Heterogeneity by global and textural feature analysis in F-18 FP-CIT brain PET images for diagnosis of Parkinson's disease

Hyun Jin Yoon, PhD[a,b] ⓘ, Kook Cho, PhD[c], Woong Gon Kim, BS[d], Young-Jin Jeong, MD[a,b], Ji-Eun Jeong, MD[a,b], Do-Young Kang, MD PhD[a,b,e,*]

Abstract

Background: The quantification of heterogeneity for the striatum and whole brain with F-18 FP-CIT PET images will be useful for diagnosis. The index obtained from texture analysis on PET images is related to pathological change that the neuronal loss of the nigrostriatal tract is heterogeneous according to the disease state. The aim of this study is to evaluate various heterogeneity indices of F-18 FP-CIT PET images in the diagnosis of Parkinson's disease (PD) patients and to access the diagnostic accuracy of the indices using machine learning (ML).

Methods: This retrospective study included F-18 FP-CIT PET images of 31 PD and 31 age-matched health controls (HC). The volume of interest was delineated according to iso-contour lines around standardized uptake value (SUV) 3.0 g/ml for each region of the striatum by PMod 3.603. One hundred eight heterogeneity indices were calculated using CGITA to find indices from which the PD and HC were classified using statistical significance. PD group was classified by constructing a 2-dimensional or 3-dimensional phase space quantifier using these heterogeneity indices. We used 71 heterogeneity indices to classify PD from HC using ML for dimensional reduction.

Results: The heterogeneity indices for classifying PD from HC were size-zone variability, contrast, inverse difference-moment, and homogeneity in the order of low *P* value. Three-dimensional quantifiers composed of normalized-contrast, code-similarity, and contrast were more clearly classified than 2-dimensional ones. After 71-dimensional reduction using PCA, classification was possible by logistic regression with 91.3% accuracy. The 2 groups were classified with an accuracy of 85.5% using the support vector machine and 88.4% using the random forest. The classification accuracy using the eXtreme Gradient Boosting was 95.7%, and feature importance was highest in order of SUV bias-corrected kurtosis, size-zone-variability, intensity-variability, and high-intensity-zone-variability.

Conclusion: It was confirmed that PD patients is more clearly classified than the conventional 2-dimensional quantifier by introducing a 3-dimensional phase space quantifier. We observed that ML can be used to classify the 2 groups in an easy and explanatory manner. For the discrimination of the disease, 24 heterogeneity indices were found to be statistically useful, and the major cut-off values of 3 heterogeneity indices were size-zone variability (1906.44), intensity variability (129.21), and high intensity zone emphasis (800.29).

Abbreviations: CGITA = Chang-Gung image texture analysis, FP-CIT = F-18 fluoro-propyl-carbomethoxyiodophenyl-tropane, HC = health controls, LR = logistic regression, ML = machine learning, PD = Parkinson's disease, RF = random forest, SD = standard deviation, SUV = standardized uptake value, SVM = support vector machine, VOI = volume of interest, XGBoost = eXtreme Gradient Boosting.

Keywords: F-18 FP-CIT PET, heterogeneity, machine learning (ML), Parkinson's disease (PD), textural feature analysis

---

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder caused by degenerative changes in dopamine neurons in the *substantia nigra* region and the loss of dopamine neuron terminals in the striatum.[1] Degenerative brain diseases, such as PD, are found after the disease progresses, because the symptoms are initially hidden by a compensatory mechanism, so early detection and differentiation from similar diseases are important for treatment. At present, the diagnosis of PD depends on clinical findings and lacks quantitative indicators for early diagnosis, so there can be failure of diagnosis and evaluation and tracking of disease progression. Non-invasive measurement of neurochemical changes due to degeneration of dopamine neurons using positron emission tomography (PET) images of PD patients is now possible.[2] The radio-pharmaceutical F-18 fluoro-propyl-carbo-methoxyiodophenyl-tropane (FP-CIT), which has a high affinity for dopamine transporters located in the pre-synaptic dopamine neuron membrane, has been used in Korea for the diagnosis of PD since 2008.

F-18 FP-CIT PET dopamine transporter images of PD show reduced uptake of radio-pharmaceuticals in the striatum. The loss of *putamen* dopamine transporters also occurs first, while the caudate nucleus changes later, as shown in Figure 1.[3,4] In the normal group, uptake in the striatum, from dopamine transporter images, decreased by 5% to 8% per 10 years as age increased.[5,6] 5.7% to 14.7% of patients with early PD were reported as normal on dopamine transporter imaging.[7,8] Some patients have symptoms similar to PD, which are mistaken for PD and are not amenable to treatment with anti-PD drugs.[9] With dopamine transporter imaging alone, it is difficult to distinguish PD from nonspecific degenerative Parkinson's syndrome.[10] Accurately classifying PD from normal can be helpful in treatment.

Recently, tumor heterogeneity has been quantified by molecular imaging PET because it is an important biomarker for aggressive outcomes and disease.[11] The "heterogeneity index" was calculated numerically for quantification of tumor heterogeneity. Evaluation of the clinical application of the heterogeneity index should be validated through several trials.[12] Although Chang-Gung image texture analysis (CGITA) software has been mainly studied in oncology applications, it can also be applied to nervous system applications.[11] CGITA analysis consists of 10 feature "parents" and 108 sub-features.[13–20] Feature parents have 10 items: co-occurrence (7 sub-features)[21] voxel-alignment (11),[14] neighborhood intensity-difference (5),[15] intensity-size-zone (11)[16] normalized co-occurrence (6),[21] standardized uptake value (SUV) statistics (49), texture spectrum (2),[17] texture feature coding (4),[18] texture feature coding co-occurrence (8)[18] and neighboring gray level dependence (5);[19] they are summarized in Table 1. In the present study, various heterogeneity indices were applied to F-18 FP-CIT brain PET images to classify the PD group from the normal group.

