applied sciences

MDPI

Article

Multi-Class Classifier in Parkinson's Disease Using an Evolutionary Multi-Objective Optimization Algorithm

Ignacio Rojas-Valenzuela [1,*], Olga Valenzuela [2,*], Elvira Delgado-Marquez [3] and Fernando Rojas [1]

[1] School of Technology and Telecommunications Engineering, University of Granada, 18071 Granada, Spain; frojas@ugr.es
[2] Department of Applied Mathematics, University of Granada, 18071 Granada, Spain
[3] Department of Economics and Statistics, University of Leon, 24004 Leon, Spain; elvira.delgado@unileon.es
* Correspondence: e.ignaciorojas@go.ugr.es (I.R.-V.); olgavc@ugr.es (O.V.)

Abstract: In this contribution, a novel methodology for multi-class classification in the field of Parkinson's disease is proposed. The methodology is structured in two phases. In a first phase, the most relevant volumes of interest (VOI) of the brain are selected by means of an evolutionary multi-objective optimization (MOE) algorithm. Each of these VOIs are subjected to volumetric feature extraction using the Three-Dimensional Discrete Wavelet Transform (3D-DWT). When applying 3D-DWT, a high number of coefficients is obtained, requiring the use of feature selection/reduction algorithms to find the most relevant features. The method used in this contribution is based on Mutual Redundancy (MI) and Minimum Maximum Relevance (mRMR) and PCA. To optimize the VOI selection, a first group of 550 MRI was used for the 5 classes: PD, SWEDD, Prodromal, GeneCohort and Normal. Once the Pareto Front of the solutions is obtained (with varying degrees of complexity, reflected in the number of selected VOIs), these solutions are tested in a second phase. In order to analyze the SVM classifier accuracy, a test set of 367 MRI was used. The methodology obtains relevant results in multi-class classification, presenting several solutions with different levels of complexity and precision (Pareto Front solutions), reaching a result of 97% as the highest precision in the test data.

Keywords: Parkinson's disease (PD); 3D-discrete wavelet transform (3D-DWT); support vector machine (SVM); multi-objective optimization evolutionary algorithm (MOE); minimum redundancy maximum relevance (mRMR)

Citation: Rojas-Valenzuela, I.; Valenzuela, O.; Delgado-Marquez, E.; Rojas, F. Multi-Class Classifier in Parkinson's Disease Using an Evolutionary Multi-Objective Optimization Algorithm. *Appl. Sci.* **2022**, *12*, 3048. https://doi.org/10.3390/app12063048

Academic Editor: Zimi Sawacha

Received: 10 February 2022
Accepted: 10 March 2022
Published: 16 March 2022

1. Introduction

Dementia is a clinical syndrome characterized by a progressive deterioration of cognitive functions. Currently, research in brain diseases and dementia is one of the most important fields of investigation. According to the Parkinson Foundation (https://www.parkinson.org/, accessed on 1 February 2022), 46.8 million people worldwide are living with some form of dementia. The same report predicts that the number of people with dementia will unfortunately double every 20 years.

Parkinson's disease (PD) is the second most common neuro-degenerative disorder [1]. PD has an incidence rate of 16 to 19 per 100,000 people per year, according to the World Health Organization [2], and it is estimated that this value will unfortunately increase in the coming years due to the ageing of the population [3]. According to recent studies, there are currently more than 10 million people worldwide who suffer from Parkinson's disease [4], and by the year 2040, it will be an important cause of death in the elderly population.

These evidences show the urgent need to advance in Parkinson's research, since an early diagnosis helps to delay the development of the disease in the patient.

Unfortunately, at present, a cure for Parkinson's has not been obtained, although the pharmaceutical industry has developed therapies that slow down the development

of PD, giving patients a minimum quality of life for a more extended period of time [5]. Neurological damage and loss of neurons are triggers in Parkinson's disease, resulting in cognitive dysfunction. In this disease, neurons degenerate in an area of the basal ganglia (called the substantia nigra) [3], which are groups of neurons found deep in the brain that help smooth muscle movements and coordinate muscle movements postural changes. When neurons degenerate in one area of the basal ganglia, dopamine production and the number of connections between neurons in the basal ganglia decrease. As a consequence, the basal ganglia do not smooth or harmonize movements as they normally do, leading to tremors, lack of coordination, slow movements (bradykinesia), a tendency to move less (hypokinesia), and postural problems associated with gait.

As has been analyzed in the literature, not all Parkinson's patients suffer the same symptoms or with the same intensity [6], affecting people who suffer from this disease differently. However, typical patterns of progression in Parkinson's disease have been established in an attempt to establish the stages of this disease [7]. The first or initial stage is denoted as "Stage One", mainly defined by patients that have mild symptoms of the disease. The last stage is "Stage Five", which is the most advanced and incapacitating stage, wherein rigidity and inflexibility in the legs makes the patient barely have the motor ability to walk [8].

The methodology usually used in PD diagnosis consists of using clinical criteria, which are based on medical evaluation and include a series of tests performed on patients to address their stage. Current diagnoses are based on the presence of motor signs, which means that patients with PD are inevitably diagnosed at an advanced stage. It is important to highlight that a neuro-protective therapy initiated in a late stage of the disease may have less substantial effects on limiting disease progression. Therefore, it is essential to find methodologies and tools that make it possible to distinguish, classify and diagnose patients in early stages of PD from the healthy population [8].

In clinical practice, various objective measures have been adopted for the differential diagnosis of PD, including a series of olfactory, electrophysiological and neuro-physiological tests. However, and with special emphasis in the last decade, the most developed area for providing an objective evaluation is neuroimaging [9].

The main diagnostic methods of PD consist of using human expert clinical criteria. These criteria are based on medical evaluation (including a range of olfactory, motor/biomechanical ability, electrophysiological and neuro-physiological tests) to diagnose the patient's disease state.

In addition to these criteria, nowadays, neuroimaging is of great importance for the diagnosis of the disease [1,9].

One of the most relevant techniques in the field of neuroimaging is Magnetic Resonance Imaging (MRI), being a non-invasive, high spatial and temporal resolution technology. MRI is a powerful tool in the field of medical imaging that has been fruitfully applied for the analysis and diagnosis of abnormalities in the brain for different diseases.

In this study, we combined MRI with machine learning algorithms in order to develop a methodology for the automatic classification of the MRI. The procedure implements three-dimensional discrete wavelet transform (3D-DWT) as a powerful tool to extract 3D characteristics from the brain and an optimization algorithm (based on multi-objective genetic algorithms) to determine and analyze which brain regions have the greatest impact on the disease. Using the most relevant regions, an intelligent classifier based on Support Vector Machine (SVM), is constructed with the aim of being able to automatically classify different types of patients (multiclass classification: Parkinson's disease, prodomal, SWEDD, gene cohort and normal patients)

2. State-of-the-Art

2.1. Diagnosis Tools

As mentioned in the Introduction section, the ability and knowledge of the expert neurologist, together with the external clinical symptoms that the patient presents, are the classically used tools in the diagnosis of PD.

One of the non-invasive methods commonly analyzed with precision for the evaluation and diagnosis of the status of PD is the analysis of initial walking patterns. Also known as gait cycles, these walking patterns measure the initial to final contact of the same foot when walking and were observed as a consequence of central pattern generators established by spinal circuits [10]. However, other parts of our brain, such as the basal ganglia, the cerebral cortex and the descending pathways to the spinal cord, also have a great influence on human locomotion.

Currently, neuroimaging has been implemented as a powerful source of information for decision making (mainly relevant in the early diagnosis of brain disorders), being an important medical test for neurologists.

Among the best known and most used technologies, it is worth mentioning structure MRI (sMRI), functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). In the field of PD, MRI has been used for the discrimination of PD and other Parkinsonian syndromes that present relevant structural modifications/changes [11]. MRI has also been used for the identification of possible patterns associated with volumetric alterations that may indicate the severity of the disease [12].

Single-Photon Emission Computed Tomography (SPECT) is based on the Dopamine Transporter scan (DaTscan) method by using the radioligand Ioflupane I 123 (123I-FP-CIT). This technology has been successfully used to differentiate PD from other motor diseases non-associated with dopaminergic deficit [13,14].

There are studies in the literature that show how glucose hypo-metabolism occurs prior to localized atrophy of brain matter and that the conversion time is correlated with this hypo-metabolism. The use of FDG-PET images [15] has become an effective tool for this information.

The numerous contributions presented in the literature analyzing the genetic factors underlying Parkinson's disease with dementia and with mild/moderate cognitive impairment are very promising and clearly demonstrate the importance of genetic for cognitive impairment in PD [16].

As presented in the Introduction section, one of the reasons that characterize Parkinson's disease comes from the deterioration of dopamine-producing cells located in the substantia nigra of the brain [17], resulting in motor and non-motor symptoms (tremors at rest, motor retardation, bradykinesia, rigidity, etc). For the elderly, the identification and proper treatment of PD as early as possible is of great importance to be able to act medically and clinically in the prevention of the patient, with the intention of slowing down the progress of the disease as much as possible.

2.2. Computer Studies and Algorithms in PD Classification

MRI analysis for PD classification is a very fruitful area of research; indeed, in recent years, many relevant papers have been published developing novel methodologies [18–21].

Most of the articles cited in the bibliography focus on the realization of binary classifiers in order to differentiate between control and PD patients. The use of intelligent systems in the field of neuroimaging has shown excellent synergy, developing precise automatic diagnostic systems [12,14,22].

For example, in Long et al. [23], 19 early PD patients and 27 normal patients (volunteers) recruited from Zhejiang University were used for the classification using a SVM classifier. For each patient, resting-state functional magnetic resonance imaging (rsfMRI) and structural images were used. Three different types of characteristics were extracted from the rsfMRI images: ReHo (regional homogeneity), ALFF (amplitude of low-frequency fluctuations) and RFCS (regional functional connectivity strength). For the structural im-

ages, the volume characteristics from the gray matter, the white matter and the cerebrospinal fluid were computed. With this reduced data set, the accuracy was 87%, sensitivity was 79% and specificity was 93%.

In Lei et al. [24], a multi-class classification system for PD analysis, based on a sparse discriminative feature selection methodology, is presented. The authors presented a framework to create a least square regression model based on Fisher's linear discriminant analysis and locality preserving projection. Input data were acquired from the Parkinson's Progression Markers Initiative (PPMI) database, using 123 PD, 56 normal and 29 SWEDD images and obtaining an accuracy of 78.4%.

A methodology based on complex networks for precise early diagnoses using MRI is presented by Amoroso et al. [18]. The authors defined a network model of brain regions and associate proper connectivity measures to each region, with the brain being characterized through a feature vector encoding the local relations where brain regions interweave. Feature selection and learning is performed by using Random Forests. SVM is used for classification normal (169 patients) and PD (374 patients) from the PPMI database, obtaining an accuracy of 93%.

An interesting approach is presented by Oliveira et al. [25], in which the potential of a set of features obtained from [123I]FP-CIT SPECT brain images are used. The authors selected seven features (five associated to uptake ratios on the striatum and two features correlated to the estimated volume and length of the striatal region with normal uptake) from each brain hemisphere. The classification capacities of each individual feature and groups of features were analyzed using a data set of 209 normal and 443 PD patients from PPMI. The best classification performance was achieved using the SVM classifier, using all the features (accuracy = 97.9%, sensitivity = 98% and specificity = 97.6%).

In recent years, Convolutional Neural Networks (CNN) have been extensively used in the automatic classification of images using Deep Learning (DL) neural networks and in particular in PD diagnosis [26,27]. For example, in Sivaranjini et al. [27], MR images of Parkinson's disease subjects and healthy control volunteers were classified by a DL neural network, showing an accuracy of 88.9%, based on the PPMI data set. The CNN architecture, AlexNet, is used to improve the diagnosis of PD via the transfer learning methodology.

In [28], a deep learning framework for the simultaneous classification and regression of PD patients using personal information (i.e., age, gender) and MRI was proposed. The data set consists of 452 Parkinson patients, including 292 males and 160 females and 204 images from people in healthy conditions (134 males and 70 females). The authors achieved a 100% accuracy in both the test and validation sets and verified that the Basal Ganglia and the Substantia Nigra parts of the brain, as was already known by medical experts, are important in the diagnosis of Parkinson's.

A methodology for creating predictive markers of PD using CNNs based on NeuroMelanin Sensitive Magnetic Resonance Imaging (NMS-MRI) is presented in [19]. NMS-MRI has been decisive in identifying abnormalities in the substantia Nigra pars compacta (SNc) in patients suffering Parkinson, analyzing the loss of dopaminergic neurons in the SNc. The proposed framework supports discriminating PD from atypical Parkinsonian syndromes, achieving an accuracy of 85.7% in test patients.

In this paper, a useful framework for multi-class classification is proposed for PD recognition using a Multi-Objective Optimization Evolutionary Algorithm. To the best of our knowledge, there are very few, if any, references in the literature that perform 3-D feature extraction for VOI optimization related to Parkinson 's disease, as 2D images have instead been more frequently used for binary classification. The high accuracy achieved during the testing phase, reaffirms the usefulness of the use of the different hybridized methodologies and techniques: 3D-DWT for feature extraction in MRI, dimensionality reduction in the problem using the most relevant features applying mRMR and PCA and the use of a multi-objective optimization algorithm for the selection of the relevant brain regions, which will be used by the SVM classifier.

