Article

A Shape Approximation for Medical Imaging Data

Shih-Feng Huang [1],[*] , Yung-Hsuan Wen [2], Chi-Hsiang Chu [3] and Chien-Chin Hsu [4]

[1] Department of Applied Mathematics, National University of Kaohsiung, Kaohsiung 811, Taiwan
[2] Institute of Statistics, National University of Kaohsiung, Kaohsiung 811, Taiwan; m1074405@mail.nuk.edu.tw
[3] Department of Statistics, National Cheng Kung University, Tainan 701, Taiwan; 10902042@gs.ncku.edu.tw
[4] Department of Nuclear Medicine, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung 833, Taiwan; cchsu@cgmh.org.tw
[*] Correspondence: huangsf@nuk.edu.tw; Tel.:+886-7-5919167

check for
updates

Abstract: This study proposes a shape approximation approach to portray the regions of interest (ROI) from medical imaging data. An effective algorithm to achieve an optimal approximation is proposed based on the framework of Particle Swarm Optimization. The convergence of the proposed algorithm is derived under mild assumptions on the selected family of shape equations. The issue of detecting Parkinson's disease (PD) based on the Tc-99m TRODAT-1 brain SPECT/CT images of 634 subjects, with 305 female and an average age of 68.3 years old from Kaohsiung Chang Gung Memorial Hospital, Taiwan, is employed to demonstrate the proposed procedure by fitting optimal ellipse and cashew-shaped equations in the 2D and 3D spaces, respectively. According to the visual interpretation of 3 experienced board-certified nuclear medicine physicians, 256 subjects are determined to be abnormal, 77 subjects are potentially abnormal, 174 are normal, and 127 are nearly normal. The coefficients of the ellipse and cashew-shaped equations, together with some well-known features of PD existing in the literature, are employed to learn PD classifiers under various machine learning approaches. A repeated hold-out with 100 rounds of 5-fold cross-validation and stratified sampling scheme is adopted to investigate the classification performances of different machine learning methods and different sets of features. The empirical results reveal that our method obtains 0.88 ± 0.04 classification accuracy, 0.87 ± 0.06 sensitivity, and 0.88 ± 0.08 specificity for test data when including the coefficients of the ellipse and cashew-shaped equations. Our findings indicate that more constructive and useful features can be extracted from proper mathematical representations of the 2D and 3D shapes for a specific ROI in medical imaging data, which shows their potential for improving the accuracy of automated PD identification.

Keywords: imaging data; Parkinson's disease; PSO algorithm; shape equation

1. Introduction

In the medical field, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and single-photon emission computed tomography (SPECT) have been used for preventive medicine, screening for disease, and confirming a diagnosis. These techniques are capable of providing the three-dimensional (3D) structure of a region of interest (ROI) by a series of two-dimensional (2D) images sequentially ordered in time.

In clinical practice, doctors use their domain knowledge to select a subset of the 2D images from the series. The selected images are superimposed into a single 2D image, which is called a 2D-combined image hereafter, for disease diagnosis. For example, Prashanth et al. [1], Taylor et al. [2], Oliveira et al. [3], Nicastro et al. [4], and Iwabuchi et al. [5] constructed classifiers for identifying Parkinson's disease (PD) by extracting features from 2D-combined SPECT images. However,

this approach may lose some useful information or features for classification since a 2D-combined image is a projection of the original 3D structure on a specific 2D space. Recently, some methodologies have been proposed to draw information from medical 3D images. For instance, Ziegler et al. [6] used multispectral structural magnetic resonance imaging (MRI) tools to measure substantia nigra volume loss before basal forebrain degeneration in early PD. Palumbo et al. [7] proposed a volumetric 3D region of interest (ROI) of putamen and caudate nucleus by BasGan software. Prashanth et al. [8] proposed to draw 'indirect 3D features' by building a regression surface of an ROI based on the 2D-combined image. Hsu et al. [9] found that the classification performance for identifying PD is improved by calculating the volumes of an ROI in 3D space. Cheng et al. [10] adopted the 3D features such as volume, surface area, and diameter to describe the geometric characteristics of the volumes of interest to assist the diagnosis of idiopathic PD. Xu et al. [11] assessed the longitudinal volume change of different hippocampal subfields in PD patients with and without cognitive decline using 3D T1-weighted MRI.

Most of the aforementioned studies of drawing 3D features focused on computing the volumes of the ROI and applying the quantities derived from the volumes of the ROI to PD identification [6,7,9–11]. Although the geometric features such as volume and the volume-related quantities are valuable proxies representing the 3D shapes of the ROI, they still cannot fully capture the desired shapes since different shapes of the ROI could have the same volumes or volume-related quantities. This fact could restrict the classification performances of these volume-based methods. In addition, the 3D shapes of the ROI did affect nuclear medicine physicians' judgment in practice. In visual interpretation, the normal striatal shape looks cashew-shaped in 3D space. The common pattern of abnormalities in PD is usually a decrease uptake in the putamen (comma-shaped) initially and progressing anteriorly to caudate nucleus over time. The striatal shape becomes oval-shaped gradually with disease progression. Consequently, if there is a way to capture the 3D shapes of the ROI for both normal and abnormal subjects more specifically and directly than just computing the volumes of the ROI, we might have the chance to obtain more useful shape features for classification. To fill this gap and unlike the indirect approach of [8] to obtain '3D features', we propose to directly characterize the 3D shape of an ROI by a suitable mathematical representation. The coefficients of the fitted equations are then adopted to learn classifiers to reflect the influence of the 3D shapes of the ROI. To the best of our knowledge, this is the first work using a family of mathematical equations to portray the 3D shapes of the ROI for normal and abnormal subjects simultaneously.

Throughout this paper, we adopt the issue of identifying PD to illustrate our approach. Parkinsonian syndromes (PS) are a group of movement disorders characterized by tremor, bradykinesia, and rigidity. PD is the most common neurodegenerative movement disorder and accounts for 75% of all PS. PD with early death of dopaminergic neurons in the substantia nigra pars compacta and dopamine deficiency within the basal ganglia leads to classical parkinsonian motor symptoms [12,13]. It is important to differentiate between PS caused by nigrostriatal dopaminergic degeneration and nondegenerative causes of parkinsonism such as essential tremor, vascular parkinsonism, psychogenic parkinsonism, or drug-induced parkinsonism. Dopamine transporter (DAT) imaging can detect presynaptic dopamine neuronal dysfunction and is useful in the differentiation between conditions with and without presynaptic dopaminergic deficit. Currently, SPECT radioligands for DAT, such as I-123 FP-CIT and Tc-99m TRODAT-1, have been available for daily clinical practice [14,15]. In practice, the healthy cases and abnormal cases have different appearances of uptakes of the ROI in 2D-combined and 3D imaging data. For example, a healthy subject has comma-shaped uptakes in his/her 2D-combined SPECT image while a subject with PD has dotted-shaped uptakes. In addition, a healthy subject has cashew-shaped uptakes around his/her striatum in 3D space while a subject with PD has sphere-shaped uptakes [8]. As a result, we propose to depict the appearances of uptakes of the ROI by proper mathematical formulas for the corresponding 2D and 3D shapes, which create a possibility to identify the differences between healthy and abnormal subjects by the associated coefficients contained in the mathematical representations. Specifically, for the issue of identifying PD,

we fit ellipses for the uptakes in the 2D-combined image and cashew-shaped 3D equation for the uptakes in the 3D space for each subject under the framework of minimizing an associated loss function. Nevertheless, when conducting these two shape fittings, we have to face multidimensional optimization problems, which usually have no closed-form solution and one can only rely on numerical approximations [16].

In this study, we adopt the Particle Swarm Optimization (PSO) algorithm proposed by [17] to deal with the considered shape fittings. The PSO algorithm is inspired by the behavior of social organisms in groups, such as bird and fish schooling or ant colonies. It has been applied to numerous areas and has been shown to be capable of solving multidimensional optimization problems effectively (see [18–21] and the references therein). Another challenge encountered in the shape fitting process is to compute the Euclidean distance (also called L2 distance) between an observation and a given shape equation in R^2 or R^3 space. Morera et al. [22] proposed to calculate the L2 distance from a point to a quadratic surface by a partial differential equation (PDE) procedure. However, their PDE procedure is difficult to be extended to higher-order surfaces like the cashew-shaped equations considered in this study. To overcome this challenge, we propose a "flashlight-searching" (FS) algorithm to compute the distance for higher-order cases. The FS algorithm is shown to be capable of producing accurate results for the quadratic examples in [22]. In addition, under some mild assumptions on the selected family of shape equations, the convergence of the proposed FS algorithm is derived, which reveals the effectiveness and easy-for-implementation of the PSO-FS algorithm. Once conducting the proposed shape fitting procedure, the coefficients of the optimal 2D ellipse and cashew-shaped 3D equation are employed to learn classifiers for PD.

Between January 2017 and May 2018, 634 patients who underwent Tc-99m TRODAT-1 brain SPECT/CT in the Department of Nuclear Medicine, Kaohsiung Chang Gung Memorial Hospital, Taiwan, were enrolled in this retrospective study. The Chang Gung Medical Foundation Institutional Review Board approved this retrospective study and waived the requirement for written informed consent. The TRODAT SPECT images were interpreted by three experienced-board-certified nuclear medicine physicians, and were assigned into normal or abnormal accordingly. Six popular classification methods such as logistic regression (LR), linear discriminant analysis (LDA), naive Bayes (NB), random forests (RF), support vector machine (SVM), and extreme gradient boost (XGB) are adopted to learn classifiers for PD using the coefficients contained in the 2D and 3D shape-fitting equations. The 6 classifiers are usually adopted in many comparison studies for medical data [3,9,23–25]. The numerical results reveal that these coefficients did improve the classification accuracies for most of the considered classifiers. Beyond the 6 machine learning classifiers considered in this study, deep learning (DL) is one of the hottest technology applied to classification problem in many fields. Recently, Dolz et al. [26] proposed a 3D and fully convolutional neural network method for subcortical brain structure segmentation in MRI, and yielded satisfactory segmentations. The objective of this study is to show that suitable shape equations of the ROI are useful for classification. Intuitively, the shape equations of the ROI will be more precise if better segmentations are used. Therefore, it will be interesting to construct shape equations based on Dolz et al. [26]'s segmentations. Nevertheless, it is beyond the scope of this work, and we leave it to our future study.

The remainder of this paper is organized as follows. Section 2 describes the proposed criteria for determining suitable shape equations to approximate the regions where we are interested in. Section 3 introduces an easy-to-implement algorithm for obtaining the optimal solutions of the considered criteria. The convergence of the proposed algorithm is also derived. The numerical results of applying the proposed approach to PD imaging data are presented in Section 4. Discussions are presented in Section 5. Technical calculations and proofs are given in the Appendix A.