If applied to a well-defined and sufficient dataset using the machine learning (ML) technique, data classification will be
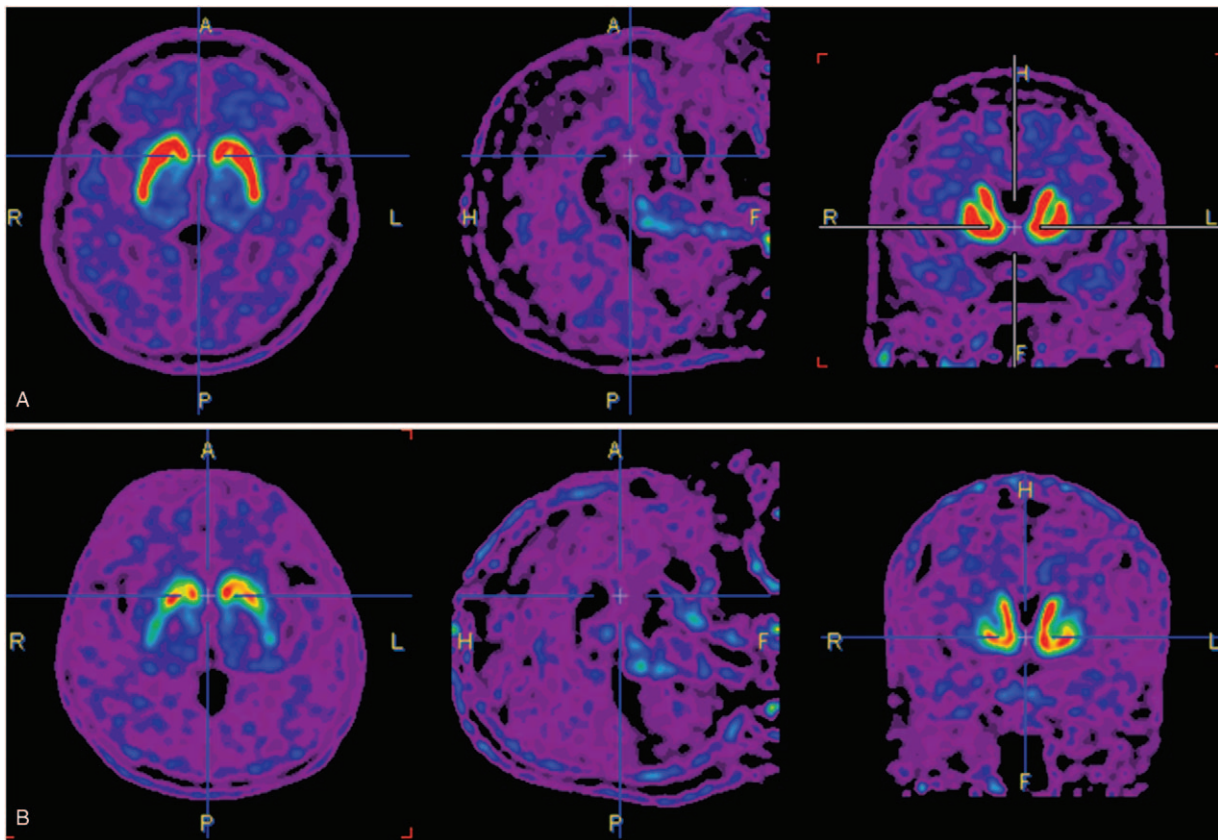

Figure 1. (A) Representative F-18 FP-CIT PET images in normal control and Parkinson's disease (PD). (B) SUV of F-18 FP-CIT PET image with PD is almost absent in both *caudate nuclei* and *putamina*.

Table 1

SUV statistics, texture feature parents and associated features of the heterogeneity indices are shown.

Feature parent	Feature name	P value	
Co-occurrence	Contrast	1.00029E–12	****
	Dissimilarity	3.75046E–10	****
Voxel-alignment	Short run emphasis	0.016343439	*
	Run-length variability	0.036887868	*
	High-intensity run emphasis	0.006793941	**
	High-intensity short-run emphasis	0.000149502	***
	High-intensity long-run emphasis	0.045119396	*
Neighborhood intensity-difference	Contrast	1.85714E–06	****
	Busyness	0.040487997	*
	Complexity	0.013769938	*
	Strength	0.000129852	***
Intensity-size-zone	Short-zone emphasis	2.6902E–09	****
	Intensity variability	0.002274678	**
	Size-zone variability	6.79209E–15	****
	Zone percentage	4.33351E–08	****
	High-intensity zone emphasis	0.000663492	***
	Low-intensity short-zone emphasis	0.018959745	*
	High-intensity short-zone emphasis	1.85965E–06	****
Normalized co-occurrence	Second angular moment	0.026808462	*
	Contrast	2.02287E–09	****
	Entropy	6.92433E–07	****
	Homogeneity	7.30236E–06	****
	Dissimilarity	4.30895E–09	****
	Inverse difference moment	1.36691E–05	****
SUV statistics	Minimum SUV	0.001157705	**
	Maximum SUV	2.7711E–07	****
	Mean SUV	5.32019E–08	****
	SUV variance	3.81441E–08	****
	SUV SD	3.22846E–10	****
	SUV skewness	4.30727E–07	****
	SUV bias-corrected skewness	4.29786E–07	****
	SUV bias-corrected kurtosis	1.29798E–05	****
	TLG	2.31906E–06	****
	Entropy	4.90619E–06	****
	SULpeak	3.19891E–07	****
	Asphericity 3	0.001787896	**
	Surface mean SUV 1	0.00029043	***
	Surface SUV variance 1	7.65935E–06	****
	Surface SUV SD 1	1.37894E–06	****
	Surface SUV NSR 1	6.50574E–07	****
	Surface SUV variance 2	1.9843E–06	****
	Surface SUV SD 2	0.013960269	*
	Surface SUV variance 3	1.6053E–05	****
	Surface SUV SD 3	0.000216963	***
	Surface SUV variance 4	6.38145E–07	****
	Surface SUV SD 4	1.23E–06	****
	SUVmean_prod_asphericity	7.95557E–05	****
	SUVmean_prod_surface_area	0.024683396	*
Texture spectrum	Max spectrum	2.37813E–08	****
	Black-white symmetry	2.7476E–08	****
Texture feature coding	Mean convergence	0.001550072	**
	Variance	0.013155872	*
Texture feature coding co-occurrence	Second angular moment	1.57702E–08	****
	Contrast	2.50733E–07	****
	Entropy	6.09593E–06	****
	Homogeneity	2.64231E–10	****
	Intensity	0.000717938	***
	Inverse difference moment	1.55045E–10	****
	Code entropy	6.09593E–06	****
	Code similarity	1.22492E–09	****
Neighboring gray level dependence	Small number emphasis	1.36827E–07	****
	Large number emphasis	0.038539334	*
	Number non-uniformity	1.53546E–06	****

Comparison of heterogeneity indices by SUV statistics and textural feature analysis in F-18 FP-CIT brain PET images with PD and HC.