3. Materials and Methods

The proposed methodology presented in this contribution for Parkinson's disease detection using MRI consists of five main stages for training the system in order to obtain the regions/volume of interest (VOI): (1) Pre-processing the NIfTI files (normalization, segmentation); (2) Feature Extraction (using 3D discrete wavelet transform (3D-DWT)); (3) Feature Selection and Reduction (using minimum Redundancy Maximum Relevance (mRMR) and PCA algorithms); (4) Detection of regions/volume of interest (VOI) (using multi-objective genetic algorithm NSGA-II); and (5) Optimization of the SVM classifier with the selected VOI. The block scheme for multi-objective optimization (VOIs selection) is shown in Figure 1. The main functional blocks are described in the following sub-sections.

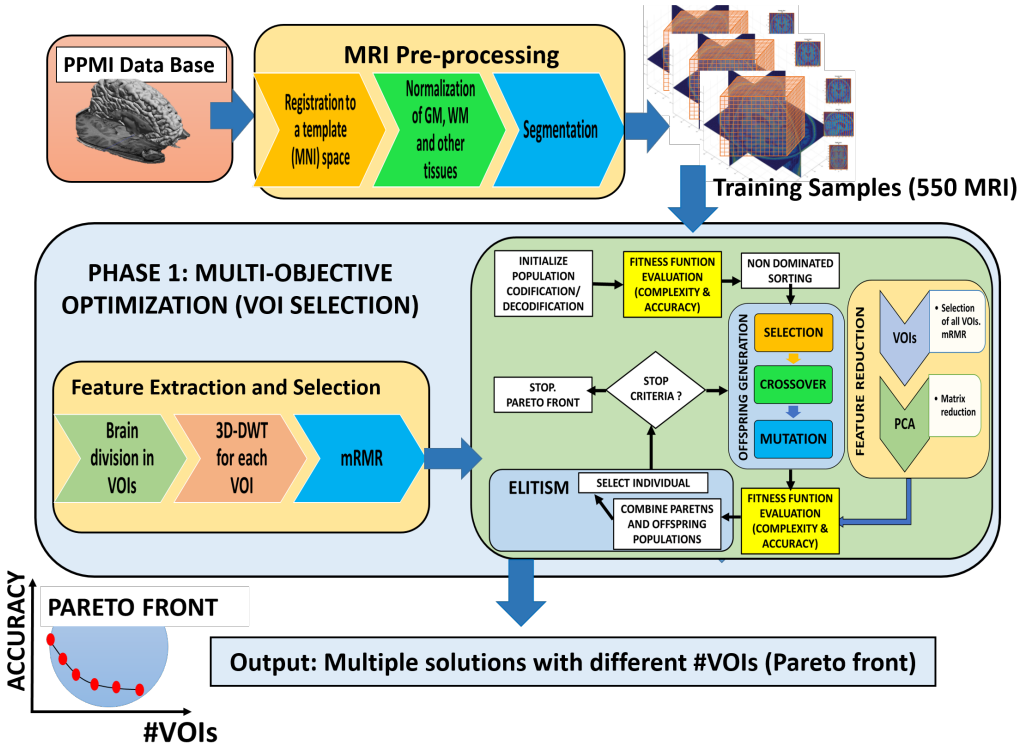


Figure 1. The block scheme for multi-objective optimization. Phase 1: VOI selection.

The phases for the performance assessment with test samples, similar to training phases, are as follows: (1) Pre-processing the NIfTI files (normalization, segmentation); (2) Feature Extraction (using 3D discrete wavelet transform (3D-DWT)); (3) Feature Selection and Reduction (using minimum Redundancy Maximum Relevance (mRMR) and PCA algorithms); (4) Classification of test samples through the SVM classifier previously trained with selected VOI; (5) Performance assessment by the confusion matrix. Figure 2 shows the block scheme of the test phase, being the results of the proposed methodology presented in Section 5.

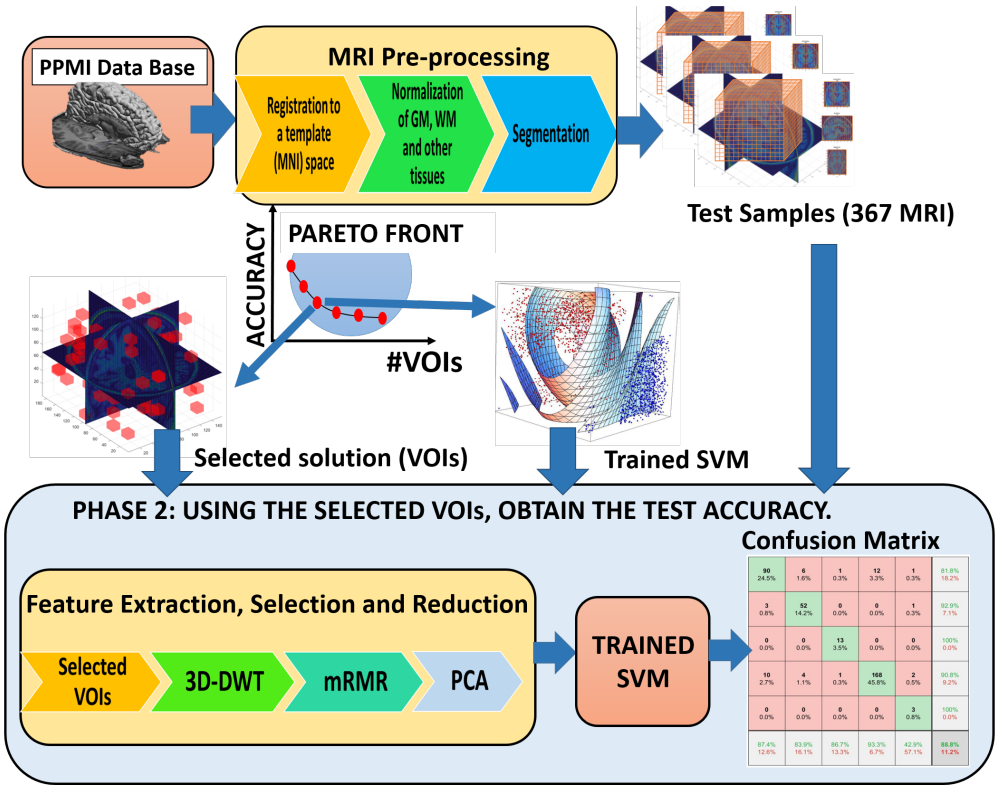


Figure 2. Phase 2: Block scheme for the test phase. Selected VOIs from the Pareto Front are used to construct the SVM classifier.

3.1. Subjects Cohort

The data used in this paper were obtained from the Parkinson's Progression Markers Initiative (PPMI) database [29]. The PPMI portal was created in 2010 supported by the Michael J. Fox Foundation and a group of scientists and academics. The main objective was the collection and analysis of data from Parkinson's patients in order to carry out a clinical study of critically necessary biological markers of the onset and progression of this disease.

The data and samples acquired from the participants (patients and volunteers) will allow the development of a complete Parkinson's database and bio-repository, with a large set of MRI images of patients of various categories.

In this contribution, different groups of patients existing in the PPMI data base have been used (Normal, SWEDD, Prodromal, PD and Gene Cohort). These groups are briefly described below.

Analyzing patients with Single Photon Emission Computed Tomography (SPECT) technology has allowed the detection of a set of patients who were initially diagnosed as suffering Parkinson's disease but who should more precisely be associated to "Scans Without Evidence of Dopaminergic Deficit (SWEDD)" [13]. The definition and determination of the term SWEDD has created some controversy in the literature, referring to the absence, rather than the presence, of a relevant abnormality in the image of a patient who has been diagnosed with PD. While there are several papers in the literature suggesting that various PD-like disorders could explain some of the SWEDD cases, others have indicated that some of them could have a benign PD sub-type. As a result, there has been controversy and confusion, and the use of this term remains unclear [30,31]. For example, there are contributions that conclude the fact that SWEDD are an initial stage of Parkinson's disease [32]. However, there are other authors who suggest that SWEDD patients do not progress to PD [33], maintaining the assumption that SWEDD patients suffer from another motor disorder [34].

A relevant analysis is also presented in [35], in which it is concluded that SPECT misinterpretation can occur in the diagnosis of SWEDD patients, and it is recommended

that suspected SWEDD patients pose not only a serious diagnostic challenge but also should prompt a detailed scan reassessment. Thus, there is a clear need for reliable and accurate systems that are capable of differentiating PD from other motor disorders and diseases [36].

The second group of patients analyzed in this contribution is logically PD. For the inclusion of this type of patient in the PPMI databases, a set of requirements is compulsory to be met: in summary, patients aged 30 or over, of both genders, the existence of a validation through presynaptic dopaminergic examinations, a scale of I or II on the Hoehn and Yahr scale at the beginning and, finally, without expectation of medication for PD within six months from baseline. It must be emphasized that there are patients having multiple NIfTI files, and all of them are used in our study. The insertion conditions for the SWEDD patients were similar as those for the PD group, except that subjects were required to have no evidence of dopaminergic deficit in presynaptic dopaminergic scans.

It is important to emphasize that the recognition and characterization of the first signs of PD disease, even in very early stages, is of great clinical utility to be able to start the treatment of a patient as soon as possible. In this sense, evidence from clinical, neuropathological and imaging research suggests the initiation of PD-specific pathology prior to the external symptoms of the disease, such as the appearance of classical motor dysfunctions (such as bradykinesia) [37,38].

The third group which was analyzed in this article are the prodromal patients. In 2013, the PPMI initiative launched a new prodromal cohort. This new cohort seeks to enroll subjects without Parkinson's disease, and as for previous groups, individuals are required to have a series of requirements. Briefly, they need to meet at least one of the following criteria: hyposmia, REM sleep behaviour disorder or LRRK2 gene mutation. Subjects enrolled in this PPMI prodromal cohort will be 60 years of age or older, and none of these subjects will have PD.

Finally, we have a class of few subjects for whom the classification is based on genetic diagnosis. This class is called Genetic Cohort Subjects (GenCohort) and is made up of subjects with and without Parkinson's disease who have a genetic mutation in gene LRRK2, GBA or SNCA.

For the construction of the database in this article, more than 1000 different MRI brain images stored in NIfTI (Neuroimaging Informatics Technology Initiative) format were used. These images were obtained at 1.5T and 3T using several medical devices, always keeping in mind the specific protocol established by PPMI [29].

The first step consisted of the elimination of duplicated and corrupted images. Afterwards, the set of images was reduced to a total of 917 MRI images, corresponding to the 5 types of patients that were analyzed. The normal (control) group contained 258 images, SWEDD group contained 154 images, Prodromal contained 38 images, PD contained 450 images and, finally, GenCohort contained 17 files.

In summary, Tables 1–3 show the number of patients in each group, indicating the number of women and men, together with the mean values and age and weight standard deviation. The volume of processed information is considerable, being around 55 GB of information (resulting in more than 85 GB after storing the pre-processed and segmented images).

Table 1. Characteristics of the cohort of normal and SWEDD images used in this contribution.

Measures	**Normal (258)**				**SWEDD (154)**			
Sex (Male/Female)	**181**		**77**		**100**		**54**	
Age (years)	61.4	11.7	59.0	8.2	64.1	9.7	57.8	10.7
Weight (kg)	83.0	10.7	67.9	17.1	88.6	12.8	75.8	12.8
	Mean	Std	Mean	Std	Mean	Std	Mean	Std

Table 2. Characteristics of the cohort of Prodromal and PD images used in this contribution.

Measures	Prodromal (38)				PD (450)			
Sex (Male/Female)	30		**8**		**288**		**162**	
Age (years)	67.1	3.9	67.8	3.1	62.4	9.8	60.2	8.7
Weight (kg)	88.8	14.5	68.5	9.7	90.8	16.1	66.3	11.9
	Mean	Std	Mean	Std	Mean	Std	Mean	Std

Table 3. Characteristics of the cohort of GeneCohort images used in this contribution.

Measures	GeneCohort (17)			
Sex (Male/Female)	**14**		**3**	
Age	62.1	10.6	72.2	9.0
Weight (kg)	87.2	11.3	64.7	5.7
	Mean	Std	Mean	Std

3.2. Pre-processing: Normalization and Segmentation

The image pre-processing process is a fundamental stage to homogenize, normalize and segment the MRIs of the different patients. This process has been carried out with the statistical tool Statistical Parametric Mapping (SPM) [39,40], in the SPM12 version. Therefore, bias correction, spatial normalization and segmentation are performed. The new version is based on the method described by J. Ashburner et al. [41] known as proposed unified segmentation algorithm, in which a probabilistic framework was defined, with which image registration, bias adjustment and tissue categorization can be merged in a unified model. The model obtained is based on a mixture of Gaussian values with the possibility to incorporate a smooth intensity adaptation and a non-linear protocol with tissue probability maps.

All the MRIs from patients and volunteers in this contribution were spatially normalized. The bounding box of 157 × 189 × 130 voxels in the X, Y and Z directions, respectively, were used. It should be kept in mind that the MRI are volumetric (three-dimensional), and in this contribution, a three-dimensional wavelet transform will also be performed, with which each voxel is adjusted to the measure of 1 mm × 1 mm × 1 mm. In the study of brain images, it is relevant to separate or segment what is called gray matter (C1 images), white matter (C2 images) and whole matter (W images). These images were available in different files.