2. Optimization Criteria

In this section, we introduce the proposed criteria for shape fitting in 2D and 3D spaces. Shape analysis and surface fitting are useful and promising methods for extracting features, which can

be used to develop diagnostic models [8,27–29]. To construct a classifier for PD identification, Prashanth et al. [8] noticed that the shapes of uptakes in the 2D-combined images are quite different between normal and PD subjects. They employed a pair of ellipses to describe the ROI for a 2D-combined PD image, and many features are then calculated from the fitted ellipse equations. For example, the lengths of major and minor axes, the eccentricity, orientation, roundness, area, etc. of an ellipse can be computed directly from the mathematical representation of the fitted ellipses. Moreover, they proposed to use the following polynomial equation of order 3 to extract 3D features from the 2D-combined images:

$$f(x,y) = p_{00} + p_{10}x + p_{01}y + p_{20}x^2 + p_{11}xy + p_{02}y^2 + p_{30}x^3 + p_{21}x^2y + p_{12}xy^2 + p_{03}y^3, \quad (1)$$

where $f(x,y)$ is the grayscale with respect to the point (x,y) on the 2D plane and the coefficients $p_{ij}, i, j \in \{0,1,2,3\}$ are estimated by minimizing the L2 loss. Figure 1 presents the fitting results of normal and PD subjects by using Prashanth et al. [8]'s approach with (1). Note that for demonstration, we present the color version of the SPECT images throughout this study. Prashanth et al. [8] used several characteristics calculated from the fitted ellipses and the coefficients in (1) as features to learn an SVM classifier. Their numerical results reveal that the classification accuracy for PD can be effectively improved. One interesting point should be noticed that Prashanth et al. [8] did not construct a model for the real shapes of uptakes in the 3D space directly, but established a regression model for the grayscale surface by using the 2D-combined image of each subject solely. This motivates us to investigate that whether we could obtain more useful features extracted from modeling the real shapes of uptakes in the 3D space according to the series of 2D scanned images of each subject. The details are illustrated in the following.


(**a**) Normal　　　　　(**b**) Fitted surface of (**a**)

(**c**) PD　　　　　(**d**) Fitted surface of (**c**)
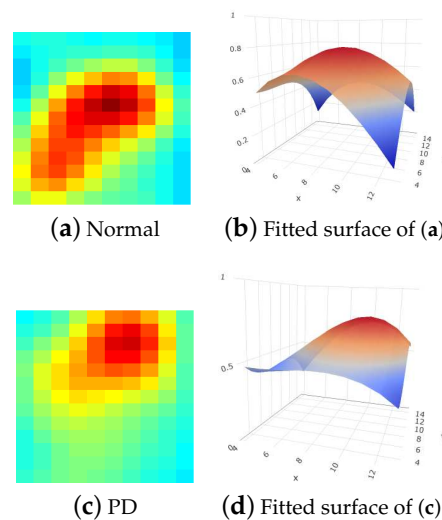
Figure 1. (**a**,**c**) are the 2D-combined SPECT images of a normal subject and a PD subject, respectively. (**b**,**d**) are the corresponding polynomial regression surfaces of order 3 for (**a**,**c**), respectively, obtained by Prashanth et al. [8]'s method.

Let $\boldsymbol{p}_i, i = 1, 2, \ldots, n$ denote the boundary points of the ROI. In Figure 2, the boundary points with grayscale value greater than 0.6 of two images are marked in black, where the left one is an image of a normal subject and the right one is an image of a subject with PD. In addition, let $\Psi(\cdot; \boldsymbol{\theta}) = 0$ denote an equation which is suitable to well depict the shape of the ROI, where $\boldsymbol{\theta}$ are unknown parameters. For example, Figure 2 shows that ellipse equations are suitable to depict the shapes of the ROI shown in a 2D-combined image and Figure 3 reveals that cashew-shaped equations are suitable to depict the structures of the ROI for a normal case or a PD case in 3D space. Intuitively, normal subjects should have similar 2D and 3D shapes while the subjects with PD should have different shapes from

normal subjects. Potentially, the coefficient $\boldsymbol{\theta}$ and the properties of the equation $\Psi(\cdot;\boldsymbol{\theta}) = 0$ could provide useful information for separating normal and abnormal cases. Nevertheless, the location of the center and the rotation angles of the surface of the uptakes for each subject (even for normal subjects) would have slight differences, which increase the complexity of fitting a suitable surface equation for the boundary points in practical implementation. In this study, we adopt the image data of PD to demonstrate the performances of the proposed method. As shown in Figures 2 and 3, the ellipse equations and cashew-shaped equations are selected to depict the ROI in 2D and 3D spaces, respectively. The details of the two families of equations for including the considerations mentioned above and the corresponding optimization criteria for 2D and 3D cases are introduced in the following.
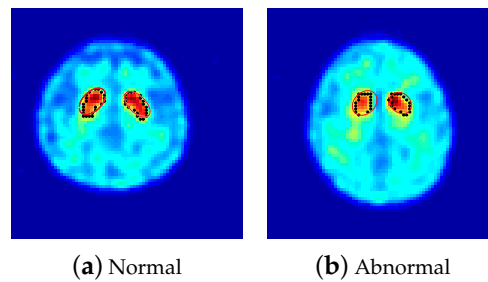
(**a**) Normal　　　　　　　(**b**) Abnormal

Figure 2. The boundary points, denoted by black points, and the approximated ellipses, denoted in red, of the region of uptakes in 2D-combined images for (**a**) a normal subject and (**b**) an abnormal subject.
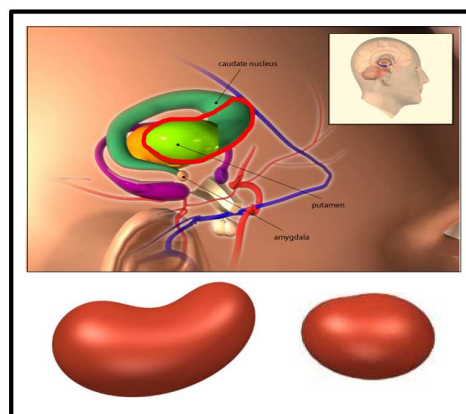
Figure 3. The 3D cashew-shaped equations to depict the ROI in 3D space, where the figure in the upper panel is downloaded from http://www.3d-brain.ki.se/atlas/images/basal_ganglia01.jpg, and the area marked with a red circle denotes the most active region of the highest-uptake. The left and right cashew-shaped structures shown in the lower panel are examples for a normal case and a PD case with two specific settings of the coefficients of the cashew-shaped equation defined in (4), respectively.

For 2D surface fitting, an ellipse equation $\Psi_{2D}(\boldsymbol{x};\boldsymbol{\theta}_{2D}) = 0$ is employed, where

$$\Psi_{2D}(\mathbf{x};\boldsymbol{\theta}_{2D}) = Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F, \qquad (2)$$

in which $\mathbf{x} = (x, y)$, $B^2 < AC$ and $\boldsymbol{\theta}_{2D} = (A, B, C, D, E, F) \in \mathbb{R}^6$. Let $H(\boldsymbol{\theta}_{2D}) = \{\mathbf{x} : \Psi_{2D}(\mathbf{x};\boldsymbol{\theta}_{2D}) = 0\}$ denote the set of points on the ellipse. An optimization criterion for finding a suitable $H(\boldsymbol{\theta}_{2D})$, abbreviated by H_{2D}, for the observed boundary points $\boldsymbol{p}_i$, $i = 1, \ldots, n$, in a 2D-combined image:

$$\boldsymbol{\theta}_{2D}^* = \arg\min_{\boldsymbol{\theta}_{2D}} \sum_{i=1}^{n} d(\boldsymbol{p}_i, H_{2D}), \qquad (3)$$

where $d(\boldsymbol{p}_i, H_{2D}) = \min_{\mathbf{x} \in H_{2D}} d(\boldsymbol{p}_i, \mathbf{x})$, and $d(\boldsymbol{p}_i, \mathbf{x}) = \|\boldsymbol{p}_i - \mathbf{x}\|$ denote the Euclidean distance between $\boldsymbol{p}_i$ and $\mathbf{x}$.

For 3D surface fitting, a cashew-shaped equation is defined as $\Psi^0_{3D}(\mathbf{x};\boldsymbol{\theta}_{3D}) = 0$, where

$$\Psi^0_{3D}(\mathbf{x};\boldsymbol{\theta}^0_{3D}) = (1.5x^2 + y^2 + z^2 + a^2 - b^2)^2 - 4a^2((x-c)^2 + (y-d)^2), \tag{4}$$

in which $\mathbf{x} = (x,y,z)$ and $\boldsymbol{\theta}^0_{3D} = (a,b,c,d) \in [0,1]^4$. In order to consider the translation, rotation, and scaling of the equation, we do the following:

1.　Rotate $\mathbf{x}$ by the following rotation matrix M with three radians $\theta_x, \theta_y, \theta_z \in [0, 2\pi]^3$:

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_x) & \sin(\theta_x) \\ 0 & -\sin(\theta_x) & \cos(\theta_x) \end{pmatrix} \times \begin{pmatrix} \cos(\theta_y) & 0 & -\sin(\theta_y) \\ 0 & 1 & 0 \\ \sin(\theta_y) & 0 & \cos(\theta_y) \end{pmatrix} \times \begin{pmatrix} \cos(\theta_z) & \sin(\theta_z) & 0 \\ -\sin(\theta_z) & \cos(\theta_z) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2.　Furthermore, by translating and scaling x by $\mathbf{t} = (t_x, t_y, t_z) \in \mathbb{R}^3$ and $s > 0$, respectively, the general form of a cashew-shaped equation is represented by

$$\Psi_{3D}(\mathbf{x};\boldsymbol{\theta}_{3D}) = \Psi^0_{3D}\Big(sM(\mathbf{x}-\mathbf{t});\boldsymbol{\theta}^0_{3D}\Big), \tag{5}$$

where $\boldsymbol{\theta}_{3D} = (a,b,c,d,s,t_x,t_y,t_z,\theta_x,\theta_y,\theta_z)$ and $\Psi^0_{3D}(\cdot)$ is defined in (4).

Based on the representation shown in (5), let $H(\boldsymbol{\theta}_{3D}) = \{\mathbf{x} : \Psi_{3D}(\mathbf{x};\boldsymbol{\theta}_{3D}) = 0\}$ denote the set of points on the cashew-shaped. Similar to (3), an optimization criterion for finding a suitable $H(\boldsymbol{\theta}_{3D})$, abbreviated by H_{3D}, for the observed boundary points $\boldsymbol{p}_i$, $i = 1,\ldots,n$, in 3D space:

$$\boldsymbol{\theta}^*_{3D} = \arg\min_{\boldsymbol{\theta}_{3D}} \sum_{i=1}^{n} d(\boldsymbol{p}_i, H_{3D}), \tag{6}$$

where $d(\boldsymbol{p}_i, H_{3D}) = \min_{\mathbf{x} \in H_{3D}} d(\boldsymbol{p}_i, \mathbf{x})$.

The idea for determining optimal shape equations through the criteria shown in Equations (3) and (6) for 2D and 3D cases, respectively, is based on minimizing the L2 distance of the observed points to an ellipse or a cashew-shaped equation, which is similar to the classical least-squares (LS) regression. However, the classical LS cannot be applied directly for the equations $\Psi_{2D} = 0$ and $\Psi_{3D} = 0$. For example, the variable y in the equation $\Psi_{2D} = 0$ is not a function of x, which makes the shape approximation problem different from classical regression. Therefore, we propose an algorithm to find the optimal solutions of (3) and (6), and the details are introduced in the next section.

3. The Proposed Algorithm

For the boundary points $\boldsymbol{p}_i$, $i = 1,\ldots,n$, and a determined family of surface equation $\Psi(\cdot;\boldsymbol{\theta}) = 0$, we propose to use an optimal surface satisfying $\Psi(\cdot;\boldsymbol{\theta}^*) = 0$ which has the minimal L2 distance to the boundary points $\boldsymbol{p}_i$, $i = 1,\ldots,n$, to approximate the shapes of the uptakes of the ROI as shown in (3) and (6). Since the optimal solution $\boldsymbol{\theta}^*$ is a high-dimensional vector and there is usually no closed-form representation for $\boldsymbol{\theta}^*$, a PSO framework is employed to obtain $\boldsymbol{\theta}^*$ and an effective scheme, called "flashlight-searching" (FS), is proposed to calculate the distance from a point to a surface in each PSO iteration. Herein, the proposed algorithm is abbreviated by "PSO-FS". In the following, we briefly introduce the PSO procedure.