NSR = radial noise-to-signal, SD = standard deviation, SULpeak = peak SUV corrected for lean body mass, SUV = standardized uptake value, TLG = total lesion glycolysis.

possible with optimal performance. Logistic regression (LR) is fast for learning, works well even in a small dataset, and can explain the reason for the prediction by formula. The support vector machine (SVM)[22] has a feature that works with both small and low and high datasets. Random forest (RF)[23] is a technique that assembles several decision trees to create an optimal model, and produces reliable results by reducing overfitting, rather than using 1 decision tree. XGBoost (eXtreme Gradient Boosting)[24,25] is a model based on classification and regression trees, and it performs best in the Kaggle competition. XGBoost, using a boosting algorithm, is capable of parallel processing, so processing speed is fast, optimization options are provided, automatic pruning using a "greedy-algorithm" is possible in order to reduce overfitting and it has excellent connectivity with other AI algorithms. In this study, the accuracy was evaluated using various ML algorithms to classify the PD group and the normal group.

2. Methods

2.1. Database of dopamine transporter brain PET images

Sixty-five patients who were clinically possible or probable IPD patients and normal health controls (HC) were retrospectively enrolled in the study. F-18 FP-CIT brain PET scans were used for analysis in 31 PD patients (mean age 67.8 ± 5.8 years old) and 31 age-matched HC (70.5 ± 8.5 years old). The number of data in 2 groups was matched based on age, and 3 PET images were excluded. PD patients were assessed with the British Parkinson's Disease Association Brain Bank clinical diagnostic criteria.[26] PD patients underwent the Seoul Neuropsychological Screening Battery consisting of 5 domains: a digit span test (forward and backward) for attention, a Korean version of the Boston naming test for language, the Rey complex figure test for visuo-spatial recognition, the Seoul verbal learning test for memory and the Korean – color ward Stroop test for executive function.[27] The criteria for the normal control group were: no abnormalities on neurological examinations, no underlying medical history or medication, and no abnormal findings on both F-18 FP-CIT PET and brain magnetic resonance imaging. The Institutional Review Board of our hospital reviewed and approved the study protocol and informed consent form.

Scanning was performed without interruption of dopamine agonist anti-Parkinson medications, such as *levodopa* to control motor symptoms of PD and especially for bradykinesia (slowness of movement) and rigidity. Drugs affecting cognitive function, such as anticholinergic drugs, were stopped for 2 days before the F-18 FP-CIT PET scan. This study was approved by the Dong-A Medical Center Ethics Committee and did not require prior consent because of its retrospective nature. Scanning was performed with a Biograph mCT flow PET/CT (Siemens), providing an in-plane spatial resolution of up to 5.0 mm fullwidth at half maximum at the center of the field of view. Patients were injected intravenously with 185 MBq F-18 FP-CIT and PET/CT acquisition was started 2 hours after radioisotope tracer injection.[28–30] A helical CT scan was performed with a rotational time of 0.5 seconds at 120 kVp and 100 mA without intravenous contrast agent. PET scan was followed by CT scan and acquired for 10 minutes in 3D mode. All images were acquired from the skull vertex to the base of the skull. Image reconstruction was performed with a point spread function and time-of-flight algorithm with a 400×400 matrix, 5 iterations and

21 subsets. Patients were able to continue anti-Parkinson medication.

2.2. Image analysis

The volume of interest (VOI) was delineated according to iso-contour lines around SUV 3.0 g/ml for each region of the striatum by PMod 3.603.[31] The heterogeneity index is calculated by loading VOI extracted using PMOD in CGITA.[32] CGITA 1.4 contains 108 heterogeneity indices that represent spatial non-uniformity. The values of textural feature indices were analyzed within VOI of 62 patients, with and without PD. The quantification of heterogeneity by molecular images has great clinical potential for diagnosis using medical images by analyzing local or global changes in the spatial arrangement of pixel intensities by texture analysis.[11] CGITA is a free open source software package for quantifying tumor heterogeneity in molecular images. Heterogeneity indices in F-18 FP-CIT brain PET images were calculated using CGITA.[32] One hundred eight heterogeneity indices from textural feature analysis were calculated using a co-occurrence matrix, voxel-alignment matrix, neighborhood intensity difference matrix, intensity-size-zone matrix, normalized co-occurrence matrix, texture spectrum, texture feature coding, texture feature coding co-occurrence matrix and neighborhood gray-level dependence, as shown in Table 1. One hundred eight heterogeneity indices are derived from the ten feature parent matrices. The values of global and textural feature indices (local and regional algorithms) were analyzed within VOI of 62 patients, with and without PD. The global indices were SUVmax, SUVmean, SUVSD, SUV skewness, SUV kurtosis, total lesion glycolysis, tumor volume, entropy, SUVpeak, as shown in Table 1.

2.3. Statistical and ML analysis

To classify the 2 groups, the correlations between each variable were examined and *t* tests were performed to calculate *P* values. One hundred eight heterogeneity indices were calculated and compared to indices, clearly distinguishing normal controls from PD patients. Two-dimensional scatter plots and histograms, based on heterogeneity indices, were used to classify the 2 groups. Two-dimensional scatter plots consist of a combination of 2 heterogeneity indices; they find the optimal indices for classification. To more clearly classify the 2 groups, we created a quantifier of 3-dimensional phase space.[33] Three-dimensional axes were constructed by finding 3-indices with small *P* values in the statistical comparison. Statistical comparative analysis could classify up to 3 heterogeneity indices. Apart from the 3 heterogeneity indices, with other indices it was difficult to influence the decision for a final diagnosis. ML was introduced to determine all heterogeneity indices. Seventy-one heterogeneity indices were used to distinguish PD patients from normal controls using ML, dimensional reduction and decision trees. After the 71-dimensional reduction by the PCA algorithm, the LR algorithm[34] was used to determine the accuracy. In addition to the LR algorithm, the SVM,[22] RF,[23] and XGBoost algorithm[24,25] were used. XGBoost algorithm calculates feature importance and arranges each according to importance of the heterogeneity indices.