It is relevant to note that MRI images are often corrupted by an artifact or noise that varies spatially and changes the intensity level of the brain image (bias). Although the noise that can be present in an image may not excessively affect the visual inspection by an expert neurologist, it can adversely influence the performance of an automatic image pre-processing methodology, and therefore, it requires removal.

3.3. Feature Extraction

Each volume of interest, VOI, must be characterized by a set of characteristics or features. These features (as a real number vector) will represent each one of the brain images.

A powerful transform based on the windowing technique with variable size is the Wavelet Transform (WT). The main feature of WT is that it maintains both the time and frequency information of the analyzed signal, adopting ”scale” instead of traditional ”frequency”; namely, it does not generate a time-frequency interpretation but a time-scale understanding of the signal. The discrete wavelet transform (DWT) is a helpful implementation of the wavelet transform, frequently used in numerical analysis and functional analysis, for which the wavelets are discretely sampled. The procedure to perform the

wavelet transform is that a fully scalable modulated window is shifted along the signal for every position. This procedure is repeated modifying the amplitude of the window, which leads to obtaining a multi-resolution analysis. In the case of its application in images, the decomposing of the original image in its wavelet coefficients will generate a series of images with different scales. The computation of these coefficients is calculated as follows.

First, the mother or kernel wavelets define a Hilbert basis, which is constructed as the family of functions $\{\psi_{jk}; j, k \in \mathbb{Z}\}$ by means of translations and dilations of the mother wavelet ψ:

$$\psi_{j,k}(x) = 2^{\frac{j}{2}}\psi(2^j x - k) \tag{1}$$

The integral wavelet transform of $f(x)$ relative to an specific mother wavelet is defined by the following equation:

$$[W_\psi f](a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi\left(\frac{x-b}{a}\right) f(x)dx \tag{2}$$

In this way, the wavelet coefficients c_{jk} are given by:

$$c_{jk} = [W_\psi f](2^{-j}, k2^{-j}) \tag{3}$$

These operations are discretized to implement the Discrete Wavelet Transform, where translations and dilations convert to dyadic scales and positions by limiting a and b to a discrete lattice $\{a = 2^b; a > 0; a, b \in \mathbb{R}\}$ so that the DWT coefficients are now defined as [42]:

$$c_{j,k}^{app}(n) = DS\left[\sum_n x(n) g_j^*(n - 2^j k)\right] \tag{4}$$

$$c_{j,k}^{det}(n) = DS\left[\sum_n x(n) h_j^*(n - 2^j k)\right] \tag{5}$$

where "app" refers to "approximation", hence approximation coefficients; "det" refers to "detail", hence detail coefficients; "DS" refers to down sampling; and $g(n)$ and $h(n)$ denote a low-pass filter and a high-pass filter, respectively.

The parameters j, k presented in the previous equations are interpreted as the wavelet scale and translation factors.

In the case of having a one-dimensional signal, 1D-DWT can be applied. Each filter yields one sub-band, low-filter outputs the low-frequency sub-band L and high-filter outputs the high-frequency sub-band.

In the case of having images, 1D-DWT can be applied iteratively, obtaining four sub-bands: Low–Low, Low–High, High–High and High–Low, denoted as LL, LH, HH and HL, respectively, LL being the approximation coefficients of the image, while LH, HH and HL are the detailed coefficients. Different levels of decomposition can be obtained, performing the transformation to the approximation coefficients. As the level of decomposition is increased, further sub-bands are computed and more compact yet coarser approximation coefficients are acquired. For example, Figure 3 shows a two-level decomposition for a brain image.

If we assume that the initial image has a size of N × N, in the first level decomposition, the size of the approximation coefficients would be reduced to N/2 × N/2, and the size of the second level will be reduced to N/4 × N/4.

Traditionally, slices of the brain are fetched and used to extract features [20,43,44], using, for example, the 2D-DWT Daubechies family on a sagital slice [45].

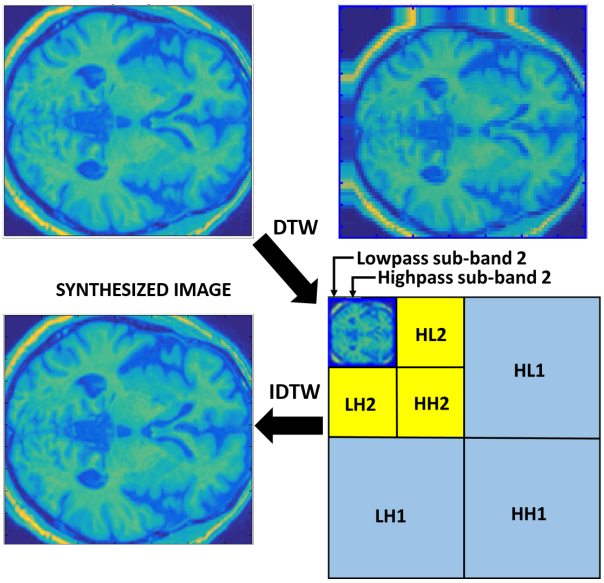


Figure 3. Operating mode of the 2D wavelet transform applied to images.

One of the main novelties of the methodology presented in this document is the extension of the 2-D image field to the 3-D volumetric analysis, using a three-dimensional volumetric feature extraction (wavelet transform), instead of two-dimensional wavelet transform of the slices selected by an expert neurologist. The main goal of the volumetric analysis is finding those regions of the brain that will have more relevance in the characterization of the different classes analyzed in order to obtain a precise classifier. Additionally, a better understanding of the disease is expected by characterizing the most relevant regions of the brain for the construction of the classifier.

In this paper, the three-dimensional wavelet transformation (3D-DWT) has been used in order to extract features for each of the selected volumes of interest (VOI). The 3D-DWT wavelet transform can be considered as an extension to an additional dimension [46], of the one existing for two-dimensional spaces (being the two-dimensional, an extension of spaces with a single dimension).

Figure 4 shows a block diagram of how the 3D-DWT transform is performed on a selected brain volume (VOI). Initially, there is a decomposition, using the 1D-DWT transform along the columns, which assumes the wavelet transform in the X dimension. This gives the volumes L and H (green boxes) corresponding to the low-pass volume and a volume high pass, respectively. The amount of data in each of these volumes is reduced. If we assume that the original volume has a data number of N_x, N_y and N_z pixels, corresponding to each of the dimensions (x, y, z) in the low-pass volume L(x,y,z), the number of pixels would be $(N_x/2 * N_y * N_z)$. After this application of the wavelet transform in the x dimension, the decomposition is performed in the rows (y axis), with the generation of new subvolumes (the orange boxes): HH, Hl, LH and LL. As in the previous decomposition, the amount of data is also reduced, being for these volumes: $(N_x/2 * N_y/2 * N_z)$. The final phase is the decomposition along the cuts (z axis), obtaining eight sub-bands (blue boxes): HHH, HHL, HLH, HLL, LHH, LHL, LLH and LLL, with the number of data: $(N_x/2 * N_y/2 * N_z/2)$.

In the specialized wavelet literature and its applications in images [47,48], there are a large number of variants for the determination of the mother wavelet, opting in this paper for the use of the biorthogonal 3.3 up to level 2. When a volumetric magnetic resonance image is being processed, various filtering procedures must be performed, some of them to select the type of matter to be analyzed. In this work, the whole matter (W) and the gray matter (C1) of the brain are used, thus performing a wavelet transform on each of these matters for each of the volumes analyzed. Figure 5 shows the feature matrix of 3D-DWT wavelet coefficients obtained for each specific VOI, in whole and gray matter.

Each feature corresponds to the coefficients of the wavelet transform, obtaining M coefficients for each of the volumes analyzed in each patient and for each type of matter (W and C1). The brain of each patient is divided into 1728 different, non-overlapping VOIs, each one with a dimension of $12 \times 14 \times 10$ voxels.

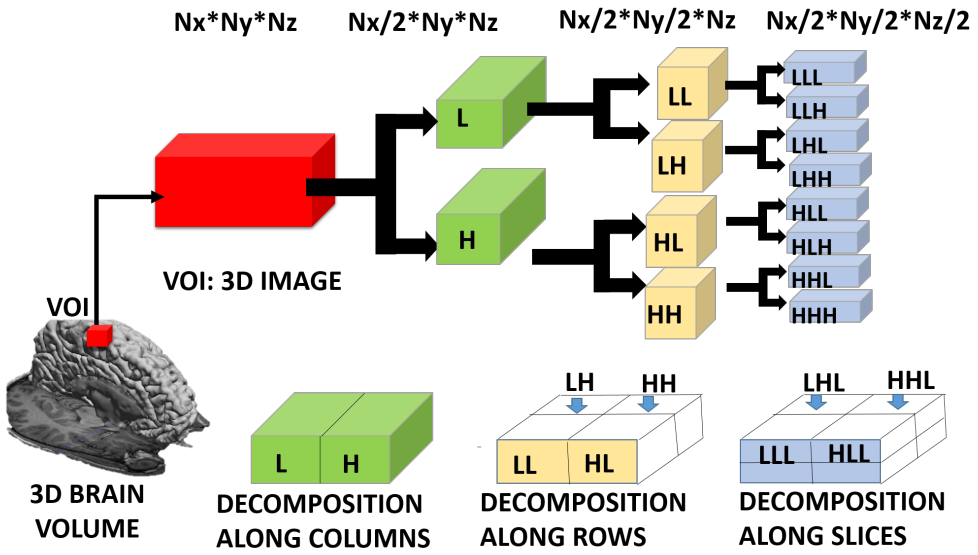
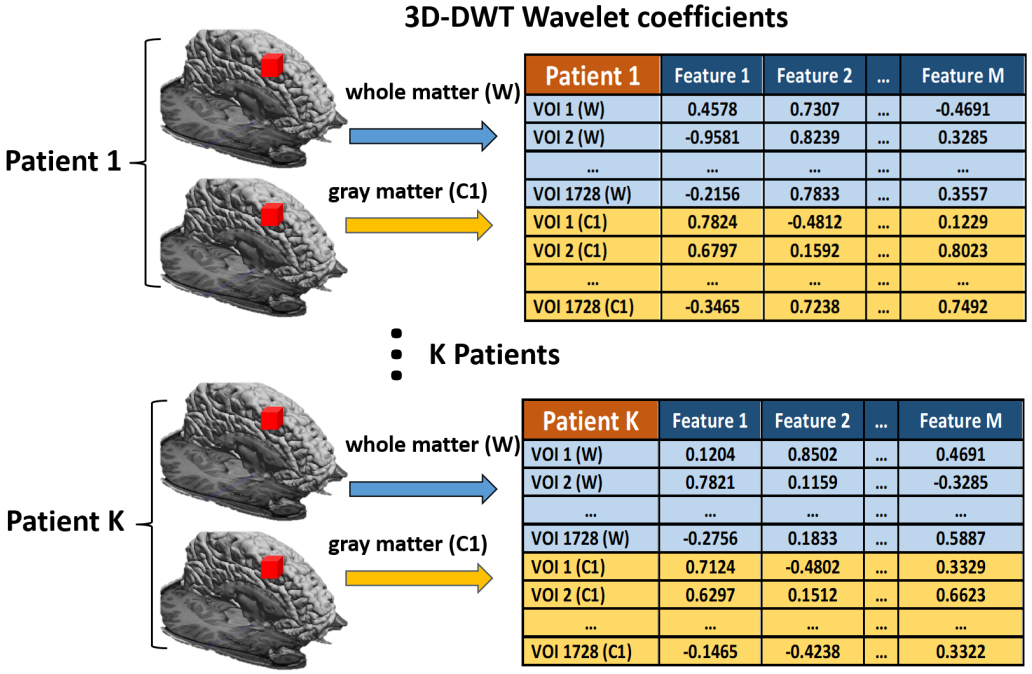


Figure 4. Operating mode of the 3D-DWT wavelet decomposition applied to a specific VOI.



Patient 1	Feature 1	Feature 2	...	Feature M
VOI 1 (W)	0.4578	0.7307	...	-0.4691
VOI 2 (W)	-0.9581	0.8239	...	0.3285
...	...	...	...	...
VOI 1728 (W)	-0.2156	0.7833	...	0.3557
VOI 1 (C1)	0.7824	-0.4812	...	0.1229
VOI 2 (C1)	0.6797	0.1592	...	0.8023
...	...	...	...	...
VOI 1728 (C1)	-0.3465	0.7238	...	0.7492

Patient K	Feature 1	Feature 2	...	Feature M
VOI 1 (W)	0.1204	0.8502	...	0.4691
VOI 2 (W)	0.7821	0.1159	...	-0.3285
...	...	...	...	...
VOI 1728 (W)	-0.2756	0.1833	...	0.5887
VOI 1 (C1)	0.7124	-0.4802	...	0.3329
VOI 2 (C1)	0.6297	0.1512	...	0.6623
...	...	...	...	...
VOI 1728 (C1)	-0.1465	-0.4238	...	0.3322

Figure 5. Feature matrix of 3D-DWT wavelet coefficients for each specific VOI in whole and gray matter.

3.4. Feature Selection

This paper addresses a classification problem where there exist multi-classes (C classes) for each pattern. Each pattern (patient) is represented by the vector formed by the set of M variables or features (as presented in Figure 5) and an additional value that indicates the class to which it belongs, that is, $Patient_k = v_i, c_k$ with $i = 1, \ldots, M$ and $c_k = 1, \ldots, C-1$.

In this contribution the number of non-balanced classes is 5, represented by 1 = Control, 2 = SWEDD, 3 = Prodomal, 4 = PD and 5 = GenCohort.