Kennedy et al. [17] proposed an optimization algorithm, called PSO, which searches the optimal solution through particles and emulates the interaction of information sharing among particles iteratively. In each iteration, the updated search position of each particle is determined by the best position in its historical trajectory and the best position found in the trajectories of all particles. Specifically, we denote the position of the i-th particle in R^d by a d-dimensional vector $\mathbf{x}_i$ and denote

its velocity vector by $\mathbf{v}_i$, $i = 1, \ldots, n$. In the $(t+1)$-th iteration, each particle changes its position according to the new velocity. That is, the position $\mathbf{x}_i^{t+1}$ is defined as:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \mathbf{v}_i^{t+1} \tag{7}$$

with the new velocity $\mathbf{v}_i^{t+1}$ defined by

$$\mathbf{v}_i^{t+1} = \omega \mathbf{v}_i^t + c_1 r_1 (P_i^t - \mathbf{x}_i^t) + c_2 r_2 (G^t - \mathbf{x}_i^t), \tag{8}$$

where P_i^t and G^t denote the best position of the i-th particle and the best position of all particles up to the t-th iteration, respectively. The parameters ω, (c_1, c_2) and (r_1, r_2) are inertia weight, two positive constants and two random parameters within $[0, 1]$, respectively. The PSO algorithm is repeated until a predetermined stopping criterion is reached or the change rate of the particles approaches zero.

For example, suppose that the objective is to find the minimum of a surface $f : R^2 \rightarrow R$, where the contour plots of f with the paths of 5 particles for the first 4 iterations of PSO are shown in Figure 4. In the four panels of Figure 4, the solid points denote the locations of the 5 particles in each iteration and the gray points combined with dashed lines denote the corresponding path of each particle. For each particle, the red arrow denotes the velocity of the i-th particle $\mathbf{v}_i^{t+1}$ in the t-th iteration, and the black, green, and yellow dotted-arrows denote the 3 components, $\mathbf{v}_i^t$, $(P_i^t - \mathbf{x}_i^t)$, and $(G^t - \mathbf{x}_i^t)$, respectively, on the right-hand side of (8) to obtain $\mathbf{v}_i^{t+1}$. In addition, the green circle represents P_i^t and the particle with red cross represents G^t. From Figure 4, one can find that the particles did share their information with others to determine their searching directions in each iteration and all of them tend to approach the point with minimal f gradually.

In this study, we employ the PSO procedure to obtain the optimal solutions of the criteria (3) and (6) in 2D and 3D spaces, respectively. For example, to proceed with the PSO procedure for searching θ_{2D}^* defined in (3), the vector $\mathbf{x}_i^t$ defined in (7) is set to be $\mathbf{x}_i^t = \theta_{2D,i}^t$, where $\theta_{2D,i}^t$ denotes the position of the i-th particle at the t-th PSO iteration. In addition, the P_i^t and G^t defined in (8) are computed from $\Psi_{2D}(\cdot)$ defined in (2).

Nevertheless, to calculate (3) and (6) in each PSO iteration, we have to calculate $d(\boldsymbol{p}_i, H)$, which usually does not have a closed-form representation and a numerical method can help to compute the distance. For example, Morera et al. [22] proposed a numerical method under the framework of PDE to compute the L2 distance from a point to a quadric surface. However, their approach heavily relies on the properties of quadric equations when handling the corresponding PDEs and is not trivial to be extended to the cases of higher order surfaces, like the cashew-shaped equations considered in this study. Therefore, we propose an easy-to-implement FS algorithm instead to overcome the difficulty. The details of the FS algorithm are illustrated in the following.

For any given $\boldsymbol{p} \in \mathbb{R}^d$, $d \geq 2$, let $D_p \subseteq H$ denote a set satisfying

$$D_p = \{\mathbf{x}^* \in H : d(\boldsymbol{p}, \mathbf{x}^*) = d(\boldsymbol{p}, H)\}, \tag{9}$$

which allows the case of more than one $\mathbf{x}^* \in H$ such that $d(p, \mathbf{x}^*) = d(p, H)$. Further, let $\overline{B}_\epsilon(\mathbf{x}^*) = \{\mathbf{x} : ||\mathbf{x} - \mathbf{x}^*|| \leq \epsilon, \mathbf{x}^* \in D_p\}$ denote a closed ball of radius $\epsilon > 0$ centered at an $\mathbf{x}^* \in D_p$, and $J = \{\bigcup_{\mathbf{x}^* \in D_p} \overline{B}_\epsilon(\mathbf{x}^*)\} \cap H$ denote the arc of H associated with $\bigcup_{\mathbf{x}^* \in D_p} \overline{B}_\epsilon(\mathbf{x}^*)$. Since J is allowed to be a disconnected set, let $\{J_i, i = 1, 2, \ldots\}$ denote a partition of J, where $\bigcup_{i=1,2,\ldots} J_i = J$, $J_i \cap J_k = $ for $i \neq k$, and each J_i is a connected set. A schematic diagram of these notations is given in Figure 5 for the special case of D_p containing only one $\mathbf{x}^*$. The FS algorithm for computing $d(\boldsymbol{p}_i, H)$ proceeds as follows.

(**a**) iterate = 1　　　　　　　　　　　　　　　　　(**b**) iterate = 2

(**c**) iterate = 3　　　　　　　　　　　　　　　　　(**d**) iterate = 4
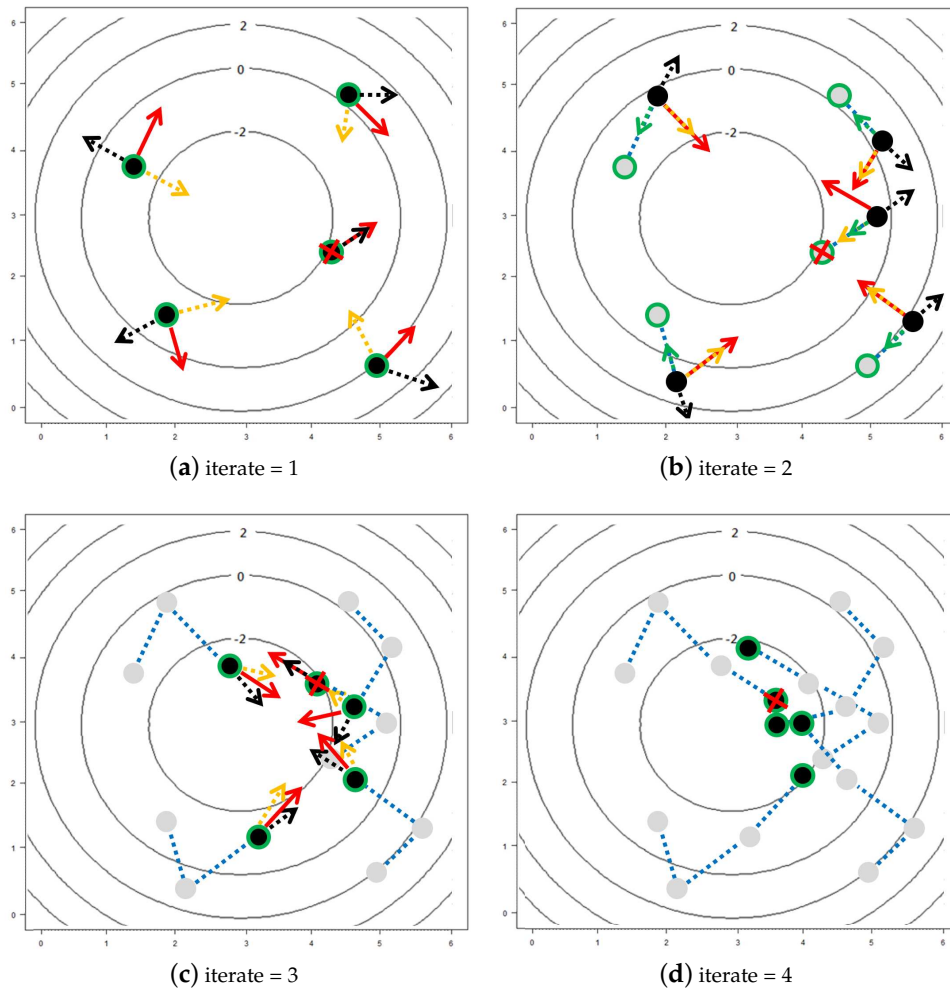
Figure 4. A schematic diagram of the Particle Swarm Optimization (PSO) algorithm with 5 particles in the first 4 iterations, where black points denote the current positions of the particles, gray points combined with dashed lines denote the corresponding path of each particle, and the red arrows denote the velocity of each particle determined by the black, green, and yellow dotted-arrows, which represent the three directions of the components on the right-hand side of (8). In the *t*-th iteration, the points with green circle denote P_i^t for $i = 1, \ldots, 5$, and the point with red cross denotes G^t, $t = 1, \ldots, 4$, where P_i^t and G^t are defined in (8).
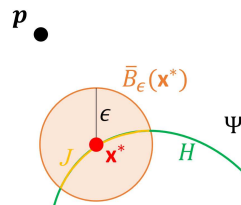
Figure 5. A schematic diagram of the flashlight-searching (FS) algorithm, where p is a point in $\mathbb{R}^d$, $d \geq 2$, H is the set of points satisfying $\Psi = 0$, $\mathbf{x}^*$ is the point such that $d(p, \mathbf{x}^*) = d(p, H)$, $\overline{B}_\epsilon(\mathbf{x}^*) = \{\mathbf{x} : ||\mathbf{x} - \mathbf{x}^*|| \leq \epsilon\}$ denotes a closed ball of radius $\epsilon > 0$ centered at an $\mathbf{x}^*$, and $J = \overline{B}_\epsilon(\mathbf{x}^*) \bigcap H$ denotes the arc of H associated with $\overline{B}_\epsilon(\mathbf{x}^*)$.

1.　Set an $r > 0$ such that for all $y \in \overline{B}_r(p)$, we have $\Psi(y) > 0$ or $\Psi(y) < 0$, where $\overline{B}_r(p)$ is defined similarly to $\overline{B}_\epsilon(\mathbf{x}^*)$. Let $\mathbf{x}_0 \in H$ denote the initial point for estimating $d(p, H)$ and $\mathbf{x}_0$ can be obtained by letting

$$\mathbf{x}_0 = c(\mathbf{y}_r - p) + p, \ c \in \mathbb{R}, \tag{10}$$

where $\mathbf{y}_r = \arg\min_y \{\Psi(y) : y \in \overline{B}_r^b(p)\}$ and $\overline{B}_r^b(p)$ denotes the set of boundary points of $\overline{B}_r(p)$.

2. Let τ be a predetermined tolerance, and set $m = 0$.

　(i) Compute the tangent plane to Ψ at $\mathbf{x}_m$ and denote it by Γ_m. Let $\{\mathbf{x}_{m,j}, j = 1, \ldots, k\}$ be a set of points on Γ_m and $||\mathbf{x}_{m,j} - \mathbf{x}_m|| = \delta$, where δ is a predetermined positive constant.

　(ii) Let $\tilde{\mathbf{x}}_{m,j} \in J$ and $\tilde{\mathbf{x}}_{m,j} = c_j(\mathbf{x}_{m,j} - p) + p$ as defined in (10), where $c_j \in \mathbb{R}$, $j = 1, \ldots, k$. Let

$$\mathbf{x}_{m+1} = \arg\min_{j=1,\ldots,k} ||p - \tilde{\mathbf{x}}_{m,j}||.$$

　(iii) Let $G(\mathbf{x}) = \cos(\theta_\mathbf{x})$, where $\mathbf{x} \in J$, $\theta_\mathbf{x} \in [0, \pi/2]$ is the acute angle of two vectors, $v_\mathbf{x} = (p - \mathbf{x})$ and $n_\mathbf{x}$, which denotes the normal vector to Ψ at $\mathbf{x}$. If $1 - G(\mathbf{x}_{m+1}) < \tau$, set $\hat{d}(p, H) = d(p, \mathbf{x}_{m+1})$ and stop, where $\hat{d}(p, H)$ denotes the estimate of $d(p, H)$. Otherwise, set $m = m + 1$ and repeat Steps (i) and (ii).