3. Results

In the image of F-18 FP-CIT with PD, a decreased dopamine transporter uptake was observed in the basal ganglia, as shown in
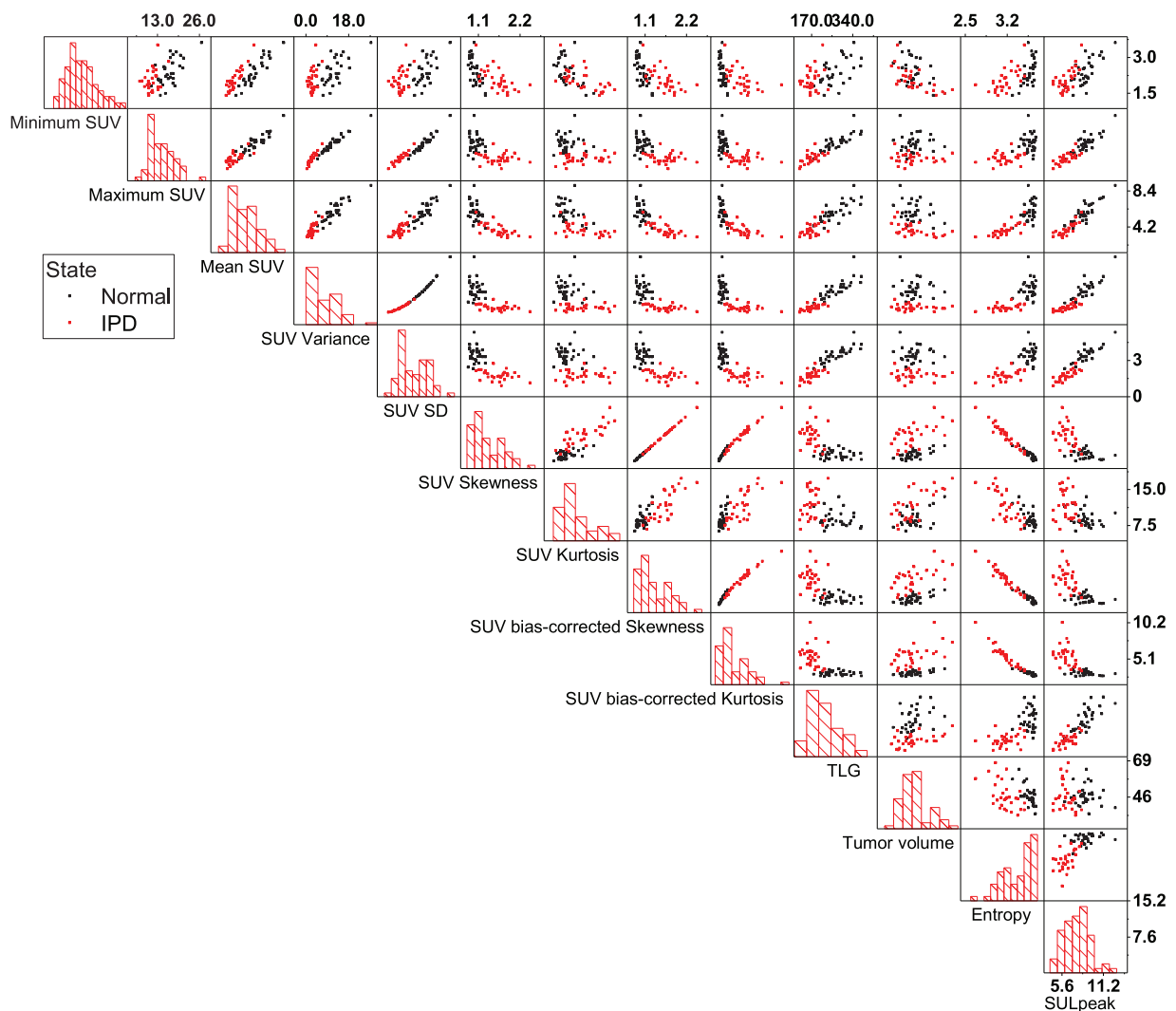
Figure 2. Scatter plots and histograms for normal and PD in heterogeneity indices derived from global SUV statistics in F-18 FP-CIT brain PET images. The values of indices by P values discriminated between patients with PD and non-PD.

Figure 1. The heterogeneous decrease in striatal dopamine transporters, indirectly indicates an excessive uptake of F-18 FP-CIT in the background. Compared to the normal group, the PD patient group generally had reduced F-18 FP-CIT uptake in both *putamina* and *nuclear nuclei*, as shown in Figure 1. The heterogeneity indices of co-occurrence, voxel-alignment, neighborhood intensity-difference, intensity-size-zone, normalized co-occurrence, SUV statistics, texture spectrum, texture feature coding, texture feature coding co-occurrence, and neighboring gray level dependence of feature parent matrix form are summarized in Table 1. Out of 10 feature-parent items, 71 sub-features satisfy $P < .05$ condition through t test, co-occurrence (2/7), voxel-alignment (5/11), neighborhood intensity-difference (4/5), intensity-size-zone (7/11), normalized occurrence (6/6), SUV statistics (24/49), texture spectrum (2/2), texture feature coding (2/4), texture feature coding co-occurrence (8/8), neighboring gray level dependence (3/5). Heterogeneity feature indices with large P values: size-zone variability (intensity-size-zone), contrast (co-occurrence), and inverse difference moment (texture feature coding co-occurrence) are summarized in Table 1.

One hundred eight indices related to heterogeneity were analyzed to distinguish between "with and without" PD, by global and textural feature analysis. The distinction of metastasis in the striatal region of the patients, with and without PD, was possible by using the 24 global indices (SUVmax, SUVmean, SUVSD, SUV variance, SUV kurtosis, SUVpeak, etc) from SUV statistics feature analysis. As shown in Table 1, there were a total of 49 heterogeneity indices for SUV statistics. There were 24 heterogeneity indices with $P < .05$ that distinguished normal from PD. The heterogeneity indices of SUV SD, Maximum SUV, mean SUV, and SUV skewness were good enough to distinguish between the 2 groups (normal and PD), as shown in Figure 2. Two groups were separated in 2-dimensional scatter plot clearly by the values of heterogeneity of SUV SD and maximum SUV or mean SUV. The histograms and scatter plots of the indices of SUV SD and maximum SUV were good shapes for separation. SUV SD and maximum SUV indices of PD patients can be used to distinguish them from the normal group.

