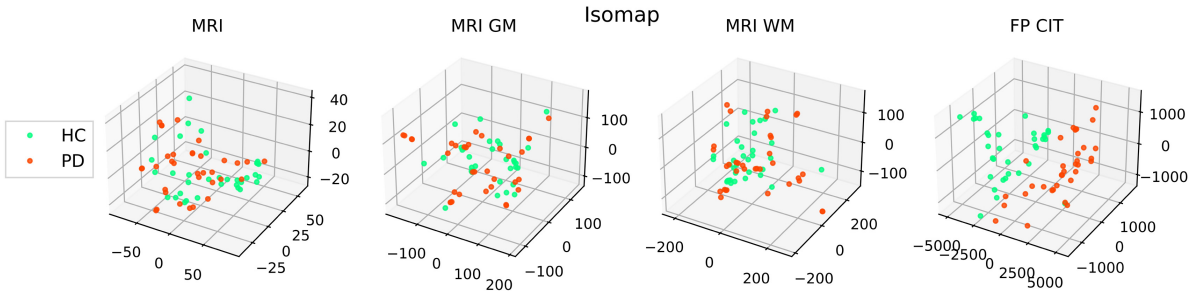

Fig. 17. Example showing the separation between PD and HC classes using three components for Isomap.

nonlinear classifier such as SVM with RBF (Radial Basis Function) Kernel.[73] In this case, the averaged improvement of using Isomap and SVM with Linear Kernel was[e]: −4.09% (GM); −3.44% (WM); −3.05% (MRI-T1) and +1.27% (FP-CIT). These results seem to be quite interesting since they highlight that nonlinear classifiers such as SVM with RBF Kernel get a better fit to structural changes in the volume or shape of some brain areas, whereas SMV classifiers with Linear Kernel are more likely to model changes related to cerebral blood flow and/or metabolic activity.

To complete this part of our analysis, we have included in Fig. 17 the projections of Isomap only considering three-components for the manifold referred to a random classification using SVM with Linear Kernel and Isomap. As it can be seen, the class separation between the HC and PD groups was unclear for any structural imaging modality. In contrast, when we examine the FP-CIT SPECT modality, we observe that PD and HC classes are completely separated even despite the low dimension.

Regardless of the chosen method for dimensionality reduction or classification, it can be stated that individual classifications handling functional data far outperform the results obtained by the structural modality in any of its forms (original RM-T1 image or its segmented versions of GM and WM). This is consistent with the existing literature.[8,18,74,75] In fact, only a few models such as the one published in Solana–Lavalle *et al.*[50] reported classification rates of 96% when applying Voxel-Based-Morphometry in scenarios with small sample sizes. In our case, none of our experiments involving structural data resulted in classification rates greater than the 77.97% ± 4.15% when classifying GM tissue referred to the intensity preservation of the concentration approach using Isomap and SVM with linear Kernel, or the 82.06 ± 3.97% for SVM with RBF Kernel. This also makes clinical sense, as brain atrophy due to PD progression typically affects the region of the nigrostriatal pathway where the largest concentration of dopaminergic neurons are located.[76,77] Although it would have also been interesting to compare the results that could have been obtained with MRI-T3 structural imaging, it is worth mentioning the work by Chakraborty *et al.*[78] where the authors classified samples of patients with PD with an accuracy of 95 .3% when using Convolutional Neural Networks (CNN).

Regarding to our classification results when analyzing the FP-CIT SPECT modality, it is worth mentioning that in most cases, they improve or equal those achieved in many other previous works.[72,79] This also makes sense since despite the limitations of our database, Isomap is able to explain the underlying nature of the input data which facilitates the work of SVM classifiers as explained in Simon-Rodriguez *et al.*[42]

In relation to the ROC curves depicted in Figs. 9 and 10, with a total AUC of 0.97 for the intensity preservation of the concentration when evaluating the FP-CIT SPECT scans, these representations confirm the findings in Castillo–Barnes *et al.*[11] where it was stated that due to the degree of deformation after the spatial normalization was greater in case of patients with PD, the intensity preservation of the concentration tended to an artificially increasing in separability between HC and PD classes, which ultimately leads to better classification results.