Figure 6 presents a schematic diagram of the FS algorithm at the m-th iteration. The idea of the FS algorithm comes from the gradient descent approach. In Step 1 of the algorithm, the initial point $\mathbf{x}_0$ is determined by looking around the neighborhood of p and selecting the searching direction by $\mathbf{y}_r$, which can be treated as a hint of the direction with high possibilities to reach an $\mathbf{x}^* \in D_p$. Similarly, the algorithm selects the next searching direction around $\mathbf{x}_m$ at the m-th iteration in the Step 2(i). This searching method mimics the scenario when one looks for a way out by a flashlight in a dark room, where r and δ in Steps 1 and 2(i), respectively, play the roles of different irradiation distances of the flashlight. Therefore, we name the above scheme by FS algorithm.
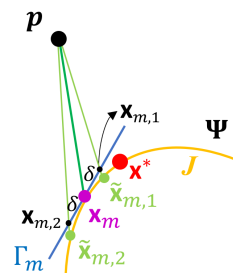
Figure 6. A schematic diagram of the m-th iteration of the FS algorithm, where p, Ψ, J, and $\mathbf{x}^*$ are defined the same as in Figure 5, $\mathbf{x}_m$ is the point at m-th iteration, Γ_m is the tangent plane of Ψ at $\mathbf{x}_m$, $\mathbf{x}_{m,1}$, and $\mathbf{x}_{m,2}$ are points satisfying $||\mathbf{x}_{m,j} - \mathbf{x}_m|| = \delta$, δ is a predetermined positive constant, and $\tilde{\mathbf{x}}_{m,j}$ is the points in J satisfying $\tilde{\mathbf{x}}_{m,j} = c_j(\mathbf{x}_{m,j} - p) + p$, and $c_j \in \mathbb{R}$, $j = 1, \ldots, k$.

By using the FS algorithm in each iteration of the PSO procedure, the optimal solutions of the criteria (3) and (6) can be obtained with the boundary points p_i, $i = 1, \ldots, n$. In the rest of this section, we derive the convergence of the FS algorithm. To begin, assume the following conditions.

C1. Ψ is continuously differentiable, denoted by $\Psi \in C^1$.

C2. For any fixed $p \in \mathbb{R}^3$, $\mathbf{x} \in J$, and $d(p, \mathbf{x}) > d(p, H)$, there exists a 2-dimensional neighborhood $\kappa_\mathbf{x}$ of $\mathbf{x}$, where $\kappa_\mathbf{x} = B_\zeta^{(2)}(\mathbf{x}) \cap J$ with $B_\zeta^{(2)}(\mathbf{x})$ being a 2-dimensional open ball of radius $\zeta \in (0, \epsilon]$ centered at $\mathbf{x}$, such that $d(p, \mathbf{x}) > d(p, \kappa_\mathbf{x})$ and for any $\mathbf{x}_a$ and $\mathbf{x}_b \in \kappa_\mathbf{x} \cap J_i$, $i = 1, 2, \ldots$, and for any $\alpha \in [0, 1]$, $\Psi(\cdot)$ satisfies $\Psi\left(\alpha\mathbf{x}_a + (1 - \alpha)\mathbf{x}_b\right)\Psi(p) \leq 0$ or $\Psi\left(\alpha\mathbf{x}_a + (1 - \alpha)\mathbf{x}_b\right)\Psi(p) > 0$.

The above two conditions are set on the function Ψ, which means that these two conditions can be checked for any selected family of shape equations. Specifically, C1 is a smoothness condition, which is satisfied by the equations with polynomial, exponential, logarithmic, and trigonometric components. C2 is a local convexity condition defined on a 2D subspace in $\mathbb{R}^3$. An intuitive way to understand C2 is to cut a cashew, for example, with a knife and check whether C2 is satisfied by the shape of the boundary of the cross-section.

In this study, we employ the ellipse and cashew-shaped equations to depict the uptake regions in the 2D and 3D spaces, respectively. Apparently, the families of equations defined in (2) and (4) both satisfy C1. Next, we demonstrate that the family of ellipse equations satisfy C2. Without loss of generality, we consider the equation of the normal to the ellipse, that is, $K(x, y) = ax^2 + by^2 = 1$, $a, b > 0$. For any $(x_1, y_1) \neq (x_2, y_2)$ such that $K(x_1, y_1) = K(x_2, y_2) = 1$, let $(u, v) = \alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2)$, where $0 \leq \alpha \leq 1$. By Jensen's inequality and since $f(x) = x^2$ is a convex function, we have $K(u, v) \leq \alpha K(x_1, y_1) + (1 - \alpha)K(x_2, y_2) = 1$, which implies that $(u, v) \in \{(x, y) : K(x, y) \leq 1\}$ for all $0 \leq \alpha \leq 1$ and, therefore, $\{(x, y) : K(x, y) \leq 1\}$ is a convex set. As a result, let $\Psi(\mathbf{x}) = K(\mathbf{x}) - 1$ and if $\Psi(\mathbf{x}_a) = \Psi(\mathbf{x}_b) = 0$, where $\mathbf{x}_a \neq \mathbf{x}_b$, we have $\Psi\left(\alpha\mathbf{x}_a + (1 - \alpha)\mathbf{x}_b\right) \leq 0$ for $\alpha \in [0, 1]$, and C2 holds no matter $\mathbf{p}$ is an interior or exterior point of $\{\mathbf{x} : \Psi(\mathbf{x}) = 0\}$. For the cashew-shaped equations, C2 is also satisfied. However, the calculation is more tedious than that for the ellipse equations. We put the details in the Appendix A.

Based on C1 and C2, we have the following result.

Theorem 1. *Let $\mathbf{x}_m$, $m = 0, 1, 2, \ldots$, be a sequence of points obtained from the FS algorithm. If C1 and C2 hold, then the sequence $d(\mathbf{p}, \mathbf{x}_m)$ converges to $d(\mathbf{p}, H)$ as $\delta \to 0$ and $\tau \to 0$.*

The proof the Theorem 1 is given in the Appendix A. Most importantly, Theorem 1 indicates that the proposed FS algorithm is capable of computing an accurate distance from a fixed point to a given surface when the surface equation satisfies the smoothness and local convexity conditions as shown in C1 and C2, respectively.

4. Numerical Results

In this section, we introduce the details of extracting features from the ellipses and cashew-shaped surfaces. In addition, 6 popular classification methods are employed to investigate whether these features are useful for identifying PD subjects from the SPECT images. The SPECT images of 634 subjects with 305 female and an average age of 68.3 years old collected from Kaohsiung Chang Gung Memorial Hospital between January 2017 and May 2018 are employed for our investigation. For each subject, 3 SPECT images are selected and combined to a 2D-combined image, and 4 SPECT images are selected for fitting cashew-shaped surfaces according to physicians' expertise and experience. Three experienced board-certified nuclear medicine physicians are invited to give their opinions on the 2D-combined image of each subject. Blinded to patient clinical symptoms and histories, the physicians determine the images independently according to the guideline. The final consensus result of each subject was assigned when at least 2 physicians achieved a common agreement, where 256 subjects were assigned to be abnormal (grade 2 or 3), 77 subjects were potentially abnormal (grade 1), 174 were normal, and 127 were nearly normal, which includes the slightly degenerated cases without neurodegenerative disorder of PD. For simplicity, we combine the first 2 categories as abnormal and the last two categories as normal, denoted by 1 and 0, respectively. Figure 2 shows examples of 2D-combined images for the above 2 labels.

Before collecting the shape features, a sensitivity study for the settings of (c_1, c_2, ω) in (8) is provided. In the literature, the parameters c_1, c_2, and w are set differently for different data [30–32]. In these studies, c_1 and c_2 are tuned between 1 and 3, and ω is tuned between 0.1 and 1.2. In the following, we conduct a sensitivity analysis with setting $c_1 \in \{0.8, 1.0, 1.2\}$, $c_2 \in \{1.6, 2.0, 2.4\}$, and $\omega \in \{0.4, 0.5, 0.6\}$ based on our experience when applying PSO to our data for the subjects shown in Figures 2 and 7. Table 1 reports the results of these 27 different combinations of (c_1, c_2, ω). From the means and standard deviations presented in Table 1, we found that the values of $n^{-1}\sum_{i=1}^{n} d(p_i, \Psi_{2D})$ and $n^{-1}\sum_{i=1}^{n} d(p_i, \Psi_{3D})$, which are the values of the objective functions defined in (3) and (6), respectively, are not sensitive to the 27 different settings. To save the computational costs, we adopt $c_1 = 1$, $c_2 = 2$ and $\omega = 0.5$ in the following comparison study.

Table 1. The means and standard deviations (SDs) of $n^{-1} \sum_{i=1}^{n} d(p_i, \Psi_{2D})$ and $n^{-1} \sum_{i=1}^{n} d(p_i, \Psi_{3D})$ for the subjects in Figures 2 and 7 based on 27 different settings of (c_1, c_2, ω) in (8), where $c_1 = 0.8, 1.0, 1.2$, $c_2 = 1.6, 2.0, 2.4$, and $\omega = 0.4, 0.5, 0.6$.

	Normal				Abnormal			
	2D		3D		2D		3D	
	L	R	L	R	L	R	L	R
mean	0.55	0.69	33.75	42.22	0.40	0.40	34.28	30.52
SD	0.01	0.14	1.43	3.81	0.01	0.04	0.97	0.95

Once the parameters in Equations (3) and (6) are estimated by the convergent values of the proposed algorithm with the SPECT images for each subject, we treat these parameters of the ellipse and cashew-shaped equations as the features extracted from 2D and 3D ROIs, respectively. To investigate whether the features obtained from the ellipse and cashew-shaped equations are useful, we consider the following 6 different sets of features and compare their classification performances for PD identification.

Set 1 Grayscale features (9 features): For each 2D-combined image, we compute the empirical density of the grayscale values, which are normalized into $[0, 1]$, and calculate 8 probabilities between two adjacent percentiles in $\{0.60, 0.65, \ldots, 1\}$ to represent the ratios of areas having different magnitudes of uptakes. In addition, we also compute the ratios of average uptake values near striatum to the average uptake values of the background.

Set 2 The features extracted from the ellipse equations (12 features): We employed the PSO-FS algorithm to approximate the shape of striatum for each 2D-combined image by the criterion defined in (3), where 100 particles are randomly set centered on an initial estimator of θ_{2D}. Figure 2 presents approximated ellipses in red of a normal subject and a subject with PD, which reveal that the PSO-FS algorithm is capable of obtaining satisfactory ellipses to approximate the shapes of uptakes.

Set 3 The features extracted from the cashew-shaped equations (22 features): The procedure for fitting a 3D structure is similar to the above 2D shape fitting process. Figure 7 presents the sequences of 2D images in the upper panel and the corresponding cashew-shaped surfaces in 3D space in the lower panel of a normal subject and a subject with PD, where the 2D images in the red rectangles are recommended by the physicians and are used to construct the cashew-shaped surfaces. The fitting results shown in Figure 7 display that the PSO-FS algorithm is capable of obtaining satisfactory cashew-shaped equations to approximate the shapes of uptakes.

Set 4 The union of Sets 1 and 2.

Set 5 The union of Sets 1, 2, and 3.

Set 6 The 34 features defined in Table 1 of [8].