The heterogeneity indices of co-occurrence and normalized co-occurrence of parent matrix form are summarized in Table 1. Second angular moment, contrast, entropy, homogeneity,

Figure 3. Scatter plots and histograms for normal and PD in heterogeneity indices derived from 2- and 3-dimensional of (A) normalized co-occurrence and (B) co-occurrence feature parents in F-18 FP-CIT brain PET images. The histogram was displayed the 2-dimensional scatter plot shows the possibility of distinction between normal and PD.

Figure 4. Scatter plots and histograms for normal and PD in heterogeneity indices derived from 2- and 3-dimensional intensity-size-zone parents in F-18 FP-CIT brain PET images. Histograms and 2-dimensional scatter plots show the possibility of distinguishing between normal and PD.
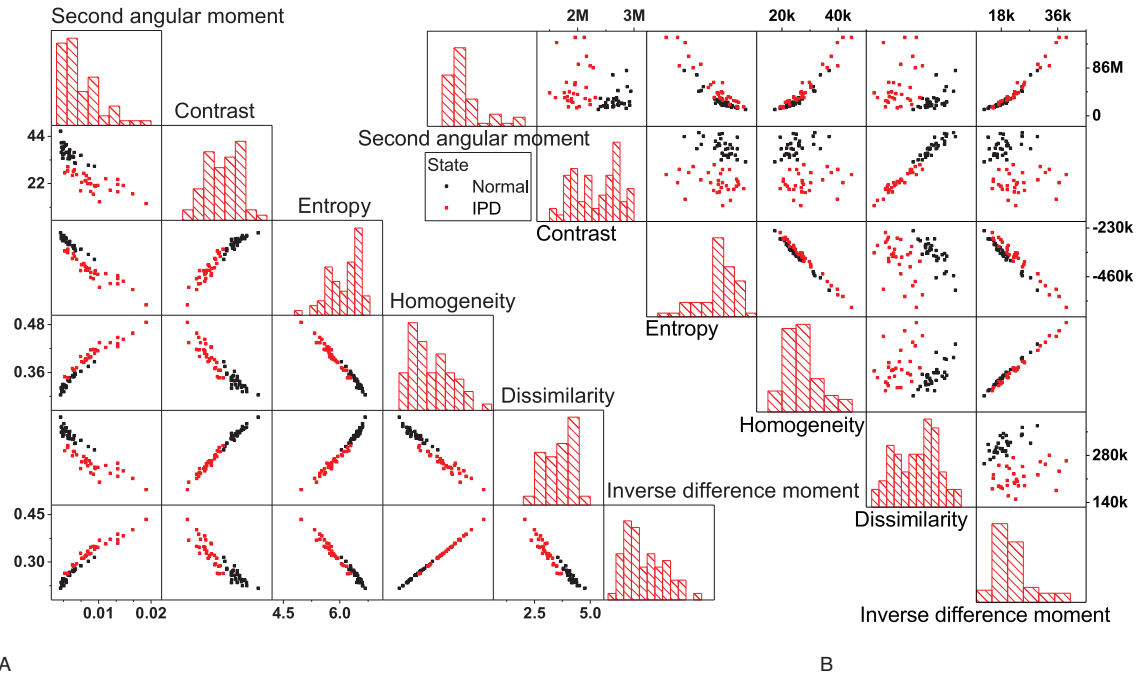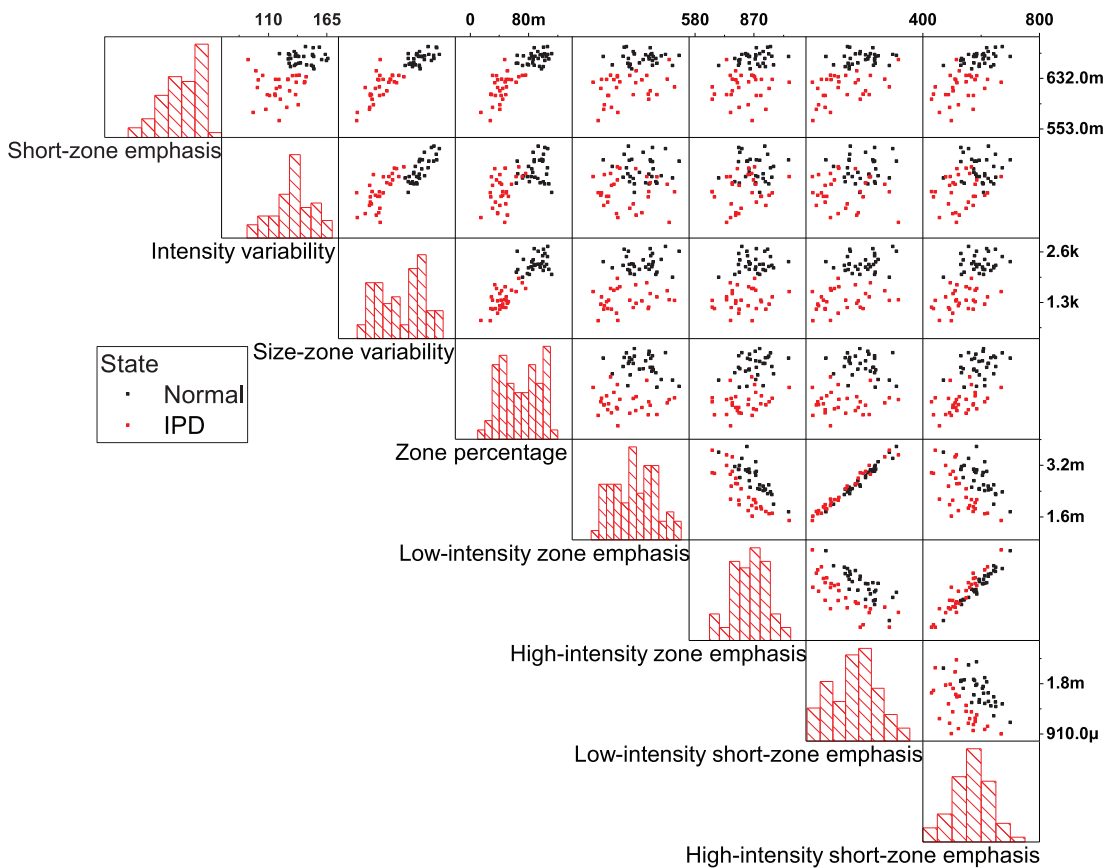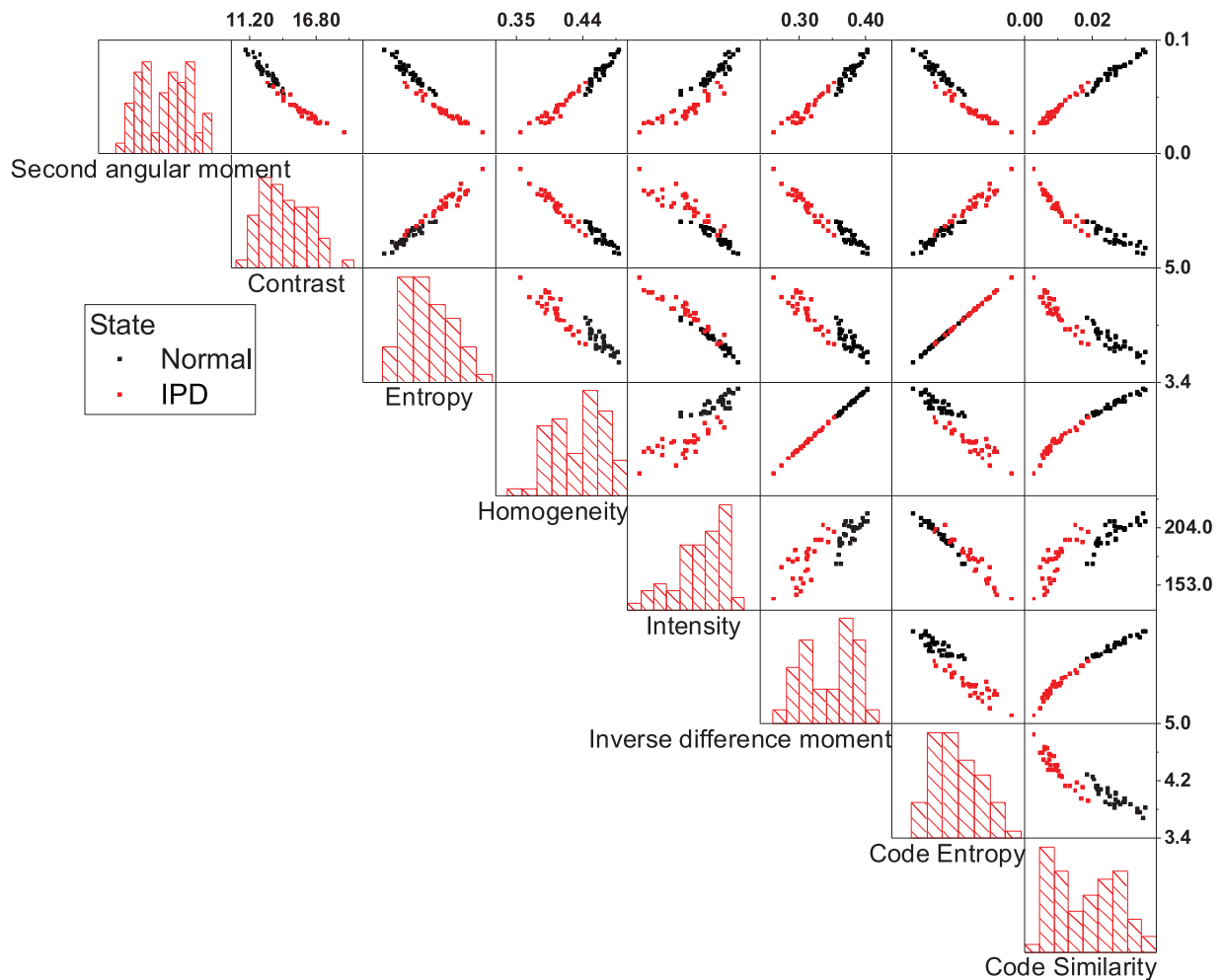
Figure 5. Scatter plots and histograms for normal and PD in heterogeneity indices derived from 2- and 3-dimensional texture feature coding co-occurrence parents in F-18 FP-CIT brain PET images.