Returning to our classification scheme in Fig. 5, the main objective of this work was to develop a

[e]In comparison to our results shown in Fig. 8.

nonlinear WMV ensemble learning model that can combine any number of input data sources and penalize the contribution of sources with poor significance and/or high variability. Owing to results in Figs. 11 and 12 (whose parameters are summarized in Table 4), we can observe that our proposal outperforms the results (in terms of balanced accuracy) achieved by other ensemble learning schemas such as the traditional MV by 11.97% and the WMV by 3.37% for the intensity preservation of the concentration[f] while reducing the variability of our classifications by one-third in the case of intensity preservation of the amount, and one-fifth in the case of the intensity preservation of the concentration. Although they look really promising, these results are too identical to those obtained for the FP-CIT SPECT modality. However, this may be explained by the low contribution of structural imaging, which hardly influences the overall classification.[41,80]

4.2. *Synthetic data*

The ensemble learning methodology proposed for this work has also been tested on a synthetic dataset for two main reasons. First, because synthetic datasets provide a controlled environment for testing the performance of the classification model. This allows for a more comprehensive evaluation of the model's capabilities, as it can be tested under various conditions and scenarios that may not be feasible or practical to test with real clinical data (e.g. the background noise, inaccuracies in the measurements captured by the medical imaging, outliers, missing data, etc). Second, because they can help us to identify potential biases and limitations in our model's performance. By comparing the results obtained with synthetic datasets to those obtained with clinical data, it is possible to determine if our model is biased towards certain features or datasets, which could have implications for the generalizability and validity of the model. With this in mind, we move on to discuss the results obtained when evaluating the synthetic data sets.

As stated for the neuroimaging data, our model enhances the contribution of reliable input sources by modifying nonlinearly the traditional schema of a WMV approach regardless of the input data nature (imaging, genetic tests, blood markers, etc). Therefore, it is expected that the evaluation of any synthetic dataset (as the one generated as explained in Sec. 2.5.2) will lead us to the same conclusions that we have already pointed out.

Considering the classification rates shown in Fig. 13, one would expect that their combination would be very similar to those obtained when only evaluating the subset with the highest `CSep` (class separation) parameter. Indeed, similar to what happened with functional neuroimaging, since this subset is less affected by the nonlinear weighting of our ensemble, it becomes the predominant when trying to predict the label of any new (unknown) input data. As shown in Table 5, if our approach were based on a simple MV schema, we would obtained classification rates (measured by the balanced accuracy parameter) of 80.3%. However, by weighting the input data sources based on their individual classifications (WMV), this margin is increased up to 87.57% or, when also including the nonlinear windowing technique with variability penalization, to 97.44%, which supposes a total increase of 17.14%.

Though a further analysis might also be included, we will focus only on few details:

- As shown in Fig. 16, the variability of MV and WMV methods is quite longer than the one achieved by the WMV-NLW. This was expected because of $\Lambda(w_k)$ term in (9) that enhances the contribution of weights on input sources with higher classification rates.[22]
- If we had only penalized for the effect of variability by dividing the averaged classification rates by their standard deviation, data sources with poor classification rates but low variance could have had a greater impact on the final decision. To avoid this, it was decided to model this contribution by using the logarithm of the weight and setting up a minimum standard deviation limit of $\pm 1\%$ when calculating the $\Lambda(w_k)$ term in (9). In this regard, it would have been interesting to compare the effect of this term on final classification rates of many other related works.[33,59,81,82]

[f]When referred to intensity preservation of the amount, this increase was of 12.28% for MV and 4.23% for WMV.

4.3. *Limitations and future lines*

One of the most important limitations of this work is the small sample size of the classification experiments using imaging data.[83] Since the PPMI dataset does not include many HC subjects who have had both functional and structural imaging testing performed at the same time (or at least, in the appropriate time range as explained in Sec. 2.1), it was necessary to carry out multiple classification experiments to obtain a more generalizable solution. In this sense, it would be desirable that PPMI initiative provide us more data from HC subjects to re-evaluate our results. This means adding other types of biomarkers such as genetics, proteomics, metabolomics or even electroencephalography –among others–;[41,84–90] but also new imaging modalities like the MRI-T3.[78] In our view, this would not only help us to create a more accurate model of PD, but also allow us to better justify the use of ensemble learning-based solutions.