Mutual information is the most general measure of statistical independence between two random variables v_i and v_j and quantifies the information that v_i and v_j share (how knowledge of one of them affects the uncertainty about the other).

Formally, the mutual information (I) between two random variables is defined from the marginal density functions $p(v_i)$ and $p(v_j)$, and the joint density function $p(v_i, v_j)$ as follows:

$$I(v_i, v_j) = \int\int p(v_i, v_j) log \frac{p(v_i, v_j)}{p(v_i)p(v_j)} d(v_i)d(v_j) \quad (6)$$

If v_i and v_j are independent, the knowledge of v_i does not give any information about v_j, being in this case $I = 0$. At the other extreme, if v_i and v_j are identically distributed, the knowledge of v_i implies full knowledge of v_j and vice versa.

The main difficulty in calculating Equation (6) lies in estimating the densities $p(v_i, v_j)$, $p(v_i)$ and $p(v_j)$ using only the data. For categorical variables, Equation (6) is reduced to a summation, and the calculation of mutual information is immediate by performing a simple count. Therefore, instead of the evaluation of the integral in Equation (6) for continuous variables, one possible solution (simplification) is to incorporate a discretization step in the data pre-processing. Determining the appropriate discretization is not an easy task for this specific problem. In this case, other estimation methods of density functions as maximum likelihood, Bayesian estimation or non-parametric techniques were used.

The objective of a variable selection method based on mutual information is to find a subset S of m variables $S = v_1, \ldots, v_m$ that jointly have the maximum dependency (MaxDep) on the target variable c:

$$maxD(S, c), D = I(v_i, i = 1 \ldots, m; c) \quad (7)$$

Although MaxDep is an optimal theoretical criterion, the estimation of $p(v_i, \ldots, v_m)$ and $p(v_1, \ldots, v_m, c)$ is difficult to be computed due to the fact that the number of available examples is often insufficient and/or is computationally expensive. It is possible to use MaxDep when a small number of variables are wanted to be selected and the number of patterns is large. In this way, the lack of data for the estimation of the density functions would be solved, and their calculation would be carried out in a reduced feature space where the computational cost is affordable.

As an alternative to MaxDep, the set of variables based on the dependency with the class (relevance) can be selected. The criterion of Maximum Relevance consists in finding the subset of variables S that maximize $D(S, c)$, where D has been approximated as the mean value of the mutual information between each of the variables that make up the subset S and the class, that is:

$$D = \frac{1}{|S|} \sum_{v_i \in S} I(v_i; c) \quad (8)$$

Therefore, the optimal subset selected by the algorithm will be the m variables with more mutual information with class c.

It is possible that the variables selected by MaxDep are quite redundant due to a great dependency between them. If the dependency between two variables is high, its discriminative power is not greatly affected by removing one of them. Therefore, not only is it desirable to maximize the relevance of the variables with the class, but it is also desirable to minimize the redundancy between them. To that end, the minimum redundancy criterion is defined as:

$$minR(S), R = \frac{1}{|S|^2} \sum_{v_i \in S} I(v_i; v_j) \quad (9)$$

The minimal redundancy maximal relevance (mRMR) algorithm combines Equations (8) and (9) according to the following expression:

$$\begin{array}{c} max\Psi(D,R), \Psi = D - R; (MID) \\ max\Omega(D,R), \Omega = D/R; (MIQ) \end{array} \quad (10)$$

MID stands for Mutual Information Different criteria, and the other possible implementation for the mRMR algorithm is the Mutual Information Quotient (MIQ). In this contribution, MID is used.

The authors of the algorithm [49] suggest an iterative method to find a set of variables close to the optimum since, in practice, it is unfeasible to try to evaluate the Equation (10) for all possible subsets of m variables in the original space. Assuming the suboptimal set S_{m-1} of $m-1$ variables to be found, the following variable to be added will be the one that verifies:

$$\max_{v_j \in X/S_{m-1}} \left\{ I(v_j; c) - \frac{1}{m-1} \sum_{v_i \in S_{m-1}} I(v_j; v_i) \right\} \quad (11)$$

In the first iteration of the algorithm, the redundancy term is null, so the first variable selected by mRMR is the same as the one chosen by MaxDep (the one that has the maximum relevance with the class).

The 3D-DWT transform produces a large number of coefficients (thousands), making it necessary to select them using mRMR (the 30 most relevant are selected). It should be kept in mind that reducing the number of coefficients for each region of interest reduces the computational complexity and memory size of the matrix of coefficients used. It should be noted that since there are 550 MRI images for the training phase, for each image, 1728 possible VOIs are obtained, each VOI has 30 coefficients and both whole and gray matter are treated, the total number of coefficients that it is necessary to store is 57,024,000.

3.5. Optimization of the VOIs Selection Using Multi-Objective Optimization Evolutionary Algorithm (MOE)

In this contribution, a Multi-Objective Optimization Evolutionary Algorithm (MOE) based on a controlled elitist algorithm variant of NSGA-II is used to select the most relevant region of the brain (VOIS), taking into account two objective functions: The classification accuracy and the complexity of the classifier (measure as the number of VOIs required for the system).

This optimization tool is based on Darwin's survival of the fittest principle, which states that an initial population of individuals evolves through natural selection. Therefore, a set of solution candidates is retained in MOE, performing selection processes (fitness assignment and sampling are performed on these potential solutions), and finally, several potential solutions can be mixed (in terms of recombination operators) to produce new solutions. In resemblance to Darwin's theory on survival (from an initial population of individuals, and due to a process of natural selection, the population evolves towards better individuals in terms of the fitness function used), the potential solutions are called individuals, and the set of solutions is called the population.

In this contribution, a binary coding of the individuals (the solutions) that make up the EMO population has been used. In the terminology of evolutionary algorithms, each individual is represented by a chromosome, which is made up of different genes. The genes in our case represent the VOIs used, indicated by a 1 if the VOI is selected and by a 0 when that VOI is not used.

Therefore, the number of brain regions that are selected in a solution of the Multi-Objective Optimization Evolutionary Algorithm is equal to the number of genes that have a value of 1. In our case, as we worked with whole matter and gray matter, the total number of genes that can be activated in a solution of the MOE (equal to the number of VOIs that can be selected for the two volumetric images of the brain) is 3456. The upper part of the Figure 6 shows an example of codification for an individual-j of the population

of the evolutionary algorithm. The total size of the population is 250 individuals in the simulations carried out, with a coding example of the entire population shown in the lower part of Figure 6.

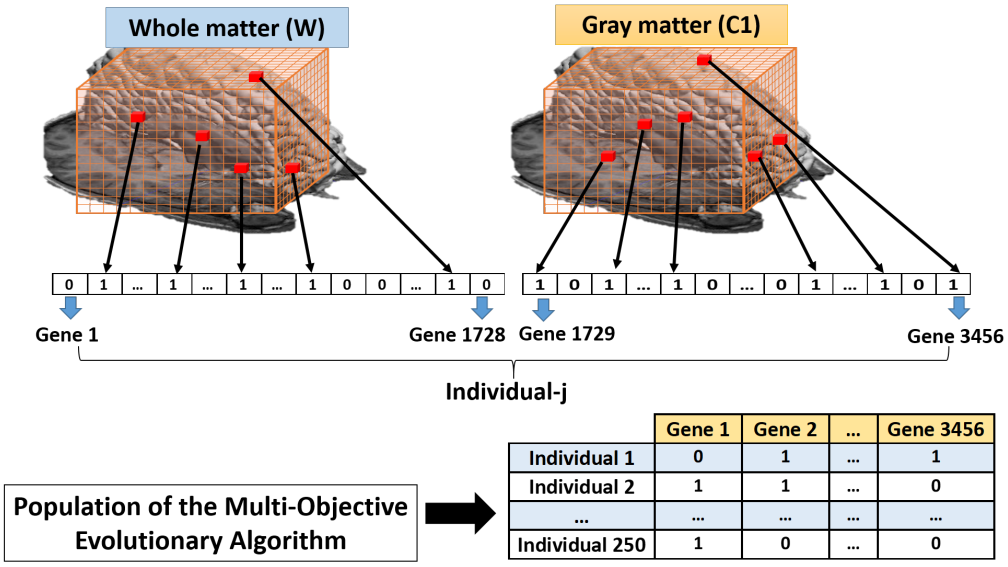


Figure 6. Binary coding used by the MOE algorithm, with the aim of selecting the most relevant VOI of the brain (both in gray and complete matter), and these regions being used for the construction of an accurate multi-class classifier.

The objective of this research is twofold. The first objective is to deepen the knowledge of which the most relevant regions of the brain are (taking into account both gray and whole matter) that are related to Parkinson's disease and the different types of patients analyzed in this contribution. Obtaining these VOIs can be of great relevance for expert doctors and neurologists and specialists in Parkinson's disease. The second objective is the realization of an accurate multi-class classifier, whose inputs are the 3D-DWT coefficients of the brain regions that have been selected as most relevant. For the construction of the classifier, in this paper, a multi-class classification system based on SVM will be used.The algorithm has the following steps:

1. Initialization of the Multi-Objective Optimization Evolutionary Algorithm (MOE), through an initial population P_0, and the creation of an initial (empty) set of non-dominated solutions (ND_0). The population size that was used in the simulation was 250.
2. Calculate the fitness value for each individual. For this calculation, using the set of training patients, the mean precision value (10-k-fold is used) of the SVM classifier used in the MOE fitness function must be computed, together with the number of genes (complexity of the system) that are set to one.
3. In the NSGA-II algorithm, there are two components that are relevant in its elitist strategy: sorting the non-dominated solutions and calculating the crowding distance. With this methodology, individuals from the population will be selected based on their fitness function.
4. The evolutionary algorithm generates a new population of individuals. To achieve this, a copy of the non-dominated individuals in P_t and ND_t is made to ND_{t+1}
5. During the execution of a new generation, operators of the evolutionary algorithm are applied to the individuals of the population. These include the binary tournament selection with replacement in ND_{t+1}. As we have indicated before, there is an elitist process where the best solutions are copied from generation to generation, and MOE operators complete the population (until the 250 individuals used in this role are

obtained). Binary crossover and mutation are also applied, and P_{t+1} is set on the resulting population.

6. Evaluation of fitness and Pareto functions from the construction. Our fitness function has two goals: Complexity of the obtained solution (measured as the number of VOIs selected) and accuracy of the classifier (using an SVM).
7. Repeat until Stopping Condition is reached (typically, a maximum number of generations is reached).

The final solution of the multi-objective optimization is the so-called Pareto Front, which is a set of non-dominated solutions, chosen as being optimal, in which the two objectives were simultaneously optimized: the accuracy of the classifier (using SVM to perform a classification between SWEDD, PD, Prodromal, GenCohort and normal subjects) and the complexity of the system (measured as the number of needed VOIs).

Figure 7 shows an example of coding for the solutions of the Pareto frontier. Assuming that there are M non-dominated solutions, each of the solutions have both a different number of selected VOIs (genes in our case), as well as a different accuracy. For each VOI, the number of features used is called N and is extracted using 3D-DWT. As can be seen in the bottom part of Figure 7, for a given patient K', using the individual–j of the Pareto Front, only a subset of the VOIs is used, each having N features. Therefore, each patient K' can be represented by a matrix of real numbers such as the one presented in Figure 7. This process must be carried out on all the patients in the database, as illustrated in Figure 8.

As previously mentioned, the number of representative wavelet coefficients for each VOI is reduced thanks to the use of the mRMR algorithm. However, when using a high number of VOIs, the number of coefficients that are input to the SVM classifier is high. For this reason, a feature reduction is performed by means of PCA, reducing the dimensionality of the input space for the classifier of the fitness function. Therefore, for each patient, the matrix of real numbers is simplified, as represented in Figure 9.

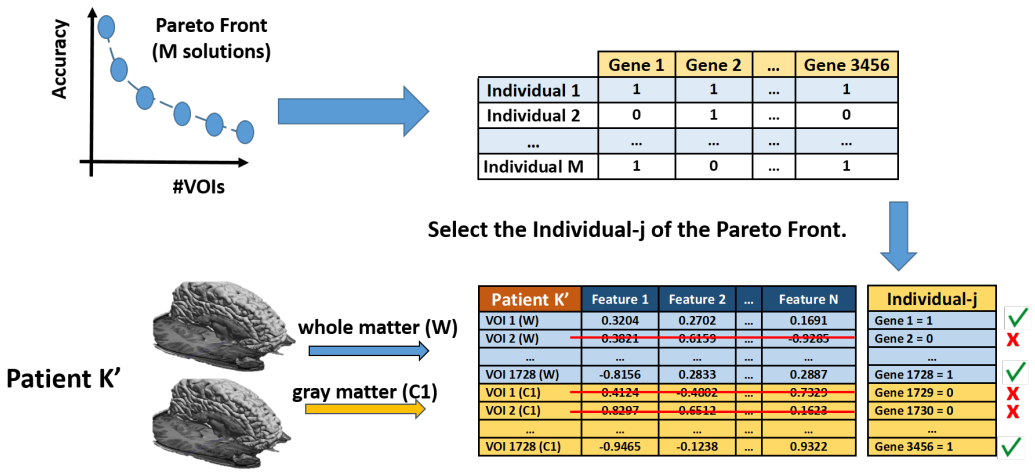


Figure 7. Binary coding used by the MOE algorithm for the selection of the most relevant VOIs in gray and whole matter.