Next, 6 well-known classification methods, which are support vector machine (SVM), naive Bayes (NB), random forests (RF), XGboost (XGB), logistic regression (LR), and linear discriminant analysis (LDA), are used to learn classifiers for PD imaging classification with the above 6 sets of features separately [3,9,23–25]. The comparison study is conducted under the following repeated hold-out with 100 rounds of 5-fold cross-validation (CV) and stratified sampling scheme.

1. We randomly split the data into training and test sets with the proportions of 80% and 20%, respectively, under a stratified sampling scheme, where the proportions of normal, nearly normal, potentially abnormal, and abnormal subjects in the training and test sets are the same.

2. In the training set, we conduct a stratified 5-fold CV framework on each of the SVM, NB, RF, XGB, LR, and LDA classifiers by R-packages svm(), naiveBayes(), randomForest(), xgboost(), glm(), and lda(), respectively. The tuning parameters in each of the SVM, RF, and XGB methods are determined by computing the validation accuracies under several settings of tuning parameters and selecting the one with the highest validation accuracy. For each of the SVM, RF, and XGB

　　　classifiers, we set 5 candidates for each tuning parameters centered at the default values of the corresponding R packages. Therefore, we have 6 candidate classifiers, where each of them has the best 5-fold CV performances for the training data.

3.　Let the classifier with the best 5-fold CV accuracy from the 6 candidates be our final selection. Learn this classifier again with all the training data and use it to compute the classification performances of the test data.

4.　Repeat the above 3 steps 100 times. Compute the mean and standard deviation (SD) of accuracy (ACC), sensitivity (SEN), specificity (SPE), and GM under the 5-fold CV framework of the 6 classifiers for training data, and compute the mean, SD, and 95% confidence interval (CI) of the 4 measurements for test data based on the results of the 100 rounds.
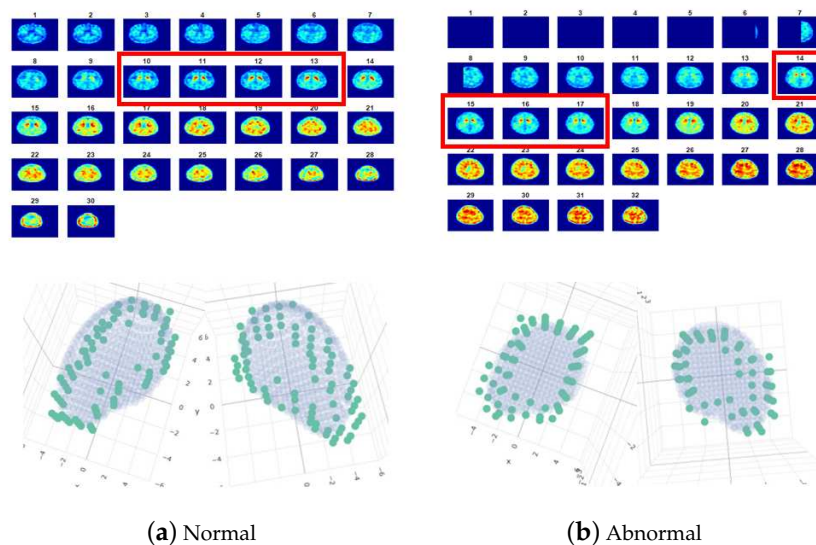
(**a**) Normal　　　　　　　　　　　　　　　(**b**) Abnormal

Figure 7. The sequences of 2D images (in the upper panel) and the corresponding cashew-shaped surfaces (in the lower panel) of (**a**) a normal subject and (**b**) an abnormal subject in 3D space. In the upper panel, the four 2D images in the red rectangles are suggested by the physicians and are used to construct the cashew-shaped surfaces. In the lower panel, the green points denote the boundary points collected from the four 2D images for left/right striatum in 3D space, and the gray areas are the corresponding fitted cashew-shaped surfaces.

In the last step of the above procedure, the ACC, SEN, SPE, and GM [33] are defined by

$$
\begin{aligned}
\text{ACC} &= (\text{TP} + \text{TN})/\text{Total population} \\
\text{GM} &= (\text{SEN} \times \text{SPE})^{0.5},
\end{aligned}
$$

where SEN = TP/True condition positive, SPE = TN/True condition negative, and 'TP' and 'TN' are short for 'True Positive' and 'True Negative', respectively.

　　　The results of 5-fold CV and test sets are reported in Tables 2 and 3, respectively. In Table 2, the highest AVEs of the ACC, SEN, SPE, and GM values among the 6 classifiers for each set of features are marked in bold, where the RF classifier has more robust performances than other classifiers for the 6 sets of features on the average. In Table 3, the highest AVEs of the ACC, SEN, SPE, and GM values among the 6 sets of features are marked in bold, where Set 5 has the highest ACC, SEN, and GM values. Interestingly, in the 100 rounds, the SVM and RF classifiers are selected to be our final classifier for test data most frequently with the features contained in Set 1. However, this is not the case for other sets of features. For example, the RF and LR classifiers are preferred for Set 3; the RF and LDA classifiers are preferred for Set 6. This phenomenon reveals that different sets of features or different data might

need different classifiers to highlight their impact of classification, which is also the main reason why we use a 5-fold CV scheme to select our final classifier in step 3 of the above procedure.

Table 2. The AVEs and SDs of the ACC, SEN, SPE and GM values of the SVM, NB, RF, XGB, LR, and LDA classifiers with 6 collections of features for the 5-fold CV results based on 100 rounds, where the highest AVEs of the ACC, SEN, SPE, and GM values among the 6 classifiers for each set of features are marked in bold.

			Classifier					
Measure	**Set**		**SVM**	**NB**	**RF**	**XGB**	**LR**	**LDA**
ACC	1	AVE	0.865	0.753	**0.866**	0.836	0.842	0.848
		SD	0.008	0.008	0.009	0.013	0.009	0.008
	2	AVE	**0.798**	0.760	0.793	0.747	0.789	0.776
		SD	0.008	0.007	0.010	0.015	0.010	0.010
	3	AVE	0.798	0.805	**0.826**	0.788	0.824	0.819
		SD	0.012	0.009	0.010	0.016	0.010	0.009
	4	AVE	0.873	0.788	**0.880**	0.845	0.854	0.862
		SD	0.007	0.008	0.007	0.013	0.009	0.008
	5	AVE	0.824	0.824	**0.880**	0.835	0.855	0.859
		SD	0.010	0.008	0.007	0.012	0.011	0.009
	6	AVE	0.619	0.620	**0.865**	0.830	0.849	0.857
		SD	0.011	0.012	0.009	0.014	0.004	0.009
SEN	1	AVE	0.823	0.564	**0.846**	0.825	0.819	0.765
		SD	0.013	0.014	0.011	0.018	0.012	0.015
	2	AVE	0.760	0.642	**0.785**	0.734	0.772	0.700
		SD	0.015	0.011	0.013	0.018	0.013	0.016
	3	AVE	0.715	0.706	0.775	0.760	**0.799**	0.771
		SD	0.026	0.017	0.014	0.021	0.013	0.013
	4	AVE	0.845	0.656	**0.870**	0.828	0.836	0.802
		SD	0.010	0.013	0.008	0.017	0.013	0.011
	5	AVE	0.766	0.733	**0.874**	0.813	0.843	0.818
		SD	0.017	0.012	0.009	0.016	0.013	0.012
	6	AVE	**0.975**	0.286	0.852	0.815	0.846	0.812
		SD	0.006	0.024	0.010	0.017	0.007	0.013
SPE	1	AVE	0.911	**0.963**	0.888	0.848	0.867	0.939
		SD	0.014	0.005	0.012	0.019	0.013	0.010
	2	AVE	0.841	**0.891**	0.802	0.761	0.808	0.861
		SD	0.015	0.009	0.016	0.020	0.013	0.013
	3	AVE	0.891	**0.915**	0.882	0.818	0.853	0.873
		SD	0.020	0.011	0.012	0.019	0.014	0.012
	4	AVE	0.903	**0.935**	0.890	0.864	0.873	0.928
		SD	0.011	0.006	0.011	0.019	0.013	0.011
	5	AVE	0.888	**0.925**	0.888	0.858	0.869	0.904
		SD	0.016	0.009	0.009	0.018	0.014	0.011
	6	AVE	0.224	**0.990**	0.879	0.847	0.851	0.907
		SD	0.022	0.004	0.014	0.021	0.006	0.013
GM	1	AVE	0.866	0.737	**0.867**	0.837	0.843	0.848
		SD	0.008	0.010	0.009	0.013	0.009	0.008
	2	AVE	**0.799**	0.756	0.793	0.747	0.790	0.776
		SD	0.008	0.008	0.010	0.015	0.010	0.010
	3	AVE	0.798	0.804	**0.827**	0.788	0.825	0.820
		SD	0.012	0.010	0.010	0.016	0.010	0.009
	4	AVE	0.874	0.783	**0.880**	0.845	0.854	0.863
		SD	0.007	0.008	0.007	0.013	0.009	0.008
	5	AVE	0.825	0.824	**0.881**	0.835	0.856	0.860
		SD	0.010	0.008	0.007	0.012	0.011	0.009
	6	AVE	0.467	0.531	**0.865**	0.831	0.849	0.858
		SD	0.023	0.022	0.009	0.014	0.004	0.009

In view of Tables 2 and 3, we found that the classification performances did not depend on the number of features solely. For example, the ACCs and GMs of Set 1 are higher than those of Sets 2 and 3 in most cases although Set 1 contains less number of features than Set 2 or Set 3. Nevertheless, when we combine the features in Sets 1 and 2 into Set 4 or the features in Sets 1-3 into Set 5, the classification performances are improved significantly in terms of paired-wise *t*-tests as shown

in the previous question. These phenomena reveal that the features in Set 1 did provide useful and important information for identifying PD but there is still room for improvement. The numerical results in Tables 2 and 3 reveal that the 2D and 3D features in Sets 2 and 3, respectively, extracted from the proposed mathematical representations show their potential for improving the accuracy of automated PD identification.

Moreover, in Table 3, the average ACCs for test data based on 100 rounds with the 6 sets of features are 0.858, 0.792, 0.825, 0.873, 0.880, and 0.858, respectively. Since we use the same hold-out and 5-fold CV samples to learn classifiers with each of the 6 feature sets in each round, we adopt paired-wise *t*-tests on the ACCs of 4 cases, 'Set 6 vs. Set 1', 'Set 1 vs. Set 4', 'Set 6 vs. Set 5', and 'Set 4 vs. Set 5', to test whether these sets of features provide significantly different ACCs for test data. The *p*-values of these 4 cases are 0.44, 2.3×10^{-7}, 4.3×10^{-11}, and 7.3×10^{-4}, respectively, which indicate that Set $5 \succ$ Set $4 \succ$ Set $6 \approx$ Set 1, where $A \succ B$ denotes that the population mean of A is significantly higher than the population mean of B, and $A \approx B$ denotes that population means of A and B have no significant difference. Figure 8 presents the scatter plots of the ACCs of the 4 cases for test data based on 100 rounds, which reveal consistent phenomena with the statistical test results since most points in Figure 8b–d are located below the 45-degree line.

Table 3. The AVEs, SDs, and 95% CIs of the ACC, SEN, SPE, and GM values of 6 collections of features for the test data based on 100 rounds, where the highest AVEs of the ACC, SEN, SPE, and GM values among the 6 sets of features are marked in bold. In addition, $n_1, n_2, \ldots,$ and n_6, respectively, denote the numbers of times that the SVM, NB, RF, XGB, LR, and LDA classifiers being chosen in the 100 rounds.