5. Conclusions

Accurate diagnosis of PD is essential to ensure that patients receive the best possible care and treatment. However, the growing number of patients with PD and the lack of well-trained specialists is highlighting the need to develop alternative methods for PD identification that can reduce the time to diagnosis.

In this context, we have developed a weighted ensemble learning methodology for PD diagnosis that has several advantages over any other model published to date, for several reasons:

(1) Because it is able to penalize in a nonlinear way the contribution of those input data sources with low significance and/or high variability.
(2) It comprises a detailed analysis about preprocessing of the two most commonly used imaging modalities for Parkinson's assessment (functional FP-CIT SPECT and structural MRI-T1 imaging). This also includes the evaluation of different intensity preservation techniques, the spatial registration of all imaging modalities to the in MNI space, and the segmentation of MRI-T1 scans so that we could quantify the influence that WM and GM may have on final decision of our CAD model.
(3) Its pipeline includes steps such ANOVA and Isomap that reduce potential overfitting.
(4) Because this model has also been evaluated using several synthetic datasets, which help to justify the classification results and allow us to compare our proposal with any other existing model.

This proposal has been evaluated using the data available in the PPMI repository. However, due to the lack of input samples with both functional and structural images acquired within a short period of time (21 days) to overcome the progression of the neurodegenerative process, it was necessary to implement a bagging of classifications to make our model generalize as much as possible. Despite this, classification results obtained demonstrate that our CAD model is capable of identifying PD with a balanced accuracy of 96.48% and opens the door to incorporating many other sources of information that could help us to achieve even more accurate diagnosis.

Acknowledgments

This work was supported by the FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades/ Proyecto (B-TIC-586-UGR20); the MCIN/AEI/ 10.13039/501100011033/ and FEDER "Una manera de hacer Europa" under the RTI2018-098913-B100 project, by the Consejería de Economía, Innovación, Ciencia y Empleo (Junta de Andalucía) and FEDER under CV20-45250, A-TIC-080-UGR18 and P20-00525 projects. Grant by F.J.M.M. RYC2021-030875-I funded by MCIN/AEI/10.13039/ 501100011033 and European Union NextGenerationEU/PRTR. Work by D.C.B. is supported by the MCIN/AEI/FJC2021-048082-I 'Juan de la Cierva Formacion'. Work by J.E.A. is supported by NextGenerationEU Fund through a Margarita Salas Grant, and work by C.J.M. is supported by Ministerio de Universidades under the FPU18/04902 grant.

References

1. P. Gonzalez-Latapi, E. Bayram, I. Litvan and C. Marras, Cognitive impairment in parkinson's disease: Epidemiology, clinical profile, protective and risk factors, *Behavioral Sci.* **11** (2021) 74.