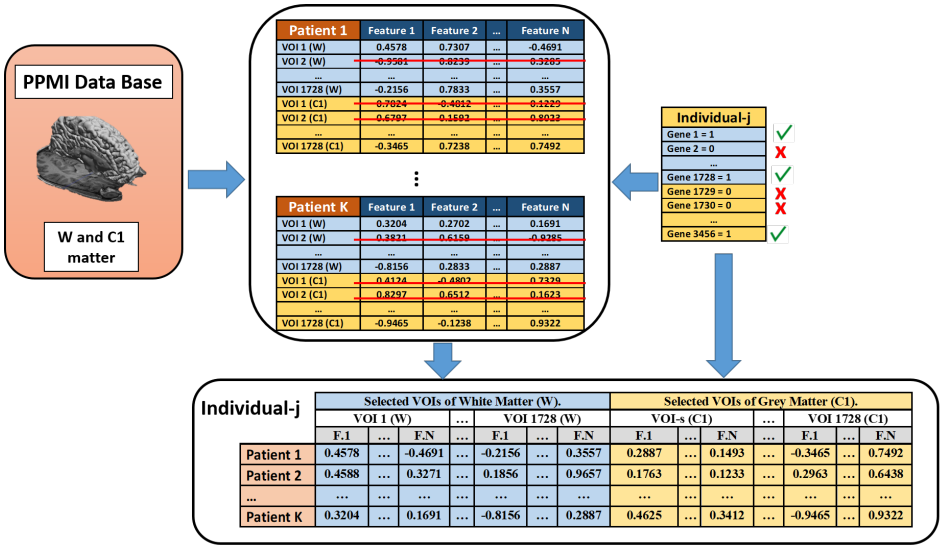


Figure 8. Matrix of coefficients of the 3D-DWT for the different VOIs selected in the individual-j of the Pareto Front.

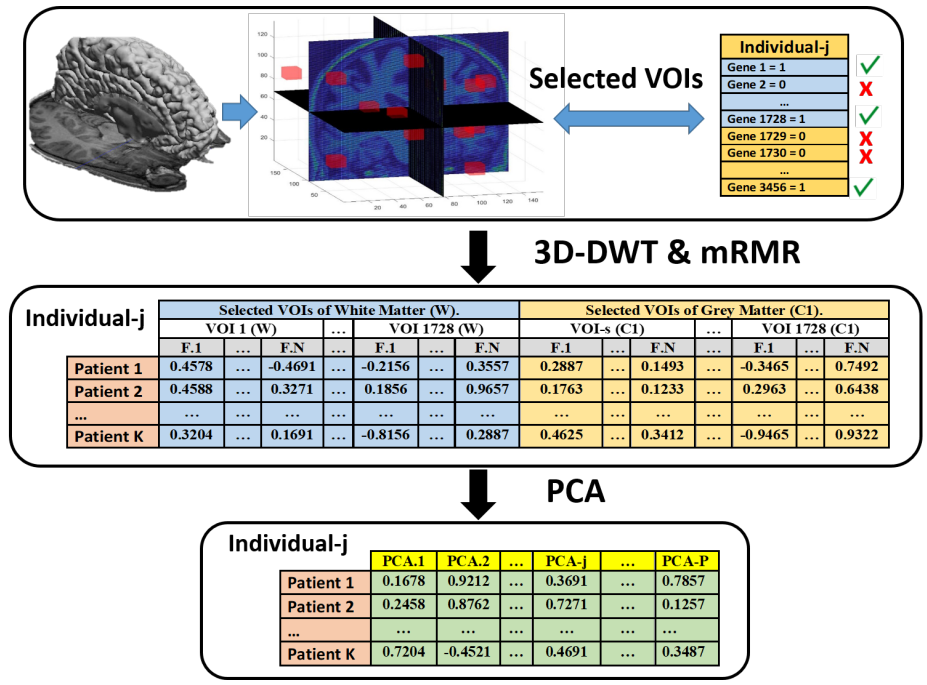


Figure 9. Block diagram of the PCA component matrix used for reducing the number of features for each patient, using a specific individual-j of the Pareto Front.

4. Pipeline

MRIs from the PPIM database were divided into two different groups with no intersection. The first group is made up of the training set, with a total of 550 MRIs, being Control (n = 155), SWEDD (n = 92), Prodromal (n = 23), PD (n = 270) and GeneCohort (n = 10). For the training set, 367 different MRI are selected, their composition being: Control (n = 103), SWEDD (n = 62), Prodromal (n = 15), PD (n = 180) and GeneCohort (n = 7).

All the MRIs used were previously normalized to the MNI space to find the best brain areas, dividing the brain into 1728 different VOIs, with a dimension of $12 \times 14 \times 10$ voxels. Once the MRIs are normalized, different filters are used to select gray matter (C1) and whole matter (W). These VOIs are characterized by the wavelet coefficients of the 3D-DWT transform using the biorthogonal function 3.3 as a reference wavelet up to level 2.

The number of coefficients obtained in the 3D-DWT transform is very high (thousand), requiring a reduction and selecting the most relevant ones (the 30 best coefficients are selected) with the mRMR algorithm.

A further step to reduce the number of coefficients is the use of PCA when all the coefficients of the selected VOIs are merged using the multi-objective optimization evolutionary algorithm. The coefficients obtained through PCA will be the input to the SVM classifier.

As indicated, the methodology presented has two different phases: obtaining the most relevant VOIs (optimization phase with the 550 training MRIs) and a test phase of the obtained SVM classifier (using the different solutions of the Pareto Frontier and the 367 MRI scan).

In the first optimization phase, each individual of the population of the MOE (i.e., potential solutions, called chromosomes) is represented by a binary code (with a length of 1728 genes, equal to the number of VOIs, for each matter W and C1, Figure 6). If the value of a gene is one, it means that the VOI is selected, and zero therefore means that the VOI was not used for classification.

The parameters used in the MOE are a population of 250 individuals and 100 generations in an optimization process based on NSGA-II. The number of individuals and generations has not been selected excessively large, since the computational complexity and the execution time increase exponentially in this type of task.

To obtain greater robustness in the optimization phase, a 10-fold cross-validation has been performed in the evaluation of the classification of the SVM classifier. The output of the MOE is the Pareto front (Figure 10), where various solutions (solutions with different numbers of VOIs for performing the classification) are obtained. In this contribution, various solutions obtained by the MOE algorithm have been analyzed, performing the test phase with the set of 367 MRI.

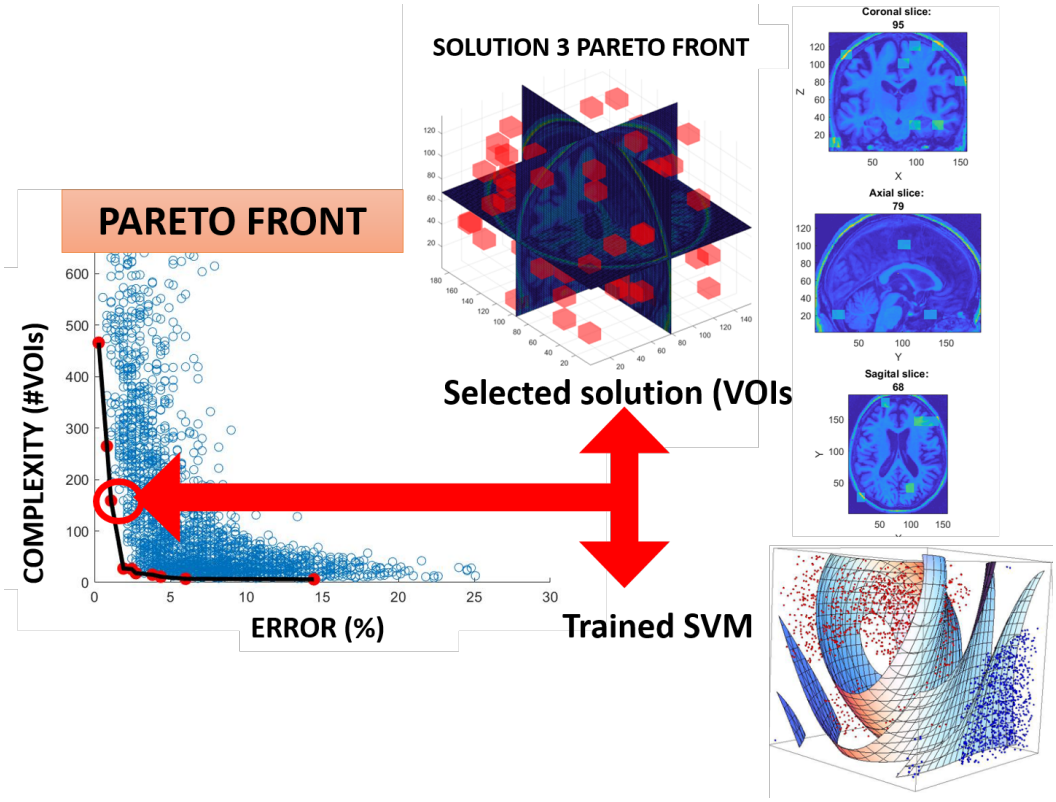


Figure 10. Non-dominated solutions from the Pareto Front, representing a sets of selected VOIs and its corresponding SVM classifier.

5. Results

5.1. Performance

In bi-class problems, the sensitivity and specificity are often used to analyze the behavior of the classifier. However, the problem addressed in this contribution is more complex, using five different classes. In multi-class problems, the confusion matrix is frequently used for the analysis of the behavior of the classifier. The confusion matrix for a problem of n class is an nxn matrix in which the columns represent the actual classes and the rows the output classes predicted by the model. The main diagonal contains the sum of all correct predictions.

5.2. Classification Results

Once the optimization process had ended, it is important to analyze the non-dominated solutions obtained from the Pareto front, observing their complexity and the accuracy obtained (Table 4). The first subject of the Pareto front is the individual who achieves the best accuracy. However, it is also the most complex and contains the most VOIs. This solution has 466 VOIs, taking into account both whole and gray sources (W + C1 sources).

Table 4. Pareto front for VOI optimization. Training phase.

Pareto Solution	Accuracy (%)	Total VOIs	W-VOIs	C1-VOIs
1	99.73	466	231	235
2	99.18	265	133	132
3	98.91	159	73	86
4	98.09	27	6	21
5	97.54	26	4	22
6	97.27	18	6	12
7	96.18	15	4	11
8	95.64	11	4	7
9	94.00	7	0	7
10	85.55	6	0	6

As can be seen from Table 4, for Pareto front solutions with a higher number of VOIs, the precision is high (It is above 98%, bearing in mind that it is a measure of accuracy obtained with training data during the first optimization phase), the ratio of the number of VOIs being similar for both gray matter and whole matter. As the number of selected VOIs is smaller, the precision decreases and most of the VOIs selected by MOE are in the volumetric information of the brain belonging to the gray matter.

This statement can be corroborated thanks to Figure 11, which shows the Pareto Genome, that is, the selected VOIs of the brain are represented on the x-axis (both for gray matter and for whole matter), and the y-axis represents the precision of the MOE solution with these selected VOIs. For solutions with low complexity, most of the sources are in the gray matter part. Remember that the brain is divided into $12 \times 12 \times 12$ partitions, with VOIs from 1 to 1728 coming from whole matter and from 1729 to 3456 from gray matter images.

Figure 12 shows the complexity (measured as the number of VOIs), together with the error obtained for the classification in the training stage, for all the solution populations of the MOE (zoomed to 700 VOIs solutions). The last generation solutions (filled blue circles) and Pareto front solutions (filled red circles) are also included in the figure.

It should be noted that the results presented in Table 4 come from the VOI optimization phase (Figure 1), and therefore, the precision results are related to the training phase. These solutions have been obtained using the MOE algorithm, within which there is an evaluation stage of the solutions selected, using an SVM classifier. This SVM classifier uses 550 training MR images.

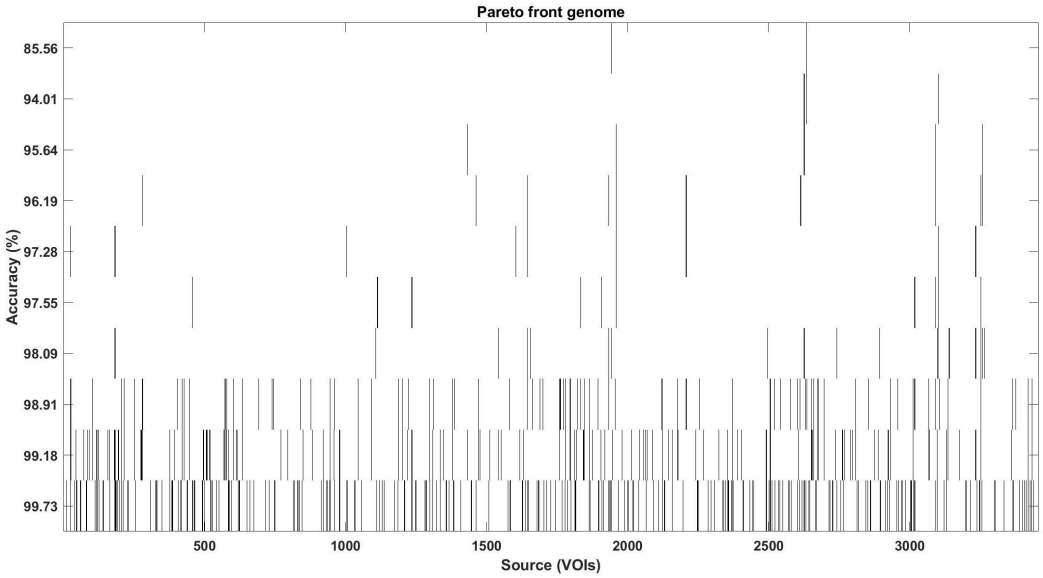


Figure 11. Pareto front genome, where the VOIs that are most relevant to obtain a high accuracy in the classifier, are highlighted.