Set		**Measure**				**# Times**					
		ACC	**SEN**	**SPE**	**GM**	n_1	n_2	n_3	n_4	n_5	n_6
1	AVE	0.858	0.822	**0.898**	0.859	45	0	54	0	1	0
	SD	0.024	0.042	0.036	0.024						
	95%CI	(0.81,0.91)	(0.74,0.91)	(0.83,0.97)	(0.81,0.91)						
2	AVE	0.792	0.759	0.828	0.792	60	0	23	0	17	0
	SD	0.024	0.045	0.047	0.024						
	95%CI	(0.74,0.84)	(0.67,0.85)	(0.73,0.92)	(0.74,0.84)						
3	AVE	0.825	0.788	0.866	0.825	0	0	54	0	41	5
	SD	0.028	0.049	0.045	0.028						
	95%CI	(0.77,0.88)	(0.69,0.89)	(0.78,0.96)	(0.77,0.88)						
4	AVE	0.873	0.855	0.892	0.873	18	0	80	0	0	2
	SD	0.023	0.039	0.038	0.024						
	95%CI	(0.83,0.92)	(0.78,0.93)	(0.82,0.97)	(0.83,0.92)						
5	AVE	**0.880**	**0.872**	0.888	**0.880**	0	0	99	0	0	1
	SD	0.023	0.031	0.038	0.023						
	95%CI	(0.83,0.93)	(0.81,0.93)	(0.81,0.96)	(0.83,0.93)						
6	AVE	0.858	0.839	0.878	0.858	0	0	78	0	3	19
	SD	0.026	0.039	0.041	0.026						
	95%CI	(0.81,0.91)	(0.76,0.92)	(0.80,0.96)	(0.81,0.91)						
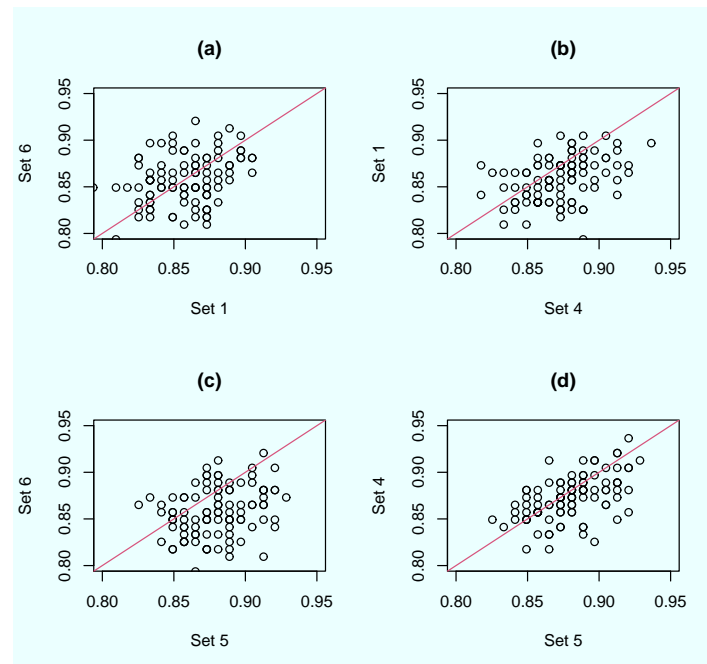
Figure 8. Scatter plots of the ACCs of different feature sets for test data based on 100 rounds: (**a**) Set 6 vs. Set 1, (**b**) Set 1 vs. Set 4, (**c**) Set 6 vs. Set 5, (**d**) Set 4 vs. Set 5, where the red line in each plot denotes the 45-degree line.

5. Discussion

A shape approximation approach is proposed to portray the ROI from medical images. An effective PSO-FS algorithm for obtaining a suitable approximation is proposed under the framework of PSO. The convergence of the PSO-FS algorithm is derived under mild assumptions on the selected family of equations. We apply the proposed procedure to the issue of detecting PD from the SPECT, where ellipse and cashew-shaped equations are selected to approximate the ROI in the 2D and 3D spaces, respectively. The coefficients of the ellipse and cashew-shaped equations, together with some well-known features of PD existing in the literature, are employed to learn PD classifiers under 6 commonly used machine learning approaches. The empirical results reveal that the coefficients of the ellipse and the cashew-shaped equations are capable of improving classification performances.

According to the findings in our empirical results, the proposed method has shown its potential to be applied to other medical imaging data to extract important features contained in the corresponding 2D and 3D images. Most importantly, once a suitable family of equations is selected to depict the regions of uptakes of a specific ROI based on the suggestions made by experts, the PSO-FS algorithm can help to obtain the optimal equation effectively if the selected family of equations satisfies some mild conditions. Therefore, more constructive and useful features can be computed according to the corresponding mathematical representations of the 2D and 3D shapes, which provide researchers the opportunities to gain more insights into the associated problems.

Moreover, although our empirical study provides numerical evidence to show the usefulness of the 2D and 3D shape approximations of the ROI for PD identification, the proposed method still has the following limitations:

- For a specific ROI, the proposed method needs to select a suitable family of mathematical representation, like the ellipse and cashew-shaped equations for identifying PD in 2D and 3D spaces, respectively. However, we may not have enough understanding or knowledge to derive useful features directly from a mathematical representation of any shape of the ROI, like the cashew-shaped equation used in this study. Therefore, we just adopt the coefficients of the shape equation as features for classification. Although these coefficients can represent the shape of the

ROI, it may not be the most effective way for a suitable classifier to learn. More studies are needed to dig more insights for this issue.

- This study and [8] both adopted the family of ellipse equations to portray the ROI observed from an 2D-combined image for PD identification. However, the ROI for normal subjects should be comma-shaped, which is not an ellipse. Due to the reason that we know ellipses better than comma-shaped equations, the family of ellipse equations is selected to approximate the ROI. This selection, of course, has systematic biases (or called model risk), which might reduce the classification performance.

- In the proposed PSO-FS algorithm, we need to compute the distance of each boundary point of the ROI to a given shape equation in each PSO iteration. This procedure is computationally expensive, especially when training a classifier with a huge number of boundary points. A more efficiently computing way should be developed in the future.

Author Contributions: Conceptualization, S.-F.H., Y.-H.W., C.-H.C., and C.-C.H.; methodology, S.-F.H. and Y.-H.W.; validation, S.-F.H., Y.-H.W., C.-H.C., and C.-C.H.; formal analysis, S.-F.H., Y.-H.W., C.-H.C., and C.-C.H.; Resources, C.-H.C. and C.-C.H.; data curation, C.-H.C. and C.-C.H.; writing—original draft preparation, S.-F.H. and Y.-H.W.; writing—review and editing, S.-F.H. and Y.-H.W.; visualization, S.-F.H. and Y.-H.W.; supervision, S.-F.H. and C.-C.H.; funding acquisition, S.-F.H. and C.-H.C. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Chang Gung Medical Foundation Institutional Review Board (no. 202001334B0). The IRB approves the waiver of the participants' consent.

Appendix A. Derivation of the Cashew-Shaped Equation (5) Satisfying C2

The objective is to show that at the m-th iteration of the FS algorithm with the cashew-shaped equation $\Psi_{3D} = 0$ defined in (5), if $d(\boldsymbol{p}, \mathbf{x}_m) > d(\boldsymbol{p}, H)$, there exists an $\tilde{\mathbf{x}}_m \in J$ such that $d(\mathbf{x}_m, \tilde{\mathbf{x}}_m) = \zeta$ and $\boldsymbol{p}$, $\mathbf{x}_m$, and $\tilde{\mathbf{x}}_m$ are not collinear. Moreover, based on the $\tilde{\mathbf{x}}_m$, one can define a $\kappa_{\mathbf{x}_m}$ as the intersection of $\{\mathbf{x} \in J : d(\mathbf{x}_m, \mathbf{x}) \leq \zeta\}$ and the plane containing $\boldsymbol{p}$, $\mathbf{x}_m$, and $\tilde{\mathbf{x}}_m$, where $d(\boldsymbol{p}, \mathbf{x}_m) > d(\boldsymbol{p}, \kappa_{\mathbf{x}_m})$ and C2 holds.

Without loss of generality, we consider $\Psi_{3D}(\cdot)$ with $s = 1$ and $(t_x, t_y, t_z) = (0, 0, 0)$ in (5), and consider the case of $\boldsymbol{p}$ and $\mathbf{x}_m$ having the same y-coordinate, where $\boldsymbol{p}$ and $\mathbf{x}_m$ are denoted by (p_x, c_y, p_z) and (m_x, c_y, m_z), respectively. Since $d(\boldsymbol{p}, \mathbf{x}_m) > d(\boldsymbol{p}, H)$, which indicates that the vector $\boldsymbol{p} - \mathbf{x}_m = (p_x - m_x, 0, p_z - m_z)$ is not orthogonal to the tangent plane to $\Psi_{3D} = 0$ at $\mathbf{x}_m$. Hence, there must exist one 2D plane, denoted by L_m, such that $\boldsymbol{p}, \mathbf{x}_m \in L_m$, and the vector $(p_x - m_x, 0, p_z - m_z)$ in L_m is not orthogonal to the tangent line to $\Psi_{3D} = 0$ restricted in L_m at $\mathbf{x}_m$, which indicates that $d(\boldsymbol{p}, \mathbf{x}_m) > d(\boldsymbol{p}, \hat{J}_m)$, where $\hat{J}_m$ denotes the intersection of $\Psi_{3D} = 0$ and L_m. The rest of the proof is to show that there exists an $\tilde{\mathbf{x}}_m$ in $\hat{J}_m$ such that C2 holds.

To achieve this goal, we first rotate the coordinate system such that the equation of the plane L_m is represented by $y = c_y$ in the new coordinate system with a rotation matrix M_1, which is defined similarly to the M in (5). Therefore, $\boldsymbol{p}$, $\mathbf{x}_m$ and $\tilde{\mathbf{x}}_m$ have the same y-coordinate in the new system. For example, Figure A1a presents a schematic diagram, where $\mathbf{x}_m$ and $\tilde{\mathbf{x}}_m$ are two points on a cashew-shaped surface $\Psi_{3D} = 0$, and $\boldsymbol{p}$ is an external point of $\{\mathbf{x} : \Psi_{3D} \leq 0\}$. If we rotate $\Psi_{3D} = 0$ to a new coordinate system, denoted by $\hat{\mathbf{x}} = (\hat{x}, \hat{y}, \hat{z})$ as shown in Figure A1b, with a proper rotation matrix, we have that $\boldsymbol{p}$, $\mathbf{x}_m$, and $\tilde{\mathbf{x}}_m$ have the same $\hat{y}$ values. Apparently, to prove that $\Psi_{3D} = 0$ satisfies C2 in Figure A1a is equivalent to prove it under the scenario in Figure A1b. In addition, since $\Psi_{3D}(\cdot)$ is transformed from $\Psi_{3D}^0(\cdot)$ defined in (4), let $\hat{\mathbf{x}}^{\mathsf{T}} = M_1 \cdot M \cdot \mathbf{x}^{\mathsf{T}}$ denote the coordinate in the $\hat{\mathbf{x}}$-coordinate system and the corresponding equation of $\hat{J}_m$ is represented as $\Psi_{3D}^0(M^{-1} \cdot M_1^{-1} \cdot (\hat{x}, c_y, \hat{z})^{\mathsf{T}}) = 0$, which is a polynomial equation of $\hat{x}$ and $\hat{z}$. Consequently, for a fixed $\hat{p}_x$, which is the x-coordinate of $\boldsymbol{p}$ in the $\hat{\mathbf{x}}$-coordinate system, there exists a $\zeta > 0$ such that $\tilde{\mathbf{x}}_m = (\tilde{m}_x, c_y, \tilde{m}_z) \in \hat{J}_m$, $d(\mathbf{x}_m, \tilde{\mathbf{x}}_m) = \zeta$, and C2 holds

in the $\hat{\mathbf{x}}$-coordinate system, which indicates that C2 also holds in the original $\mathbf{x}$-coordinate system since $\mathbf{x}^\mathsf{T} = (M_1 M)^{-1} \hat{\mathbf{x}}^\mathsf{T}$. Therefore, the cashew-shaped equations defined in (5) also satisfy C2.