2. O. Darbin and E. B. Montgomery, Challenges for future theories of parkinson pathophysiology, *Neurosci. Res.* **177** (2022) 1–7.
3. G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana and G. Logroscino, Accuracy of clinical diagnosis of parkinson disease, *Neurology* **86** (2016) 566–576.
4. E. Hustad, A. H. Skogholt, K. Hveem and J. O. Aasly, The accuracy of the clinical diagnosis of parkinson disease. the HUNT study, *J. Neurol.* **265** (2018) 2120–2124.
5. N. J. Herzog and G. D. Magoulas, Convolutional neural networks-based framework for early identification of dementia using MRI of brain asymmetry, *Int. J. Neural Syst.* **32** (2022) 2250053.
6. L. J. Evers, J. H. Krijthe, M. J. Meinders, B. R. Bloem and T. M. Heskes, Measuring parkinson's disease over time: The real-world within-subject reliability of the MDS-UPDRS, *Movement Disorders* **34** (2019) 1480–1487.
7. E. Tolosa, A. Garrido, S. W. Scholz and W. Poewe, Challenges in the diagnosis of parkinson's disease, *Lancet Neurol.* **20** (2021) 385–397.
8. R. Martins, F. Oliveira, F. Moreira, A. P. Moreira, A. Abrunhosa, C. Januário and M. Castelo-Branco, Automatic classification of idiopathic parkinson's disease and atypical parkinsonian syndromes combining [11c]raclopride PET uptake and MRI grey matter morphometry, *J. Neural Eng.* **18** (2021) 046037.
9. A. Ortiz, J. E. Arco, M. A. Formoso, N. J. Gallego-Molina, I. Rodríguez-Rodríguez, J. Martínez-Murcia, J. M. Górriz and J. Ramírez, Towards mixed mode biomarkers: Combining structural and functional information by deep learning, *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications* (Springer International Publishing, 2022), pp. 95–103.
10. D. Castillo-Barnes, F. J. Martinez-Murcia, A. Ortiz, D. Salas-Gonzalez, J. RamÍrez and J. M. Górriz, Morphological characterization of functional brain imaging by isosurface analysis in parkinson's disease, *Int. J. Neural Syst.* **30** (2020) 2050044.
11. D. Castillo-Barnes, J. E. Arco, C. Jimenez-Mesa, J. Ramirez, J. M. Górriz and D. Salas-Gonzalez, Evaluating intensity concentrations during the spatial normalization of functional images for parkinson's disease, *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications*, eds. J. M. Ferrández Vicente, J. R. Œlvarez-Sánchez, F. de la Paz López and H. Adeli (Springer International Publishing, 2022), pp. 176–186.
12. D. Castillo-Barnes, C. Jimenez-Mesa, F. J. Martinez-Murcia, D. Salas-Gonzalez, J. Ramírez and J. M. Górriz, Quantifying differences between affine and nonlinear spatial normalization of FP-CIT SPECT images, *Int. J. Neural Syst.* **32** (2022) 2250019.