dissimilarity, inverse difference moment and correlation were derived by co-occurrence parent matrix. The values of heterogeneity of contrast and dissimilarity found were significantly different from those found by using a 2 sample *t* test; see Table 1. Figure 3 shows a histogram and 2-dimensional scatter distribution of heterogeneity indices. The histogram of contrast and dissimilarity shows the possibility of distinction. The distribution of combination by 2 heterogeneity indices was clearly distinguished between the 2 groups, as shown in Figure 3. As shown in Table 1, the numbers of the heterogeneity indices of co-occurrence and normalized co-occurrence feature parents that can be classified as normal and PD, under $P < .05$, are 2 and 6, respectively. The heterogeneity indices of co-occurrence feature parents, contrast and dissimilarity, had very low P values of 1.00×10^{-12} and 3.75×10^{-10}, and clarified the classification of normal and PD groups in scatter plots. The classification is clearer than for other heterogeneity indices.

Intensity-size-zone feature parents have a total of 11 heterogeneity indices, and 7 indices satisfy $P < .05$ when comparing normal and PD. The separation of the 2 groups can be determined using scatter plots and histograms, as shown in Figures 4 and 5. The indices of intensity short-zone emphasis and size-zone variability was good enough to distinguish between

normal and PD, as shown in Figure 4. The values calculated by intensity variability and low-intensity short-zone emphasis were not clearly distinguished between the 2 groups (Fig. 5). As shown in Table 1, there were 8 heterogeneity indices for the texture feature coding co-occurrence feature parent. PD and normal were classified under conditions of $P < .05$ by statistical comparison using 8 heterogeneity indices. Heterogeneity indices of entropy, homogeneity, intensity, inverse difference moment, code entropy, code similarity were very good at distinguishing 2 groups (normal and PD).