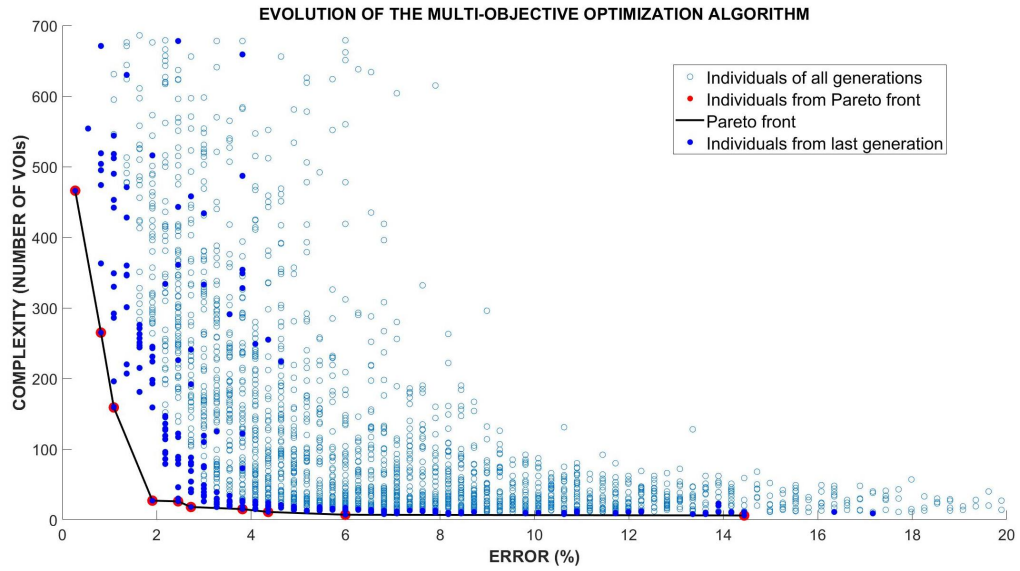


Figure 12. Evolution of all populations of individuals of the MOE algorithm (including the last generation and the Pareto front).

To analyze the robustness of the proposed methodology, it is necessary to carry out the second phase, consisting of performing the classification with test data, using the solutions proposed in the Pareto front. The Pareto front solutions obtained in the first optimization phase (VOI selection) were tested on the test images, where 367 new MRI images are classified with an SVM, the coefficients of the selected VOIs being the input of the classifier.

It is relevant to emphasize that the MRIs used during the first phase (optimization phase for VOI selection) are different and independent from the MRIs used in the testing phase. Carrying out the test phase guarantees the solidity and robustness of the results presented by the proposed methodology. Different Pareto front solutions and configurations regarding the selection of the VOIs selected (gray matter, whole matter or both) were tested, showing in Table 5 some of the most relevant results.

Table 5. Test results for several solutions of the Pareto front. Details of the number of VOIs, number of features and components of the PCA for W,C1 and W+C1 sources.

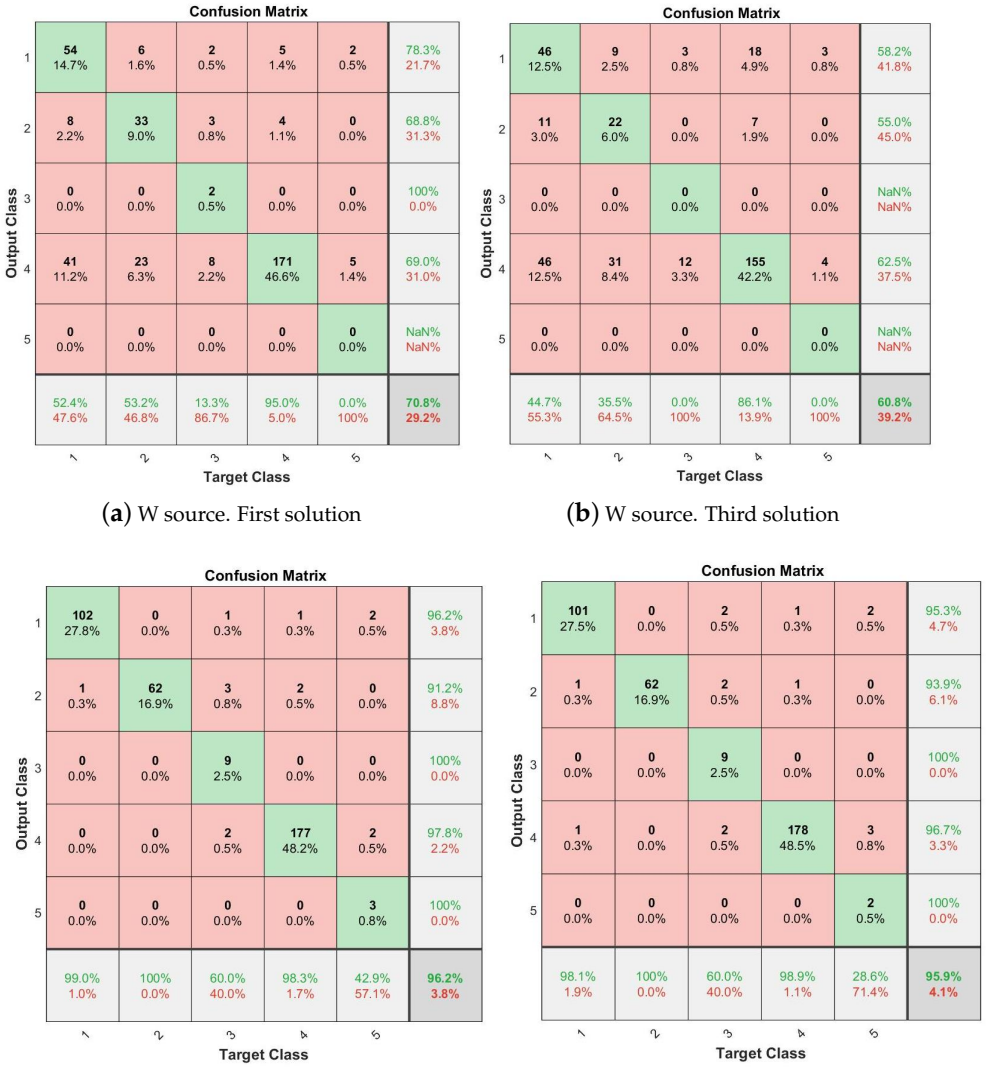
Solution (PF)	Source	#VOIs	#Features	#PCA	Accuracy	Conf.Matrix
1	W	231	6930	511	70.8	Figure 13a
	C1	235	7050	464	96.2	Figure 13c
	W+C1	466	13,980	535	97.0	
2	W	133	3990	419	63.8	
	C1	132	3960	416	95.9	
	W+C1	265	7950	498	96.2	
3	W	73	2190	339	60.8	Figure 13b
	C1	86	2580	380	95.9	Figure 13d
	W+C1	159	4770	448	96	
4	W	6	180	32	48.0	
	C1	21	630	224	93.7	
	WC1	27	810	232	93.5	
5	W	4	120	37	49.86	
	C1	22	660	221	92.9	
	WC1	26	780	230	93.18	
7	W	4	120	3	49.0	
	C1	11	330	136	88.8	
	WC1	15	450	138	90.5	

As it was described in the first optimization phase, due to the high number of coefficients, PCA is used to keep around 95% of the features variance, significantly reducing the final dimension of the number of input features to the SVM classifier. The confusion matrices of the different configurations analyzed for the first and third Pareto front solutions are presented in Figure 13 (for W, C1 and W+C1 sources used for VOIs selection).

The results obtained using an independent cohort of MRI subjects for the testing phase have yielded promising results, taking into account that the problem is complex because it has five different classes, with an imbalance between them. For the first Pareto Front solution using both whole matter and gray matter, the accuracy obtained in the confusion table is 97%, requiring 466 VOIs. Analyzing the third Pareto solution, focusing on the gray matter, using 86 VOIs reaches 95.9 % precision. Note that the brain regions used in whole matter may be different from those used in gray matter.

Table 6 shows the locations of the 73 and 86 VOIs selected by the third solution of Pareto Front, for whole and gray matter, respectively. Graphically, Figures 14 and 15 show the location in three-dimensional space of these VOIs in the brain. In the left part of the figures, the spatial location of the selected VOIs is shown. On the right side, and in order to facilitate its location in slices, three slices of the brain and the location of the VOIs are presented.

It is relevant to make a comparison of the PD disease classification results presented in the bibliography, with those obtained in the present methodology (although it should be noted that the number of classes in this paper is higher than those commonly presented in publications). Table 7 presents a summary of relevant contributions that, based on MRI images, have used machine learning and artificial intelligent techniques for the automatic classification of PD.



(**a**) W source. First solution (**b**) W source. Third solution

(**c**) C1 source. First solution (**d**) C1 source. Third solution

Figure 13. Confusion matrices obtained in the testing phase, for the first (**left**) and third (**right**) solution of the Pareto Front, using W and C1 sources.

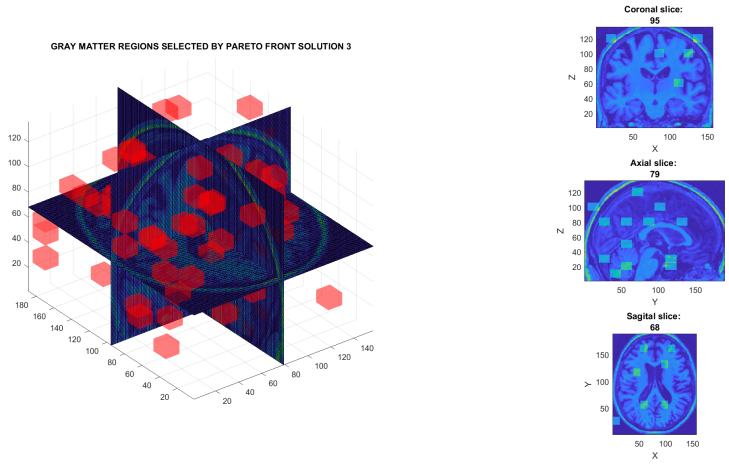


Figure 14. Selected brain regions from the third Pareto solution (whole matter source, W).

Table 6. Description of the spatial location (each VOI being described by its six integers [xmin,xmax,ymin,ymax,zmin,zmax]) for the third solution obtained by the MOE algorithm, for both W and C1 matter.