13. M. Liu, F. Zhao, X. Jiang, H. Zhang and H. Zhou, Parallel binary image cryptosystem via spiking neural networks variants, *Int. J. Neural Syst.* **32** (2021) 2150014.
14. G. Mirzaei, A. Adeli and H. Adeli, Imaging and machine learning techniques for diagnosis of alzheimer's disease, *Rev. Neurosci.* **27** (2016) 857–870.
15. G. Mirzaei and H. Adeli, Machine learning techniques for diagnosis of alzheimer disease, mild cognitive disorder, and other types of dementia, *Biomed. Signal Process. Control* **72** (2022) 103293.
16. S. Aram, D. Kornev, R. Sadeghian, S. E. Sardari, S. K. Venu, H. Dashtestani and A. Gandjbakhche, Machine learning approaches and neuroimaging in cognitive functions of the human brain: A review, in *Advances Neuroergonomics and Cognitive Engineering* (Springer International Publishing, 2020), pp. 23–29.
17. M. A. Myszczynska, P. N. Ojamies, A. M. B. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook and L. Ferraiuolo, Applications of machine learning to diagnosis and treatment of neurodegenerative diseases, *Nature Rev. Neurol.* **16** (2020) 440–456.
18. J. Mei, C. Desrosiers and J. Frasnelli, Machine learning for the diagnosis of parkinson's disease: A review of literature, *Front. Aging Neurosci.* **13** (2021) 633752.
19. S. Bhat, U. R. Acharya, Y. Hagiwara, N. Dadmehr and H. Adeli, Parkinson's disease: Cause factors, measurable indicators, and early diagnosis, *Comput. Biol. Med.* **102** (2018) 234–241.
20. C. Blauwendraat, M. A. Nalls and A. B. Singleton, The genetic architecture of parkinson's disease, *Lancet Neurol.* **19** (2020) 170–178.
21. M. B. Makarious, H. L. Leonard, D. Vitale, H. Iwaki, L. Sargent, A. Dadu, I. Violich, E. Hutchins, D. Saffo, S. Bandres-Ciga, J. J. Kim, Y. Song, M. Maleknia, M. Bookman, W. Nojopranoto, R. H. Campbell, S. H. Hashemi, J. A. Botia, J. F. Carter, D. W. Craig, K. V. Keuren-Jensen, H. R. Morris, J. A. Hardy, C. Blauwendraat, A. B. Singleton, F. Faghri and M. A. Nalls, Multi-modality machine learning predicting Parkinson's disease, *NPJ Parkinson's Disease* **8** (2022) 35.
22. L. Rokach, *Ensemble Learning* (World Scientific Publishing, 2019).
23. D. Avola, M. Cascio, L. Cinque, A. Fagioli and G. L. Foresti, LieToMe: An ensemble approach for deception detection from facial cues, *Int. J. Neural Syst.* **31** (2020) 2050068.
24. A. Savio and M. Graña, An ensemble of classifiers guided by the AAL brain atlas for alzheimer's disease detection, *Advances in Computational Intelligence*, (Springer, Berlin, Heidelberg, 2013), pp. 107–114.
25. A. Lebedev, E. Westman, G. V. Westen, M. Kramberger, A. Lundervold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsolaki, B. Vellas, S. Lovestone and A. Simmons, Random forest