(a)   (b)
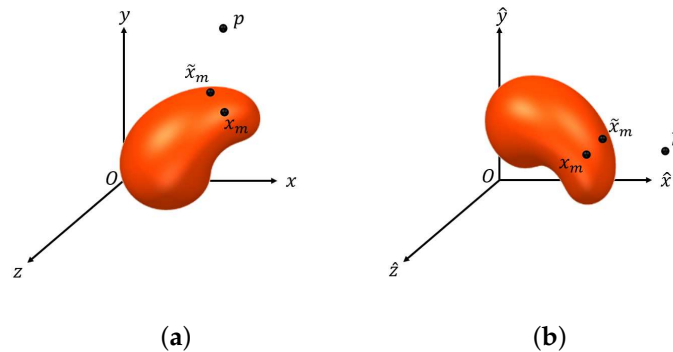
Figure A1. A schematic diagram to show the case of (**a**) $\mathbf{x}_m$ and $\tilde{\mathbf{x}}_m$ being two points on a cashew-shaped surface $\Psi_{3D} = 0$, and $\boldsymbol{p}$ being an external point of $\{\mathbf{x} : \Psi_{3D} \leq 0\}$. (**b**) If we rotate $\Psi_{3D} = 0$ in (**a**) to a new coordinate system, denoted by $(\hat{x}, \hat{y}, \hat{z})$, with a proper rotation matrix, we have that $\boldsymbol{p}$, $\mathbf{x}_m$, and $\tilde{\mathbf{x}}_m$ have the same $\hat{y}$ values.

Before presenting the proof of Theorem 1, we derive the following lemmas.

Lemma A1. *In Step 1 of the FS algorithm, if $\Psi(\cdot)$ is continuous, there exists an $r_\epsilon > 0$ such that for all $r_\epsilon < r < d(\boldsymbol{p}, H)$, $\mathbf{x}_0 \in J$.*

Proof. If $\boldsymbol{p} \in H$, then $\Psi(\boldsymbol{p}) = 0$ and we have $d(p, H) = 0$. In the following, we only proof the case of $\boldsymbol{p} \notin H$ and $\Psi(\boldsymbol{p}) > 0$ since the case of $\Psi(\boldsymbol{p}) < 0$ can be obtained by using similar arguments. For the case of $\Psi(\boldsymbol{p}) > 0$, there exist some $0 < r < d(\boldsymbol{p}, H)$ such that

$$\overline{B}_r(\boldsymbol{p}) \cap H = \varnothing \text{ and } \Psi(\mathbf{y}) > 0 \text{ for all } \mathbf{y} \in \overline{B}_r(\boldsymbol{p}), \tag{A1}$$

where $\overline{B}_r(\boldsymbol{p})$ is a closed ball of radius $r > 0$ centered at $\boldsymbol{p}$. Let

$$\mathbf{y}_r = \arg\min_{\mathbf{y}}\{\Psi(\mathbf{y}) : \mathbf{y} \in \overline{B}_r(\boldsymbol{p})\} \tag{A2}$$

denote a point in $\overline{B}_r(\boldsymbol{p})$ having the smallest value of $\Psi(\cdot)$. Note that $\mathbf{y}_r$ might not be unique and any $\mathbf{y}_r$ should be in $\overline{B}_r^b(\boldsymbol{p})$, which denotes the set of boundary points of $\overline{B}_r(\boldsymbol{p})$. Define a new equation $\Psi_r(\mathbf{x}) = \Psi(\mathbf{x}) - \Psi(\mathbf{y}_r) = 0$ and let $H_r = \{\mathbf{x} : \Psi_r(\mathbf{x}) = 0\}$. Then, we have $\mathbf{y}_r \in H_r$ and $d(\boldsymbol{p}, H_r) = d(\boldsymbol{p}, \mathbf{y}_r) = r$ for all $\mathbf{y}_r$ satisfying (A2) since $\Psi_r(\mathbf{y}) = 0$ for $\mathbf{y} \in \overline{B}_r(\boldsymbol{p}) \cap H_r$ and $\Psi_r(\mathbf{y}) > 0$ for $\mathbf{y} \in \overline{B}_r(\boldsymbol{p}) \cap H_r^c$, where H_r^c denotes the complement of H_r. In other words, we have

$$d(\boldsymbol{p}, H_r) \to d(\boldsymbol{p}, H), \text{ as } r \uparrow d, \tag{A3}$$

where $d = d(\boldsymbol{p}, \mathbf{x}^*) = d(p, H)$. Consequently, for a large enough r satisfying (A1), let $\mathbf{y}_r^* = c(\mathbf{y}_r - \boldsymbol{p}) + \boldsymbol{p}$ with $c > 1$ and $\Psi(\mathbf{y}_r^*) = 0$, which indicate that $\mathbf{y}_r^* \in H$ and $\mathbf{y}_r^*$, $\mathbf{y}_r$ and $\boldsymbol{p}$ are located in the same line as defined in (10), and $\mathbf{y}_r$ is in-between $\boldsymbol{p}$ and $\mathbf{y}_r^*$. By (A3) and the definition of $\mathbf{y}_r^*$, we have $\lim_{r \uparrow d} d(\boldsymbol{p}, \mathbf{y}_r^*) = \lim_{r \uparrow d} d(\boldsymbol{p}, \mathbf{y}_r) = d$. This fact together with the continuity of $\Psi(\cdot)$ leads to $\lim_{r \uparrow d} \mathbf{y}_r^* \in D_p$, where D_p is defined in (9). Therefore, for all $\epsilon > 0$, there exists an $r_\epsilon \in [0, d]$ such that for all $r \geq r_\epsilon$, we have $\mathbf{x}_0 = \mathbf{y}_r^* \in J$. $\quad\square$

The result of Lemma A1 indicates that the Step 1 of the FS algorithm can find an initial point in the neighborhood, $\bigcup_{\mathbf{x}^* \in D_p} \overline{B}_\epsilon(\mathbf{x}^*)$, of those $\mathbf{x}^*$ in D_p.

Lemma A2. *Assume that C1 and C2 hold. Then we have*

1. *G is continuous on J_i, $i = 1, 2, \ldots$.*
2. *If $x \in J$ and $d(p, x) > d(p, H)$, there exists another point $x' \in J_x \subseteq J$ such that $d(p, x') < d(p, x)$.*

Proof. Without loss of generality, we consider the case of $v_\mathbf{x}$ and $n_\mathbf{x}$ being normalized vectors. Hence, we have

$$G(\mathbf{x}) = \cos(\theta_\mathbf{x}) = |{<}v_\mathbf{x}, n_\mathbf{x}{>}|, \text{ for all } \mathbf{x} \in J, \tag{A4}$$

where $<a, b>$ denotes the inner product of two vectors a and b. For any $\mathbf{x}_0 \in J$, the triangular inequality leads to

$$\begin{aligned}
|G(\mathbf{x}) - G(\mathbf{x}_0)| &\leq |{<}v_\mathbf{x}, n_\mathbf{x} - n_{\mathbf{x}_0}{>}| + |{<}v_\mathbf{x} - v_{\mathbf{x}_0}, n_{\mathbf{x}_0}{>}| \\
&\leq ||n_\mathbf{x} - n_{\mathbf{x}_0}|| + ||v_\mathbf{x} - v_{\mathbf{x}_0}||.
\end{aligned} \tag{A5}$$

In addition, C1 leads to for all $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$, there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$||v_\mathbf{x} - v_{\mathbf{x}_0}|| < \varepsilon_1, \text{ for all } ||\mathbf{x} - \mathbf{x}_0|| < \delta_1, \tag{A6}$$

and

$$||n_\mathbf{x} - n_{\mathbf{x}_0}|| < \varepsilon_2, \text{ for all } ||\mathbf{x} - \mathbf{x}_0|| < \delta_2, \tag{A7}$$

where $||\mathbf{x} - \mathbf{x}_0|| < \delta_1$ (or $||\mathbf{x} - \mathbf{x}_0|| < \delta_2$) indicates that $\mathbf{x}$ belongs to the J_i containing $\mathbf{x}_0$. Consequently, let $\varepsilon_1 = \varepsilon_2 = \varepsilon/2$ and $\delta = \min(\delta_1, \delta_2)$. By (A5)–(A7), we have that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|G(\mathbf{x}) - G(\mathbf{x}_0)| < \varepsilon$ for all $||\mathbf{x} - \mathbf{x}_0|| < \delta$. Therefore, G is a continuous function defined on J_i, $i = 1, 2, \ldots$.

For a point $\mathbf{x} \in J$, let $\Gamma_\mathbf{x}$ denote the tangent plane to Ψ at $\mathbf{x}$. The existence of $\Gamma_\mathbf{x}$ is provided by C1. In addition, if $\mathbf{x}$ is an interior point of J, there exists a $\zeta > 0$ such that $B_\zeta^2(\mathbf{x}) \cap H \subseteq J$, where $B_\zeta^{(2)}(\mathbf{x})$ denotes a 2-dimensional open ball of radius ζ centered at $\mathbf{x}$. Denote $B_\zeta^{(2)}(\mathbf{x}) \cap H$ by $\kappa_\mathbf{x}$ for short. In particular, if $\mathbf{x}$ is a boundary point of J, we defined $\kappa_\mathbf{x}$ as the intersection of $B_\zeta^{(2)}(\mathbf{x}) \cap H$ and J. Furthermore, if $d(p, \mathbf{x}) \neq d(p, H)$, we have $G(\mathbf{x}) < 1$. Consequently, there exists an $\tilde{\mathbf{x}} \in \Gamma_\mathbf{x}$ such that $G(\mathbf{x}) < G(\tilde{\mathbf{x}})$, which is equivalent to

$$d(p, \tilde{\mathbf{x}}) < d(p, \mathbf{x}), \tag{A8}$$

and an $\tilde{\mathbf{x}}' \in \kappa_\mathbf{x}$ associated to $\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}}'$ denotes the intersection of $\kappa_\mathbf{x}$ and the line formed by $\tilde{\mathbf{x}}$ and p.

According to C2, we consider the following two cases:

C2.1 $\Psi\left(\alpha\mathbf{x} + (1 - \alpha)\tilde{\mathbf{x}}'\right)\Psi(p) > 0$, for $\alpha \in [0, 1]$;

C2.2 $\Psi\left(\alpha\mathbf{x} + (1 - \alpha)\tilde{\mathbf{x}}'\right)\Psi(p) \leq 0$, for $\alpha \in [0, 1]$.