Source	Description of VOI Locations
W	[1,13,35,49,16,26]; [1,13,35,49,36,46]; [1,13,95,109,6,16]; [1,13,110,124,86,96]; [1,13,110,124,106,116]; [1,13,125,139,76,86]; [1,13,155,169,46,56]; [14,26,20,34,6,16]; [14,26,20,34,66,76]; [14,26,50,64,46,56]; [14,26,80,94,36,46]; [14,26,80,94,116,126]; [14,26,95,109,106,116]; [14,26,125,139,116,126]; [14,26,170,184,36,46]; [14,26,170,184,56,66]; [27,39,5,19,86,96]; [27,39,140,154,76,86]; [27,39,170,184,16,26]; [27,39,170,184,76,86]; [40,52,20,34,36,46]; [40,52,170,184,86,96]; [40,52,170,184,96,106]; [53,65,5,19,36,46]; [53,65,5,19,86,96]; [53,65,35,49,36,46]; [53,65,65,79,116,126]; [53,65,110,124,6,16]; [53,65,125,139,46,56]; [53,65,140,154,96,106]; [53,65,155,169,6,16]; [53,65,170,184,66,76]; [53,65,170,184,106,116]; [66,78,20,34,76,86]; [66,78,35,49,16,26]; [66,78,80,94,86,96]; [66,78,155,169,6,16]; [66,78,155,169,16,26]; [79,91,20,34,16,26]; [79,91,95,109,96,106]; [79,91,125,139,16,26]; [92,104,35,49,66,76]; [92,104,50,64,16,26]; [92,104,65,79,6,16]; [92,104,65,79,86,96]; [92,104,95,109,26,36]; [92,104,95,109,116,126]; [92,104,125,139,46,56]; [105,117,35,49,116,126]; [105,117,50,64,106,116]; [105,117,65,79,16,26]; [105,117,80,94,116,126]; [105,117,125,139,46,56]; [105,117,140,154,66,76]; [118,130,5,19,16,26]; [118,130,5,19,46,56]; [118,130,20,34,36,46]; [118,130,80,94,86,96]; [118,130,95,109,26,36]; [118,130,95,109,116,126]; [118,130,110,124,56,66]; [118,130,125,139,86,96]; [118,130,140,154,26,36]; [118,130,140,154,66,76]; [118,130,155,169,26,36]; [131,143,35,49,76,86]; [131,143,140,154,66,76]; [131,143,170,184,96,106]; [144,156,65,79,116,126]; [144,156,95,109,76,86]; [144,156,125,139,96,106]; [144,156,140,154,76,86]; [144,156,155,169,106,116]
C1	[1,13,5,19,86,96]; [1,13,20,34,66,76]; [1,13,35,49,76,86]; [1,13,35,49,96,106]; [1,13,50,64,96,106]; [1,13,65,79,36,46]; [1,13,80,94,76,86]; [1,13,80,94,96,106]; [1,13,110,124,96,106]; [1,13,125,139,96,106]; [1,13,140,154,96,106]; [1,13,170,184,26,36]; [1,13,170,184,46,56]; [1,13,170,184,56,66]; [14,26,20,34,116,126]; [14,26,50,64,6,16]; [14,26,65,79,86,96]; [14,26,80,94,16,26]; [14,26,95,109,116,126]; [27,39,5,19,106,116]; [27,39,125,139,76,86]; [27,39,125,139,86,96]; [27,39,125,139,116,126]; [40,52,20,34,56,66]; [40,52,50,64,116,126]; [40,52,80,94,6,16]; [40,52,110,124,6,16]; [40,52,110,124,66,76]; [40,52,110,124,116,126]; [40,52,155,169,6,16]; [53,65,35,49,26,36]; [53,65,50,64,66,76]; [53,65,50,64,106,116]; [53,65,80,94,76,86]; [53,65,110,124,46,56]; [53,65,125,139,46,56]; [53,65,140,154,106,116]; [53,65,155,169,66,76]; [66,78,20,34,46,56]; [66,78,20,34,106,116]; [66,78,65,79,96,106]; [66,78,80,94,116,126]; [66,78,110,124,96,106]; [66,78,155,169,96,106]; [79,91,5,19,96,106]; [79,91,20,34,26,36]; [79,91,20,34,76,86]; [79,91,35,49,6,16]; [79,91,50,64,16,26]; [79,91,50,64,46,56]; [79,91,50,64,76,86]; [79,91,65,79,116,126]; [79,91,80,94,76,86]; [79,91,95,109,96,106]; [79,91,110,124,16,26]; [79,91,110,124,26,36]; [79,91,125,139,76,86]; [92,104,50,64,66,76]; [92,104,65,79,86,96]; [92,104,80,94,116,126]; [92,104,125,139,66,76]; [92,104,140,154,86,96]; [92,104,140,154,96,106]; [105,117,65,79,26,36]; [105,117,65,79,36,46]; [105,117,65,79,86,96]; [105,117,65,79,106,116]; [105,117,80,94,46,56]; [105,117,95,109,56,66]; [105,117,155,169,66,76]; [105,117,155,169,116,126]; [105,117,170,184,56,66]; [105,117,170,184,106,116]; [118,130,50,64,76,86]; [118,130,65,79,86,96]; [118,130,80,94,76,86]; [118,130,95,109,96,106]; [118,130,140,154,36,46]; [131,143,80,94,86,96]; [131,143,95,109,116,126]; [144,156,35,49,6,16]; [144,156,65,79,36,46]; [144,156,65,79,46,56]; [144,156,80,94,36,46]; [144,156,125,139,116,126]; [144,156,155,169,16,26]

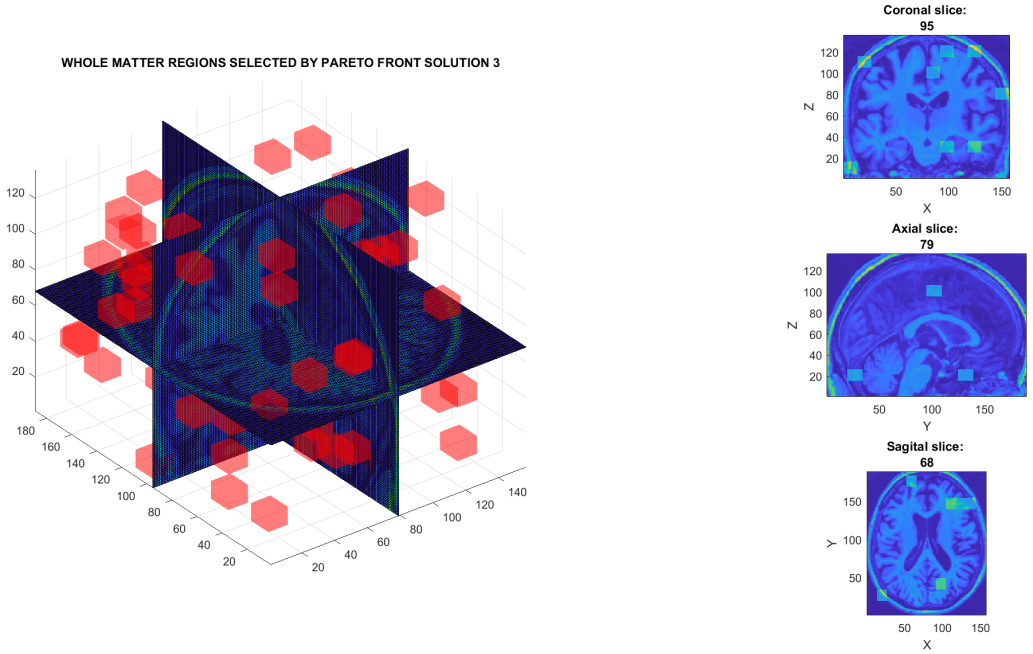


Figure 15. Selected brain regions from the third Pareto solution (gray matter source, C1).

Table 7. Comparative summary of different methodologies presented in the bibliography to predict PD from MRI.

Author, Year	Cohort Data Base	Number of Subjects	Method Employed	Accuracy (%)
Long et al., 2012 [23]	Zhejiang University.	PD (n = 19) vs. Normal (n = 26)	SVM, LOOCV	86.9
Salvatore et al., 2014 [12]	Own data-base.	PD (n = 28) vs. Normal (n = 28)	PCA, SVM, VBM	83.2
Rana et al., 2015 [22]	Own data-base.	PD (n = 30) vs. Normal (n = 30)	SVM, ROI	86.67
Adeli et al., 2016 [50]	PPMI.	PD (n = 374) vs. Normal (n = 169)	Joint feature-sample selection, RLDA	81.9
Abos et al., 2017 [51]	Own data-base.	PD (n = 27) vs. Normal (n = 38)	SVM, Functional connectome	80.0
Lei et al., 2018 [24]	PPMI.	PD (n = 123) vs. SWEDD (n = 29) vs. Normal (n = 56)	SVM, GDC,CSF, DSSM	78.37
Amoroso et al., 2018 [18]	PPMI.	PD (n = 378) vs. Normal (n = 169)	Random Forest, SVM	93
Oliveira et al., 2018 [25]	PPMI. FP-CIT SPECT	PD (n = 443) vs. Normal (n = 209)	SV, LOOCV, ROI, k-NN	97.9
Ariz et al., 2019 [52]	UK Brain Bank.	PD (n = 39) vs. Normal (n = 40)	Artificial Neural Network	79.9
Shinde et al., 2019 [19]	NIMHANS, India.	PD (n = 45) vs. Normal (n = 35)	CNNs	80.00
Pang et al., 2020 [20]	China Medical University.	IPD (n = 83) vs. MSA-P (n = 102)	Seven radiomic features, SVM	86.2
Zhang et al., 2020 [53]	PPMI.	PD (n = 201) vs. Prodromal (n = 193) vs. Normal (n = 204)	Semi-supervised, LDA Multi-view learning Clustering	74
Park et al., 2020 [54]	Yonsei Univ. HealthSystem.	PD (n = 77) vs. Normal (n = 53)	SVM, Vol + ReHo + DegCen	75.3
Pereira et al., 2020 [26]	PPMI.	(a) PD (n = 378) vs. SWEDD (n = 58) (b) PD (n = 378) vs. Normal (n = 168)	CNNs	(a) 73.3 (b) 97.4
Solana et al., 2021 [55]	PPMI	PD (n = 348) vs. Normal (n = 159)	Seven classifiers Haralick and Wrapper FS	93
Talai et al., 2021 [56]	U.Medical Center Hamburg.	PD (n = 45) vs. PSP-RS (n = 20) vs. Normal (n = 159)	Regional morphology features quantitative T2 values diffusion-tensor	95.1
Proposed method Solution 1. C1 matter Figure 13c	PPMI.	PD (n = 450) vs. Normal (n = 258) vs. SWEDD (n = 154) vs. Prodromal (n = 38) vs. GeneCohort (n = 17)	3D-DWT, MOE, SVM	96.2
Proposed method Solution 3. C1 matter Figure 13d	PPMI.	PD (n = 450) vs. Normal (n = 258) vs. vs. SWEDD (n = 154) vs. Prodromal (n = 38) vs. GeneCohort (n = 17)	3D-DWT, MOE, SVM	95.9

5.3. Conclusions

This study has presented a novel procedure for the multiple-class classification in the field of Parkinson's disease using MRI and machine learning paradigms. The novel presented methodology has two main objectives. First, obtaining the brain regions that are most relevant to the detection and classification of Parkinson's disease is of great relevance to expert neurologists. Obtaining the most relevant VOIs was accomplished due to the implementation of an optimization algorithm based on Multi-Objective Optimization Evolutionary Algorithm using an SVM classifier. The second objective of this contribution is to offer a set of solutions, with different complexity (measured as the number of VOIs) and precision for the automatic classification of patient of various classes: PD, SWEDD, Prodromal, GeneCohort and healthy patients, using a large international cohort from the PPMI database. The proposed methodology is therefore composed of two phases: optimization of the most relevant VOIs and the classifier test with an independent set of MRIs.

Both the whole matter images (dividing the brain into 1728 VOIs) and the gray matter segmented images (again with 1728 VOIs) have been used. In the first phase, the optimization phase, there are therefore 3456 different volumes of interest, extracted from 550 MRI images from different subjects of the 5 categories analyzed in this paper. The optimization was accomplished using an MOE, obtaining a diverse set of solutions with different complexities (measured with the number of VOIs) and degrees of precision (giving rise to the so-called Pareto front).

In the second phase, the testing phase, some of the solutions that provide the Pareto front were tested with a cohort of independent subjects consisting of 367 MRIs of different subjects from the five categories. The most precise, but most complex solution (we could

call it the first solution), has a total of 466 volumes of interest (231 for W matter and 235 for C1 matter). With this large number of VOIs, the classifier in the test phase obtained high accuracy: 97.0 % (70.8 % and 96.2 % using only the W and C1 matter, respectively).

Another solution (the third of the Pareto front) with less complexity was analyzed. This solution selects 159 regions (73 for W matter and 86 for C1 matter). Using the 86 volumes of gray matter, the classifier, in the test phase, achieved an accuracy of 95.9 % in a multi-class problem.

From the different solutions obtained in the Pareto Front, the high number of VOIs selected in the gray matter for the construction of the classifier can be obtained as a final conclusion.

To the best of our knowledge, there are very few, if not almost none, references in the literature that perform 3-D feature extraction for VOI optimization in a multi-class classification related to Parkinson's disease. The results from the confusion matrices for the different solutions of the analyzed MOE confirm that feature extraction using 3D-DWT, together with feature reduction using mRMR and PCA, are powerful tools for characterizing brain regions. The MOE algorithm has obtained a repertoire of solutions with different degrees of precision and complexity, providing information on which are the most relevant brain regions for the construction of the SVM classifier.

Author Contributions: Conceptualization, I.R.-V., O.V., E.D.-M., F.R.; methodology, I.R.-V., O.V., E.D.-M., F.R.; software, I.R.-V., F.R.; validation, I.R.-V., O.V., E.D.-M.; formal analysis, I.R.-V., O.V., F.R.; investigation, I.R.-V., E.D.-M., F.R.; resources, O.V., F.R.; writing—original draft preparation, I.R.-V., O.V., E.D.-M., F.R.; writing—review and editing, O.V., F.R.; supervision, I.R.-V., O.V.; project administration, O.V., F.R.; funding acquisition, O.V., F.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Spanish Ministry of Sciences, Innovation and Universities under Project RTI-2018-101674-B-I00 and the projects from Junta de Andalucia B-TIC-414, A-TIC-530-UGR20 and P20-00163.

Data Availability Statement: Data Availability. LONI Image Data Archive (IDA): https://ida.loni.usc.edu/.