ensembles for detection and prediction of alzheimer's disease with a good between-cohort robustness, *NeuroImage: Clinical* **6** (2014) 115–125.
26. A. Ortiz, J. Munilla, J. M. Górriz and J. Ramírez, Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease, *Int. J. Neural Syst.* **26** (2016) 1650025.
27. W. Wu, J. Venugopalan and M. D. Wang, 11c-PIB PET image analysis for alzheimer's diagnosis using weighted voting ensembles, in *2017 39th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2017), pp. 3914–3917.
28. L. Sørensen and M. Nielsen, Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination, *J. Neurosci. Methods* **302** (2018) 66–74.
29. Y. Chen, H. Jia, Z. Huang and Y. Xia, Early identification of alzheimer's disease using an ensemble of 3d convolutional neural networks and magnetic resonance imaging, in *Advances in Brain Inspired Cognitive Systems* (Springer International Publishing, 2018), pp. 303–311.
30. H. Lei, Z. Huang, F. Zhou, A. Elazab, E.-L. Tan, H. Li, J. Qin and B. Lei, Parkinson's disease diagnosis via joint learning from multiple modalities and relations, *IEEE J. Biomed. Health Inf.* **23** (2019) 1437–1449.
31. S. Ahmed, B. C. Kim, K. H. Lee and H. Y. J. and, Ensemble of ROI-based convolutional neural network classifiers for staging the alzheimer disease spectrum from magnetic resonance imaging, *PLOS ONE* **15** (2020) e0242712.
32. S. A. Knox, T. Chen, P. Su and G. Antoniou, A parallel machine learning framework for detecting alzheimer's disease, *Brain Informatics* (Springer International Publishing, 2021), pp. 423–432.
33. J. E. Arco, J. Ramírez, J. M. Górriz and M. Ruz, Data fusion based on searchlight analysis for the prediction of alzheimer's disease, *Expert Syst. Appl.* **185** (2021) 115549.
34. R. Logan, B. G. Williams, M. F. da Silva, A. Indani, N. Schcolnicov, A. Ganguly and S. J. Miller, Deep convolutional neural networks with ensemble learning and generative adversarial networks for alzheimer's disease image data classification, *Front. Aging Neurosci.* **13** (2021) 720226.
35. M. Merone, S. L. D'Addario, P. Mirino, F. Bertino, C. Guariglia, R. Ventura, A. Capirchio, G. Baldassarre, M. Silvetti and D. Caligiore, A multi-expert ensemble system for predicting alzheimer transition using clinical features, *Brain Inf.* **9** (2022) 20.
36. D. Castillo-Barnes, J. Merino-Chica, R. Garcia-Diaz, C. Jimenez-Mesa, J. E. Arco, J. Ramírez and J. M. Górriz, Cad system for parkinson's disease with penalization of non-significant or high-variability input data sources, *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications*, eds. J. M. Ferrández Vicente, J. R. Œlvarez-Sánchez, F. de la Paz López and H. Adeli (Springer International Publishing, 2022), pp. 23–33.
37. M. N. Haque, N. Noman, R. Berretta and P. Moscato, Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification, *PLOS ONE* **11** (2016) e0146116.
38. J. Liu, S. Shang, K. Zheng and J.-R. Wen, Multi-view ensemble learning for dementia diagnosis from neuroimaging: An artificial neural network approach, *Neurocomputing* **195** (2016) 112–116.
39. J. M. Rondina, L. K. Ferreira, F. L. de Souza Duran, R. Kubo, C. R. Ono, C. C. Leite, J. Smid, R. Nitrini, C. A. Buchpiguel and G. F. Busatto, Selecting the most relevant brain regions to discriminate alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases, *NeuroImage: Clinical* **17** (2018) 628–641.
40. L. Nanni, A. Lumini and N. Zaffonato, Ensemble based on static classifier selection for automated diagnosis of mild cognitive impairment, *J. Neurosci. Methods* **302** (2018) 42–46.
41. D. Castillo-Barnes, J. Ramírez, F. Segovia, F. J. Martínez-Murcia, D. Salas-Gonzalez and J. M. Górriz, Robust Ensemble Classification Methodology for I123-Ioflupane SPECT Images and Multiple Heterogeneous Biomarkers in the Diagnosis of Parkinson's Disease, *Front. Neuroinf.* **12** (2018) 53.
42. J. A. Simón-Rodríguez, F. J. Martinez-Murcia, J. Ramírez, D. Castillo-Barnes and J. M. Gorriz, Modelling the progression of the symptoms of parkinson's disease using a nonlinear decomposition of 123iFP-CIT SPECT images, *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications* (Springer International Publishing, 2022), pp. 104–113.
43. G. Mirzaei and H. Adeli, Resting state functional magnetic resonance imaging processing techniques in stroke studies, *Rev. Neurosci.* **27** (2016) 871–885.
44. J. Rokicki, T. Wolfers, W. Nordhøy, N. Tesli, D. S. Quintana, D. Alnæs, G. Richard, A.-M. G. Lange, M. J. Lund, L. Norbom, I. Agartz, I. Melle, T. Nærland, G. Selbæk, K. Persson, J. E. Nordvik, E. Schwarz, O. A. Andreassen, T. Kaufmann and L. T. Westlye, Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders, *Human Brain Mapping* **42** (2020) 1714–1726.
45. R. Romero-Garcia, A. S. Mandal, R. A. I. Bethlehem, B. Crespo-Facorro, M. G. Hart and J. Suckling, Transcriptomic and connectomic correlates of differential spatial patterning among gliomas, *Brain* **146** (2022) 1200–1211.
46. L. Palumbo, P. Bosco, M. Fantacci, E. Ferrari, P. Oliva, G. Spera and A. Retico, Evaluation of the intra- and inter-method agreement of brain MRI