We have found that scatter plots and histograms more clearly separate normal and PD. Scatter plots and histograms were selected in Figure 6 by selecting 12 heterogeneity indices, in consideration of low P value and scatter plots for clearer classification. The heterogeneity indices selected were code similarity (parent matrix: texture feature coding co-occurrence), contrast (co-occurrence), size-zone variability (intensity-size-zone), contrast (normalized co-occurrence), inverse difference moment (texture feature coding co-occurrence), zone-percentage (intensity-size-zone), homogeneity (texture feature coding co-occurrence), dissimilarity (co-occurrence), second angular moment (texture feature coding co-occurrence), SUV SD (SUV statistics), max spectrum (texture spectrum), and dissimilarity
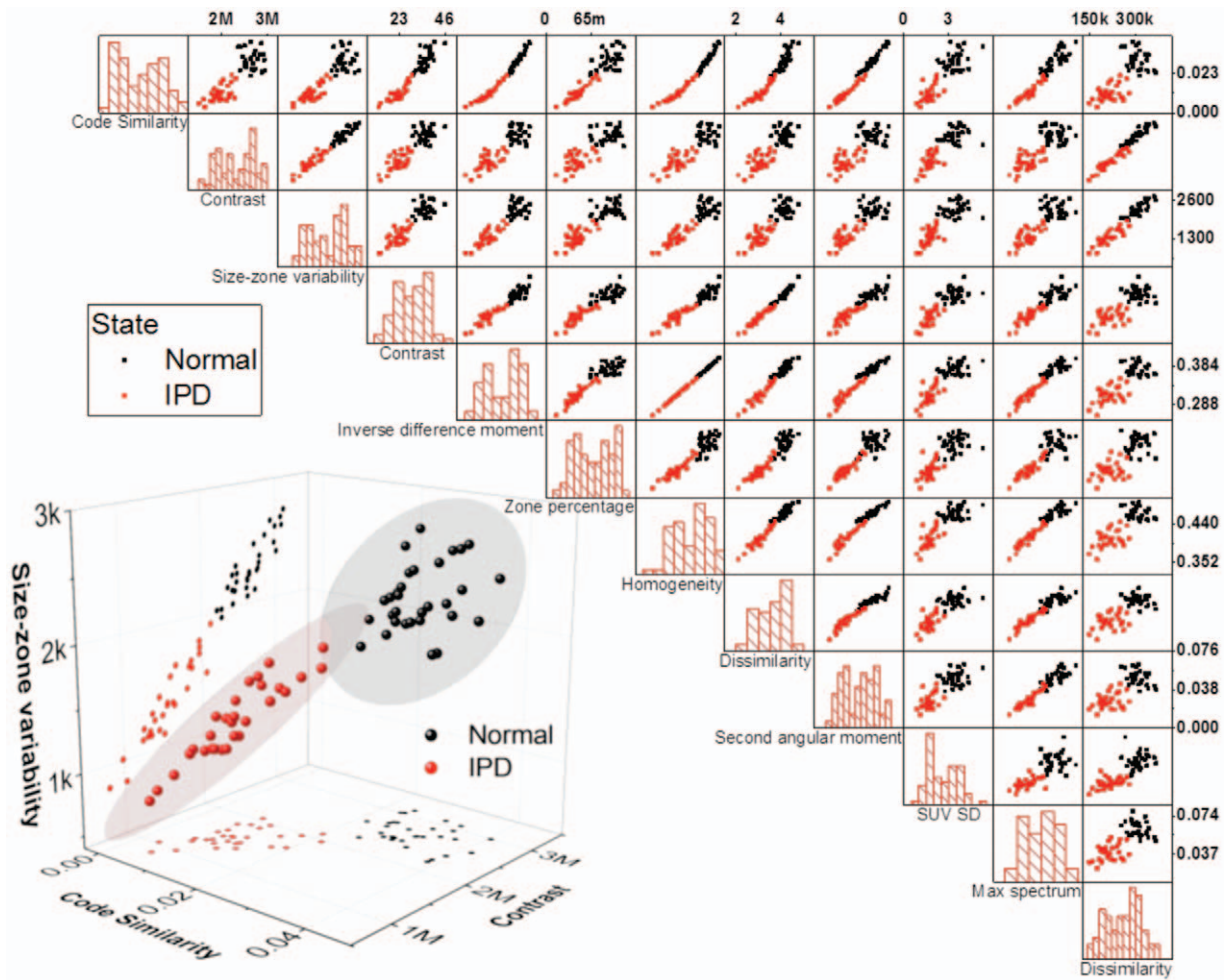
Figure 6. Scatter plots and histograms were selected by selecting 12 heterogeneity indices in order of easy classification of normal and PD. The selection was randomly chosen by visually observing the scatter plot around the low *P* value. Most heterogeneity indices show a good classification between normal and PD (black and red). A non-uniform quantitative phase space was introduced in the separation between the 2 groups. The separation between the 2 groups is clear. Three-dimensional quantifier phase-space of heterogeneity was introduced to further clarify the classification between the 2 groups.

(normalized co-occurrence). The 12 heterogeneity indices selected showed a good separation between normal and PD (black and red). To more clearly classify the 2 groups, a 3-dimensional quantifier phase-space of heterogeneity was introduced. A 3-dimensional quantifier phase-space of heterogeneity using size-zone variability (intensity-size-zone), code similarity (texture feature coding co-occurrence), and contrast (co-occurrence) was introduced to clarify the classification between the 2 groups.

As shown in Figure 7, 3 heterogeneity indices were selected to quantify the heterogeneous phase space. Three axes consist of 3 different heterogeneity indices. Each heterogeneity index was selected from a low *P* value, the shape of the scatter diagram, and the different parent features. 95% confidence ellipsoids for the classification of normal and PD are shown with a 3-dimensional phase space quantifier. For most 3-dimensional quantifiers, the separation of PD and non-PD is clearer, as shown in Figure 7. The 3-dimensional quantifier in Figure 7 is shown by (A) for normalized contrast, code similarity and contrast, (B) for SUV SD, code similarity and contrast, (C) for dissimilarity, code similarity and contrast, and (D) dissimilarity, contrast, and SUV

SD. Figure 7A, B showed the best performance when using the classification 3D quantifier. Three-dimensional heterologous quantifiers were shown to enable the diagnosis of PD patients. Different heterogeneous quantifiers, using various heterogeneity indices, could alter the diagnostic performance, as shown in Figure 7.