Acknowledgments: We would like to acknowledge the support of the Michael J. Fox Foundation for allowing access to available data from the PPMI.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Gouda, N.A.; Elkamhawy, A.; Cho, J. Emerging Therapeutic Strategies for Parkinson's Disease and Future Prospects: A 2021 Update. *Biomedicines* **2022**, *10*, 371. [CrossRef] [PubMed]
2. World Health Organization *Neurological Disorders: Public Health Challenges*; World Health Organization: Genebra, Switzerland, 2006; pp. 140–150.
3. Olanow, C.; Schapira, A.; Obeso, J. Parkinson's disease and other movement disorders. In *Harrison's Principles of Internal Medicine*, 19th ed.; McGraw-Hill Education: New York, NY, USA, 2015; pp. 2609–2626.
4. Dorsey, E.; Shererb, T.; Okunc, M.; Bloem, B. The Emerging Evidence of the Parkinson Pandemic. *J. Park. Dis.* **2018**, *8*, 140–150. [CrossRef] [PubMed]
5. Boshkoska, B.; Miljković, D.; Valmarska, A.; Gatsios, D.; Rigas, G.; Konitsiotis, S.; Tsiouris, K.; Fotiadis, D.; Bohanec, M. Decision Support for Medication Change of Parkinson's Disease Patients. *Comput. Methods Programs Biomed.* **2020**, *196*, 105552. [CrossRef] [PubMed]
6. Hughes, A.; Daniel, S.; Ben-Shlomo, Y.; Lees, A. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain* **2002**, *125*, 861–870. [CrossRef] [PubMed]
7. Foundation, P. Stages of Parkinson's. 2020. Available online: https://www.parkinson.org/ (accessed on 9 March 2020).
8. Kalia, L.; Lang, A. Parkinson's disease. *Lancet* **2015**, 896–912. [CrossRef]
9. Chougar, L.; Faouzi, J.; Pyatigorskaya, N.; Yahia-Cherif, L.; Gaurav, R.; Biondetti, E.; Villotte, M.; Valabregue, R.; Corvol, J.C.; Brice, A.; et al. Automated Categorization of Parkinsonian Syndromes Using Magnetic Resonance Imaging in a Clinical Setting. *Mov. Disord.* **2021**, *36*, 460–470. [CrossRef] [PubMed]

10. Biase, L.; Santo, A.; Caminiti, M.; Liso, A.; Shah, S.; Ricci, L.; Lazzaro, V. Gait Analysis in Parkinson's Disease: An Overview of the Most Accurate Markers for Diagnosis and Symptoms Monitoring. *Sensors* **2020**, *12*, 3529. [CrossRef] [PubMed]
11. Saeed, U.; Compagnone, J.; Aviv, R.E.A. Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: Current and emerging concepts. *Transl. Neurodegener.* **2017**, *6*, 1–25. [CrossRef] [PubMed]
12. Salvatore, C.; Cerasa, A.; Castiglioni, I.; Gallivanone, F.; Augimeri, A.; Lopez, M.; Arabia, G.; Morelli, M.; Gilardi, M.; Quattrone, A. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. *J. Neurosci. Methods* **2014**, *222*, 230–237. [CrossRef]
13. Eidelberg, D. *Imaging in Parkinson's Disease*; Oxford University Press: New York, NY, USA, 2012; pp. 3–8.
14. Chien, C.Y.; Hsu, S.W.; Lee, T.L.; Sung, P.S.; Lin, C.C. Using Artificial Neural Network to Discriminate Parkinson's Disease from Other Parkinsonisms by Focusing on Putamen of Dopamine Transporter SPECT Images. *Biomedicines* **2020**, *9*, 12. [CrossRef]
15. Spetsieris, P.; Eidelberg, D. Spectral guided sparse inverse covariance estimation of metabolic networks in Parkinson's disease. *Neuroimage* **2021**, *226*, 117568. [CrossRef] [PubMed]
16. Halliday, G.M.; Leverenz, J.S.; Schneider, R.; Adler, C.H. The neurobiological basis of cognitive impairment in Parkinson's disease. *Mov. Disord.* **2014**, *29*, 634–650. [CrossRef]
17. Simuni, T.; Uribe, L.; Cho, H.R.; Caspell-Garcia, C.; Coffey, C.S.; Siderowf, A.; Trojanowski, J.Q.; Shaw, L.M.; Seibyl, J.; Singleton, A.; et al. Clinical and dopamine transporter imaging characteristics of non-manifest LRRK2 and GBA mutation carriers in the Parkinson's Progression Markers Initiative (PPMI): A cross-sectional study. *Lancet Neurol.* **2020**, *19*, 71–80. [CrossRef]
18. Amoroso, N.; La Rocca, M.; Monaco, A.; Bellotti, R.; Tangaro, S. Complex networks reveal early MRI markers of Parkinson's disease. *Med. Image Anal.* **2018**, *48*, 12–24. [CrossRef] [PubMed]
19. Shindea, S.; Prasadb, S.; Sabooa, Y.; Kaushicka, R.; Sainid, J.; Palb, P.; Ingalhalikara, M. Predictive markers for Parkinson's disease using deep neural nets on neuromelanin sensitive MRI. *Neuroimage Clin.* **2019**, *22*, 101748. [CrossRef] [PubMed]
20. Pang, H.; Yu, Z.; Li, R.; Yang, H.; Fan, G. MRI-Based Radiomics of Basal Nuclei in Differentiating Idiopathic Parkinson's Disease From Parkinsonian Variants of Multiple System Atrophy: A Susceptibility-Weighted Imaging Study. *Front. Aging Neurosci.* **2020**, *12*, 379. [CrossRef]
21. Olivares, R.; Munoz, R.; Soto, R.; Crawford, B.; Cárdenas, D.; Ponce, A.; Taramasco, C. An Optimized Brain-Based Algorithm for Classifying Parkinson's Disease. *Appl. Sci.* **2020**, *10*, 1827. [CrossRef]
22. Rana, B.; Juneja, A.; Saxena, M.; Gudwani, S.; Kumaran, S.S.; Agrawal, R.; Behari, M. Regions-of-interest based automated diagnosis of Parkinson's disease using T1-weighted MRI. *Expert Syst. Appl.* **2015**, *42*, 4506–4516. [CrossRef]
23. Long, D.; Wang, J.; Xuan, M.; Gu, Q.; Xu, X.; Kong, D.; Zhang, M. Automatic classification of early Parkinson's disease with multimodal MR Imaging. *PLoS ONE* **2012**, *7*, e47714.
24. Lei, H.; Zhao, Y.; Wen, Y.; Luo, Q.; Cai, Y.; Liu, G.; Lei, B. Sparse feature learning for multi-class Parkinson's disease classification. *Technol. Health Care* **2018**, *26*, 193–203. [CrossRef] [PubMed]
25. Oliveira, F.; Faria, D.; Costa, D.; Castelo-Branco, M.; Tavares, J. Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [123I]FP-CIT SPECT images. *Eur. Nucl. Med. Mol. Imaging* **2018**, *45*, 1052–1062. [CrossRef]
26. Pereira, H.; Ferreira, H. Classification of Patients with Parkinson's Disease Using Medical Imaging and Artificial Intelligence Algorithms. In *MEDICON 2019, IFMBE Proceedings 76*; Springer Nature: Cham, Switzerland, 2020; pp. 2043–2056.
27. Sivaranjini, S.; Sujatha, C. Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimed. Tools Appl.* **2020**, *79*, 15467–15479. [CrossRef]
28. Esmaeilzadeh, S.; Yang, Y.; Adeli, E. End-to-End Parkinson Disease Diagnosisusing Brain MR-Images by 3D-CNN. *arXiv* **2018**, arXiv:1806.05233.
29. Marek, K.; Jennings, D.; Lasch, S.; Siderowf, A.; Tanner, C.; Simuni, T.; Coffey, C.; Kieburtz, K.; Flagg, E.; Chowdhury, S.; et al. The parkinson progression marker initiative (PPMI). *Prog. Neurobiol.* **2011**, *95*, 629–635. [CrossRef] [PubMed]
30. Akkaoui, M.; Geoffroy, P.; Roze, E.; Degos, B.; Garcin, B. Functional Motor Symptoms in Parkinson's Disease and Functional Parkinsonism: A Systematic Review. *J. Neuropsychiatry Clin. Neurosci.* **2020**, *32*, 4–13. [CrossRef] [PubMed]
31. Lee, J.; Song, Y.; Kim, H.; Ku, B.; Lee, W. Patients with scans without evidence of dopaminergic deficit (SWEDD) do not have early Parkinson's disease: Analysis of the PPMI data. *PLoS ONE* **2021**, *16*, e0246881. [CrossRef] [PubMed]
32. Erro, R.; Schneider, S.; Stamelou, M. What do patients with scans without evidence of dopaminergic deficit (SWEDD) have? New evidence and continuing controversies. *J. Neurol. Neurosurg. Psychiatry* **2016**, *87*, 319–323. [CrossRef]
33. Choi, H.; Ha, S.; Im, H. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. *NeuroImage Clin.* **2017**, *16*, 586–594. [CrossRef]
34. Schneider, S. Patients with adult-onset dystonic tremor resembling Parkinsonian tremor have scans without evidence of dopaminergic deficit (SWEDDs). *Mov. Disord.* **2007**, *22*, 2210–2215. [CrossRef]
35. Nicastro, N.; Garibotto, V.; Badoud, S.; Burkhard, P. Scan without evidence of dopaminergic deficit: A 10-year retrospective study. *Park. Relat. Disord.* **2016**, *31*, 53–58. [CrossRef]
36. Salmanpour, M.R.; Shamsaei, M.; Saberi, A.; Hajianfar, G.; Soltanian-Zadeh, H.; Rahmim, A. Robust identification of Parkinson's disease subtypes using radiomics and hybrid machine learning. *Comput. Biol. Med.* **2021**, *129*, 104142. [CrossRef] [PubMed]
37. Mahlknecht, P.; Seppi, K.; Poewe, W. The Concept of Prodromal Parkinson's Disease. *J. Parkinsons Dis.* **2015**, *5*, 681–697. [CrossRef] [PubMed]

38. Postuma, R.; Aarsland, D.; Barone, P.; Burn, D.; Hawkes, C.; Oertel, W.; Ziemssen, T. Identifying prodromal Parkinson's disease: Pre-Motor disorders in Parkinson's disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* **2012**, *27*, 617–26. [CrossRef] [PubMed]
39. Penny, W.D.; Friston, K.J.; Ashburner, J.T.; Kiebel, S.J.; Nichols, T.E. (Eds.) *Statistical Parametric Mapping: The Analysis of Functional Brain Images*; Academic Press: Cambridge, MA, USA, 2006.
40. Flandin, G.; Friston, K. Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Hum. Brain Mapp.* **2017**, *40*, 2052–205. [CrossRef] [PubMed]
41. Ashburner, J.; Friston, K. Unified segmentation. *NeuroImage* **2005**, *26*, 839–851. [CrossRef] [PubMed]
42. Nanni, L.; Lumini, A. Wavelet decomposition tree selection for palm and face authentication. *Pattern Recognit. Lett.* **2008**, *29*, 343–353. [CrossRef]
43. Challis, E.; Hurley, P.; Serra, L.; Bozzali, M.; Oliver, S.; Cercignani, M. Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI. *NeuroImage* **2015**, *112*, 232–243. [CrossRef] [PubMed]
44. Haq, E.; Huang, J.; Kang, L.; Haq, H.; Zhan, T. Image-based state-of-the-art techniques for the identification and classification of brain diseases: A review. *Med. Biol. Eng. Comput.* **2020**, *58*, 2603–2620. [CrossRef]
45. Bharati, S.; Podder, P.; Al-Masud, M. Brain Magnetic Resonance Imaging Compression Using Daubechies Biorthogonal Wavelet with the fusion of STW and SPIHT. In Proceedings of the 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, 22–24 November 2018; pp. 1–4.
46. Kovalev, V.; Kruggel, F.; Gertz, H.; Von Cramon, D.Y. Three-dimensional texture analysis of MRI brain datasets. *IEEE Trans. Med. Imaging* **2001**, *20*, 424–433. [CrossRef]
47. Valenzuela, O.; Jiang, X.; Carrillo, A.; Rojas, I. Multi-Objective Genetic Algorithms to Find Most Relevant Volumes of the Brain Related to Alzheimer's Disease and Mild Cognitive Impairment. *Int. J. Neural Syst.* **2018**, *28*, 961–972. [CrossRef]
48. He, C.; Dong, J.; Zheng, Y.; Gao, Z. 3-D coefficient tree structure for 3-D wavelet video coding. *IEEE Trans. Circuits Syst. Video Technol.* **2003**, *13*, 961–972.
49. Peng, H.C.; Long, F.H.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]
50. Adeli, E.; Shi, F.; An, L.; Wee, C.Y.; Wu, G.; Wang, T.; Shen, D. Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data. *NeuroImage* **2016**, *141*, 206–219. [CrossRef] [PubMed]
51. Abos, A.; Baggio, H.C.; Segura, B.; Garcia-Diaz, A.I.; Compta, Y.; Martí, M.J. Discriminating cognitive status in parkinson's disease through functional connectomics and machine learning. *Sci. Rep.* **2017**, *7*, 45347. [CrossRef] [PubMed]
52. Ariz, M.; Abad, R.C.; Castellanos, G.; Martínez, M.; Muñoz-Barrutia, A.; Fernández-Seara, M.A.; Pastor, P.; Pastor, M.A.; Ortiz-de Solórzano, C. Dynamic Atlas-Based Segmentation and Quantification of Neuromelanin-Rich Brainstem Structures in Parkinson Disease. *IEEE Trans. Med. Imaging* **2019**, *38*, 813–823. [CrossRef] [PubMed]
53. Zhang, X.; Zhai, D.; Yang, Y.; Zhang, Y.; Wang, C. A novel semi-supervised multi-view clustering framework for screening Parkinson's disease. *Math. Biosci. Eng.* **2020**, *17*, 3395–3411. [CrossRef] [PubMed]
54. Park, C.; Lee, P.; Lee, S.; Chung, S.; Shin, N.Y. The diagnostic potential of multimodal neuroimaging measures in Parkinson's disease and atypical parkinsonism. *Brain Behav.* **2020**, *10*, e01808. [CrossRef]
55. Solana-Lavalle, G.; Rosas-Romero, R. 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.* **2021**, *198*, 105793. [CrossRef] [PubMed]
56. Talai, A.S.; Sedlacik, J.; Boelmans, K.; Forkert, N.D. Utility of Multi-Modal MRI for Differentiating of Parkinson's Disease and Progressive Supranuclear Palsy Using Machine Learning. *Front. Neurol.* **2021**, *12*, 546. [CrossRef]