segmentation software packages: A comparison between SPM12 and FreeSurfer v6.0, *Physica Medica* **64** (2019) 261–272.
47. J. Ashburner, G. Barnes, C.-C. Chen, J. Daunizeau, G. Flandin, K. Friston, D. Gitelman, V. Glauche, R. Henson, C. Hutton, A. Jafarian, S. Kiebel, J. Kilner, V. Litvak, J. Mattout, R. Moran, W. Penny, C. Phillips, A. Razi, K. Stephan, S. Tak, A. Tyrer and P. Zeidman, SPM12 Manual. UCL Queen Square Institute of Neurology, London, UK (2021).
48. D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, P. Padilla, F. J. Martínez-Murcia and E. W. Lang, Building a FP-CIT SPECT brain template using a posterization approach, *Neuroinformatics* **13** (2015) 391–402.
49. J. Ashburner, Computational anatomy with the SPM software, *Magnetic Resonance Imaging* **27** (2009) 1163–1174.
50. G. Solana-Lavalle and R. Rosas-Romero, Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of parkinson's disease, *Comput. Methods Programs Biomed.* **198** (2021) 105793.
51. G. Mirzaei and H. Adeli, Segmentation and clustering in brain MRI imaging, *Rev. Neurosci.* **30** (2018) 31–44.
52. A. S. Talai, J. Sedlacik, K. Boelmans and N. D. Forkert, Utility of multi-modal MRI for differentiating of Parkinson's disease and progressive supranuclear palsy using machine learning, *Front. Neurol.* **12** (2021) 648548.
53. Y. Hua, X. Shu, Z. Wang and L. Zhang, Uncertainty-guided voxel-level supervised contrastive learning for semi-supervised medical image segmentation, *Int. J. Neural Syst.* **32** (2022).
54. K. Wang, Y. Wang, B. Zhan, Y. Yang, C. Zu, X. Wu, J. Zhou, D. Nie and L. Zhou, An efficient semi-supervised framework with multi-task and curriculum learning for medical image segmentation, *Int. J. Neural Syst.* **32** (2022) 2250043.
55. A. D. Nardin, P. Mishra, G. L. Foresti and C. Piciarelli, Masked transformer for image anomaly localization, *Int. J. Neural Syst.* **32** (2022) 2250030.
56. A. Brahim, J. Ramírez, J. M. Górriz, L. Khedher and D. Salas-Gonzalez, Comparison between different intensity normalization methods in 123i-ioflupane imaging for the automatic detection of parkinsonism, *PLOS ONE* **10** (2015) 1–20.
57. D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán and E. W. Lang, Linear intensity normalization of fp-cit spect brain images using the α-stable distribution, *NeuroImage* **65** (2013) 449–455.
58. P. Padilla, J. Górriz, J. Ramírez, D. Salas-Gonzalez and I. Illán, Intensity normalization in the analysis of functional DaTSCAN SPECT images: The α-stable distribution-based normalization method vs other approaches, *Neurocomputing* **150** (2015) 4–15.
59. A. Nandi, F. Xhafa, L. Subirats and S. Fort, Reward-penalty weighted ensemble for emotion state classification from multi-modal data streams, *Int. J. Neural Syst.* **32** (2022) 2250049.
60. I. A. Illan, J. M. Gorriz, J. Ramirez, F. J. Martinez-Murcia, D. Castillo-Barnes, F. Segovia and D. Salas-Gonzalez, Support vector machine failure in imbalanced datasets, *Understanding the Brain Function and Emotions* (Springer International Publishing, 2019), pp. 412–419.
61. D. R. Cox, *Principles of Statistical Inference* (Cambridge University Press, 2006).
62. J. Ramírez, J. Górriz, A. Ortiz, F. Martinez-Murcia, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, I. Illan and C. Puntonet, Ensemble of random forests one vs. rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares, *J. Neurosci. Methods* **302** (2018) 47–57.
63. J. B. Tenenbaum, V. de Silva and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* **290** (2000) 2319–2323.
64. F. Tsai, Comparative study of dimensionality reduction techniques for data visualization, *J. Artificial Intell.* **3** (2010) 119–134.
65. B. Schlkopf and A. J. Smola, *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, Mass, 2002).
66. S. Zhai, W. Shen, S. M. Graves and D. J. Surmeier, Dopaminergic modulation of striatal function and Parkinson's disease, *J. Neural Transmission* **126** (2019) 411–422.
67. M. C. Ruppert, A. Greuel, M. Tahmasian, F. Schwartz, S. Strmer, F. Maier, J. Hammes, M. Tittgemeyer, L. Timmermann, T. van Eimeren, A. Drzezga and C. Eggers, Network degeneration in parkinson's disease: multimodal imaging of nigro-striato-cortical dysfunction, *Brain* **143** (2020) 944–959.
68. E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng and M. Joliot, Automated anatomical labelling atlas 3, *NeuroImage* **206** (2020) 116189.
69. I. Guyon, J. Li, T. Mader, P. A. Pletscher, G. Schneider and M. Uhr, Competitive baseline methods set new standards for the NIPS 2003 feature selection benchmark, *Pattern Recogn. Lett.* **28** (2007) 1438–1444.
70. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Machine Learn. Res.* **12** (2011) 2825–2830.
71. M. H. Zweig and G. Campbell, Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine., *Clinical chem.* **39** (1993) 561–577.