Figure A2a,b present schematic diagrams for C2.1 and C2.2, respectively. First, C2.1 and (A8) lead to

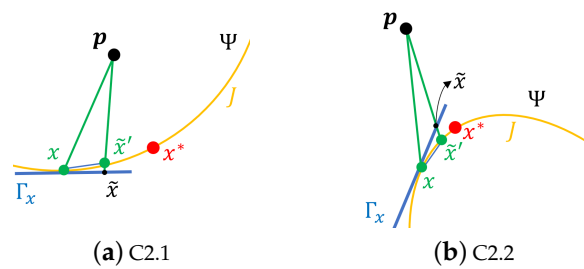$$d(p, \tilde{\mathbf{x}}') < d(p, \tilde{\mathbf{x}}) < d(p, \mathbf{x}) \tag{A9}$$

directly. On the other hand, if C2.2 holds, there exists a point $\mathbf{y} \in \overline{B}_\zeta(\mathbf{x})$, such that $\Psi(\mathbf{y})\Psi(p) < 0$ and $<p - \mathbf{y}, \mathbf{x} - \mathbf{y}> = 0$. That is, p, $\mathbf{x}$, and $\mathbf{y}$ form a right triangle, which indicates

$$d(p, \tilde{\mathbf{x}}') < d(p, \mathbf{y}) < d(p, \mathbf{x}), \tag{A10}$$

where $\tilde{\mathbf{x}}'$ is redefined as the intersection of $\kappa_\mathbf{x}$ and the line formed by p and $\mathbf{y}$. Therefore, by (A9) and (A10), we conclude that there exists an $\tilde{\mathbf{x}}' \in \kappa_\mathbf{x}$ such that $d(p, \tilde{\mathbf{x}}') < d(p, \mathbf{x})$ if C1 and C2 hold.

(a) C2.1   **(b)** C2.2

Figure A2. The 2 cases of C2. For any point $\mathbf{x}$ on J, there exists a point $\tilde{\mathbf{x}}' \in J$ associated to $\tilde{\mathbf{x}}$ which on the tangent plane Γ_x to Ψ at $\mathbf{x}$, for $\alpha \in [0,1]$, such that **(a)** $\Psi\left(\alpha\mathbf{x} + (1-\alpha)\tilde{\mathbf{x}}'\right)\Psi(\boldsymbol{p}) > 0$ or **(b)** $\Psi\left(\alpha\mathbf{x} + (1-\alpha)\tilde{\mathbf{x}}'\right)\Psi(\boldsymbol{p}) \leq 0$.
□

The second result of Lemma A2 indicates that in the Step 2 of the FS algorithm, one can find a better estimate of $d(\boldsymbol{p}, H)$ than $d(\boldsymbol{p}, \boldsymbol{x})$ if $d(\boldsymbol{p}, \boldsymbol{x}) > d(\boldsymbol{p}, H)$ by setting a small enough δ. According to Lemmas A1 and A2, we prove Theorem 1 in the following.

Proof of Theorem 1. By Lemma 1, an initial estimate $\mathbf{x}_0 \in J$ can be obtained. In addition, the second result in Lemma 2 indicates that $\{G(\mathbf{x}_m) : m = 0, 1, \ldots\}$ is a monotone increasing sequence. Due to the fact that the sequence $\{G(\mathbf{x}_m) : m = 0, 1, \ldots\}$ is also bounded above since $G(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathbb{R}^d$, Monotone Convergence Theorem indicates that the sequence $\{G(\mathbf{x}_m) : m = 0, 1, \ldots\}$ converges. This fact together with the continuity property of G given in the first result of Lemma 2 leads to $\lim_{m \to \infty} G(\mathbf{x}_m) = 1$, which is equivalent to $\lim_{m \to \infty} d(\boldsymbol{p}, \mathbf{x}_m) = d(\boldsymbol{p}, H)$. The proof is complete. □

References

1. Prashanth, R.; Roy, S.D.; Mandal, P.K.; Ghosh, S. Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. *Expert Syst. Appl.* **2014**, *41*, 3333–3342. [CrossRef]
2. Taylor, J.C.; Fenner, J.W. Comparison of machine learning and semi-quantification algorithms for (I123) FP-CIT classification: The beginning of the end for semi-quantification? *EJNMMI Phys.* **2017**, *4*, 29. [CrossRef] [PubMed]
3. Oliveira, F.P.; Faria, D.B.; Costa, D.C.; Castelo-Branco, M.; Tavares, J.M.R. Extraction, selection and comparison of features for an effective automated computer-aided diagnosis of Parkinson's disease based on [123 I] FP-CIT SPECT images. *Eur. J. Nucl. Med. Mol. Imaging* **2018**, *45*, 1052–1062. [CrossRef]
4. Nicastro, N.; Wegrzyk, J.; Preti, M.G.; Fleury, V.; Van de Ville, D.; Garibotto, V.; Burkhard, P.R. Classification of degenerative parkinsonism subtypes by support-vector-machine analysis and striatal 123 I-FP-CIT indices. *J. Neurol.* **2019**, *266*, 1771–1781. [CrossRef] [PubMed]
5. Iwabuchi, Y.; Nakahara, T.; Kameyama, M.; Yamada, Y.; Hashimoto, M.; Matsusaka, Y.; Osada, T.; Ito, D.; Tabuchi, H.; Jinzaki, M. Impact of a combination of quantitative indices representing uptake intensity, shape, and asymmetry in DAT SPECT using machine learning: Comparison of different volume of interest settings. *EJNMMI Res.* **2019**, *9*, 7. [CrossRef] [PubMed]
6. Ziegler, D.A.; Wonderlick, J.S.; Ashourian, P.; Hansen, L.A.; Young, J.C.; Murphy, A.J.; Koppuzha, C.K.; Growdon, J.H.; Corkin, S. Substantia nigra volume loss before basal forebrain degeneration in early Parkinson disease. *JAMA Neurol.* **2013**, *70*, 241–247. [CrossRef] [PubMed]
7. Palumbo, B.; Fravolini, M.L.; Buresta, T.; Pompili, F.; Forini, N.; Nigro, P.; Calabresi, P.; Tambasco, N. Diagnostic accuracy of Parkinson disease by support vector machine (SVM) analysis of 123I-FP-CIT brain SPECT data: Implications of putaminal findings and age. *Medicine* **2014**, *93*, e228. [CrossRef]
8. Prashanth, R.; Roy, S.D.; Mandal, P.K.; Ghosh, S. High-accuracy classification of Parkinson's disease through shape analysis and surface fitting in ^{123}I-Ioflupane SPECT imaging. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 794–802. [CrossRef]

9. Hsu, S.Y.; Lin, H.C.; Chen, T.B.; Du, W.C.; Hsu, Y.H.; Wu, Y.C.; Tu, P.W.; Huang, Y.H.; Chen, H.Y. Feasible classified models for Parkinson disease from 99mTc-TRODAT-1 SPECT imaging. *Sensors* **2019**, *19*, 1740. [CrossRef]

10. Cheng, Z.; Zhang, J.; He, N.; Li, Y.; Wen, Y.; Xu, H.; Tang, R.; Jin, Z.; Haacke, M.; Yan, F.; et al. Radiomic features of the Nigrosome-1 region of the substantia nigra: Using quantitative susceptibility mapping to assist the diagnosis of idiopathic Parkinson's disease. *Front. Aging Neurosci.* **2019**, *11*, 167. [CrossRef]

11. Xu, R.; Hu, X.; Jiang, X.; Zhang, Y.; Wang, J.; Zeng, X.; Longitudinal volume changes of hippocampal subfields and cognitive decline in Parkinson's disease. *Quant. Imaging Med. Surg.* **2020**, *10*, 220–232. [CrossRef] [PubMed]

12. Kalia, L.V.; Lang, A.E. Parkinson's disease. *Lancet* **2015**, *386*, 896–912. [CrossRef]

13. Hayes, M.T. Parkinson's disease and Parkinsonism. *Am. J. Med.* **2019**, *132*, 802–807. [CrossRef] [PubMed]

14. Kung, H.F.; Kung, M.P.; Wey, S.P.; Lin, K.J.; Yen, T.C. Clinical acceptance of a molecular imaging agent: A long march with [99mTc] TRODAT. *Nucl. Med. Biol.* **2007**, *34*, 787–789. [CrossRef]

15. Bajaj, N.; Hauser, R.A.; Grachev, I.D. Clinical utility of dopamine transporter single photon emission CT (DaT-SPECT) with (123I) ioflupane in diagnosis of parkinsonian syndromes. *J. Neurol. Neurosurg. Psychiatry* **2013**, *84*, 1288–1295. [CrossRef]

16. Törn, A.; Žilinskas, A. *Global Optimization*; Springer: Berlin, Germany, 1989.

17. Eberhart, R.; Kennedy, J. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; Volume 4, pp. 1942–1948.

18. Delgarm, N.; Sajadi, B.; Kowsary, F.; Delgarm, S. Multi-objective optimization of the building energy performance: A simulation-based approach by means of particle swarm optimization (PSO). *Appl. Energy* **2016**, *170*, 293–303. [CrossRef]

19. Mistry, K.; Zhang, L.; Neoh, S.C.; Lim, C.P.; Fielding, B. A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition. *IEEE Trans. Cybern.* **2016**, *47*, 1496–1509. [CrossRef] [PubMed]

20. Hasanipanah, M.; Naderi, R.; Kashir, J.; Noorani, S.A.; Qaleh, A.Z.A. Prediction of blast-produced ground vibration using particle swarm optimization. *Eng. Comput.* **2017**, *33*, 173–179. [CrossRef]

21. Kumar, A.; Pant, S.; Ram, M.; Singh, S.B. On solving complex reliability optimization problem using multi-objective particle swarm optimization. In *Mathematics Applied to Engineering*; Academic Press: London, UK, 2017; pp. 115–131.

22. Morera, D.M.; Sarlabous, J.E. On the distance from a point to a quadric surface. *Investig. Oper.* **2003**, *24*, 153–161.

23. Rovini, E.; Maremmani, C.; Moschetti, A.; Esposito, D.; Cavallo, F. Comparative motor pre-clinical assessment in Parkinson's disease using supervised machine learning approaches. *Ann Biomed. Eng.* **2018**, *46*, 2057–2068. [CrossRef]

24. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

25. Huang, S.F.; Lu, H.P. Classification of temporal data using dynamic time warping andcompressed learning. *Biomed. Signal Process. Control* **2020**, *57*, 101781. [CrossRef]

26. Dolz J.; Desrosiers, C; Ayed, I.B. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage* **2018**, *170*, 456–470. [CrossRef] [PubMed]

27. Carr, J.C.; Fright, W.R.; Beatson, R.K. Surface interpolation with radial basis functions for medical imaging. *IEEE Trans. Med. Imaging* **1997**, *16*, 96–107. [CrossRef] [PubMed]

28. Pu, J.; Zheng, B.; Leader, J.K.; Fuhrman, C.; Knollmann, F.; Klym, A.; Gur, D. Pulmonary lobe segmentation in CT examinations using implicit surface fitting. *IEEE Trans. Med. Imaging* **2009**, *28*, 1986–1996. [PubMed]

29. Huang, A.; Summers, R.M.; Hara, A.K. Surface curvature estimation for automatic colonic polyp detection. In Proceedings of the Medical Imaging 2005: Physiology, Function, and Structure from Medical Images, San Diego, CA, USA, 12–17 February 2005; pp. 393–402.

30. Hu, X.; Eberhart, R. Solving constrained nonlinear optimization problems with particle swarm optimization. In Proceedings of the Sixth World Multiconference on Systemics, Cybernetics and Informatics, Orlando, FL, USA, 14–18 July 2002; Volume 5, pp. 203–206.

31. Robinson, J.; Rahmat-Samii, Y. Particle swarm optimization in electromagnetics. *IEEE Trans. Antennas Propag.* **2004**, *52*, 397–407. [CrossRef]

32. Sumardi, M.S.; Riyadi, M.A. Particle swarm optimization (PSO)-based self tuning proportional integral derivative (PID) for bearing navigation control system on quadcopter. In Proceedings of the 4th International Conference on Information Technology, Computer, and Electrical Engineering, Semarang, Indonesia, 18–19 October 2017; pp. 181–186.

33. Kuncheva, L.I.; Arnaiz-Gonzalez, A.; Diez-Pastor, J.F.; Gunn, I.A. Instance selection improves geometric mean accuracy: A study on imbalanced data classification. *Prog. Artif. Intell.* **2019**, *8*, 215–228. [CrossRef]