72. K. Sakai and K. Yamada, Machine learning studies on major brain diseases: 5-year trends of 2014–2018, *Japanese J. Radiol.* **37** (2018) 34–72.
73. M. Musavi, W. Ahmed, K. Chan, K. Faris and D. Hummels, On the training of radial basis function classifiers, *Neural Netw.* **5** (1992) 595–603.
74. S. Badoud, D. V. D. Ville, N. Nicastro, V. Garibotto, P. R. Burkhard and S. Haller, Discriminating among degenerative parkinsonisms using advanced 123i-ioflupane spect analyses, *NeuroImage: Clinical* **12** (2016) 234–240.
75. C. Jimenez-Mesa, D. Castillo-Barnes, J. E. Arco, F. Segovia, J. Ramirez and J. M. Górriz, Analyzing statistical inference maps using MRI images for parkinson's disease, *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications* (Springer International Publishing, 2022), pp. 166–175.
76. C. Owens-Walton, D. Jakabek, X. Li, F. A. Wilkes, M. Walterfang, D. Velakoulis, D. van Westen, J. C. Looi and O. Hansson, Striatal changes in parkinson disease: An investigation of morphology, functional connectivity and their relationship to clinical symptoms, *Psychiatry Research: Neuroimaging* **275** (2018) 5–13.
77. A. M. Pagnozzi, J. Fripp and S. E. Rose, Quantifying deep grey matter atrophy using automated segmentation approaches: A systematic review of structural MRI studies, *NeuroImage* **201** (2019) 116018.
78. S. Chakraborty, S. Aich and H.-C. Kim, Detection of parkinson's disease from 3t t1 weighted MRI scans using 3d convolutional neural network, *Diagnostics* **10** (2020) 402.
79. A. Vitale, R. Villa, L. Ugga, V. Romeo, A. Stanzione and R. C. and, Artificial intelligence applied to neuroimaging data in parkinsonian syndromes: Actuality and expectations, *Math. Biosci. Eng.* **18**(2) (2021) 1753–1773.
80. J. E. Arco, A. Ortiz, D. Castillo-Barnes, J. M. Górriz and J. Ramírez, Quantifying inter-hemispheric differences in parkinson's disease using siamese networks, *Artificial Intelligence in Neuroscience: Affective Analysis and Health Applications* (Springer International Publishing, 2022), pp. 156–165.
81. M. Grassi, N. Rouleaux, D. Caldirola, D. Loewenstein, K. Schruers, G. Perna and M. D. and, A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to alzheimer's disease using socio-demographic characteristics, clinical information, and neuropsychological measures, *Front. Neurol.* **10** (2019) 756.
82. J. E. Arco, A. Ortiz, J. Ramírez, F. J. Martínez-Murcia, Y.-D. Zhang and J. M. Górriz, Uncertainty-driven ensembles of multi-scale deep architectures for image classification, *Inf. Fusion* **89** (2023) 53–65.
83. J. E. Arco, A. Ortiz, J. Ramírez, Y.-D. Zhang and J. M. Górriz, Tiled sparse coding in eigenspaces for image classification, *Int. J. Neural Syst.* **32** (2021) 2250007.
84. C. H. Hawkes, K. D. Tredici and H. Braak, Parkinson's disease: A dual-hit hypothesis, *Neuropathol. Appl. Neurobiol.* **33** (2007) 599–614.
85. S. Hulbert and H. Adeli, EEG/MEG- and imaging-based diagnosis of alzheimer's disease, *Rev. Neurosci.* **24** (2013) 563–576.
86. F. Moradi, H. Mohammadi, M. Rezaei, P. Sariaslani, N. Razazian, H. Khazaie and H. Adeli, A novel method for sleep-stage classification based on sonification of sleep electroencephalogram signals using wavelet transform and recurrent neural network, *Eur. Neurol.* **83**(5) (2020) 468–486.
87. J. P. Amezquita-Sanchez, N. Mammone, F. C. Morabito and H. Adeli, A new dispersion entropy and fuzzy logic system methodology for automated classification of dementia stages using electroencephalograms, *Clinical Neurol. Neurosurgery* **201** (2021) 106446.
88. M. C. Maya-Piedrahita, P. M. Herrera-Gomez, L. Berrío-Mesa, D. A. Cárdenas-Peña and A. A. Orozco-Gutierrez, Supported diagnosis of attention deficit and hyperactivity disorder from EEG based on interpretable kernels for hidden Markov models, *Int. J. Neural Syst.* **32** (2022) 2250008.
89. H. A. Ardakani, M. Taghizadeh and F. Shayegh, Diagnosis of autism disorder based on deep network trained by augmented EEG signals, *Int. J. Neural Syst.* **32** (2022) 2250046.
90. N. Mammone, C. Ieracitano, H. Adeli and F. C. Morabito, AutoEncoder filter bank common spatial patterns to decode motor imagery from EEG, *IEEE J. Biomed. Health Informatics* **27**(5) (2023) 1–12.