By calculating 108 heterogeneity indices, we found, from the statistically-significant indices that clearly distinguished PD patients from normal controls, an order of size-zone variability, co-occurrence contrast and inverse difference moment, as shown in Table 1. When the 2 groups were divided into 2-dimensional and 3-dimensional phase space quantifiers, they could be classified by statistical algorithms. As shown in Table 2, when the 108 dimensions were reduced by using the PCA and classified using the LR algorithm, classification was possible with 91.3% accuracy (Fig. 8). Two groups were classified with an accuracy of 85.5% using the SVM algorithm and 88.4% using the RF algorithm. When the PD patients were classified using the decision tree algorithm, co-occurrence contrast was classified as the top layer. As shown in Table 3, the classification accuracy
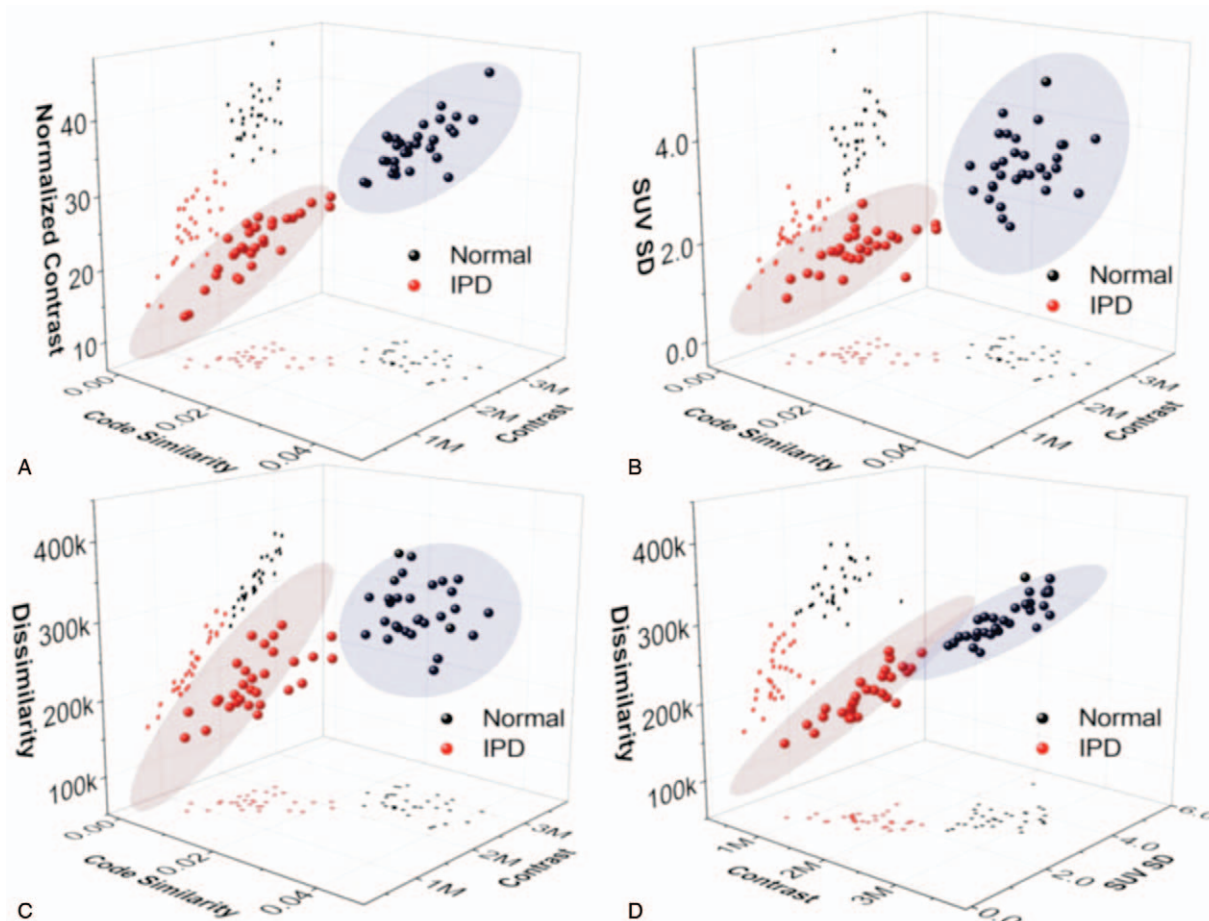
Figure 7. We selected 3 heterogeneity indices for 3-dimensional quantifier phase-space of heterogeneity. 95% confidence ellipsoid for normal and PD was displayed in 3-dimensional phase-space quantifier. Classification was clearer when using 3-dimensional quantifiers. We could help diagnosis for patients with PD applied using this 3-dimensional heterogeneity quantifier.

using the XGBoost algorithm was 95.7%, and the feature importance was SUV bias-corrected kurtosis (22), size-zone variability (20), intensity variability (11), and high intensity zone variability (10).

The classification of HC and IPD using the 4 major parameters related to SUV (SUVmax, SUVmean, SUV variance, SUV SD) shows that they have a high AUC in Figure 9A. The AUC for the 4 heterogeneity parameters, in order of feature importance using the Xgboost method, are shown in Figure 9B.

The area of AUC is higher of 2 heterogeneity parameters than the AUC found in the classification using SUV related parameters. In Table 4, when classifying HC and IPD using the maximum and minimum SUV values, the AUC showed an accuracy of 0.95 or higher. At this time, the cut-off value was 13.42 for the maximum SUV and 4.46 for the minimum SUV. On the other hand, the AUC of the heterogeneity parameter SUV bias corrected for kurtosis or size-zone variability was 0.99 or higher, and almost completely indistinguishable for the 2

Table 2

The precision, recall, F1-score and accuracy of classification by SVM, LR, RF, and XGBoost algorithms.

Method	Classification	Precision	Recall	F1-score	Accuracy
SVM	HC	0.83	0.88	0.86	0.86
	PD	0.88	0.83	0.85	
LR	HC	0.91	0.91	0.91	0.91
	PD	0.91	0.91	0.91	
RF	HC	0.86	0.91	0.89	0.88
	PD	0.91	0.86	0.88	
XGBoost	HC	0.97	0.94	0.96	0.96
	PD	0.94	0.97	0.96	

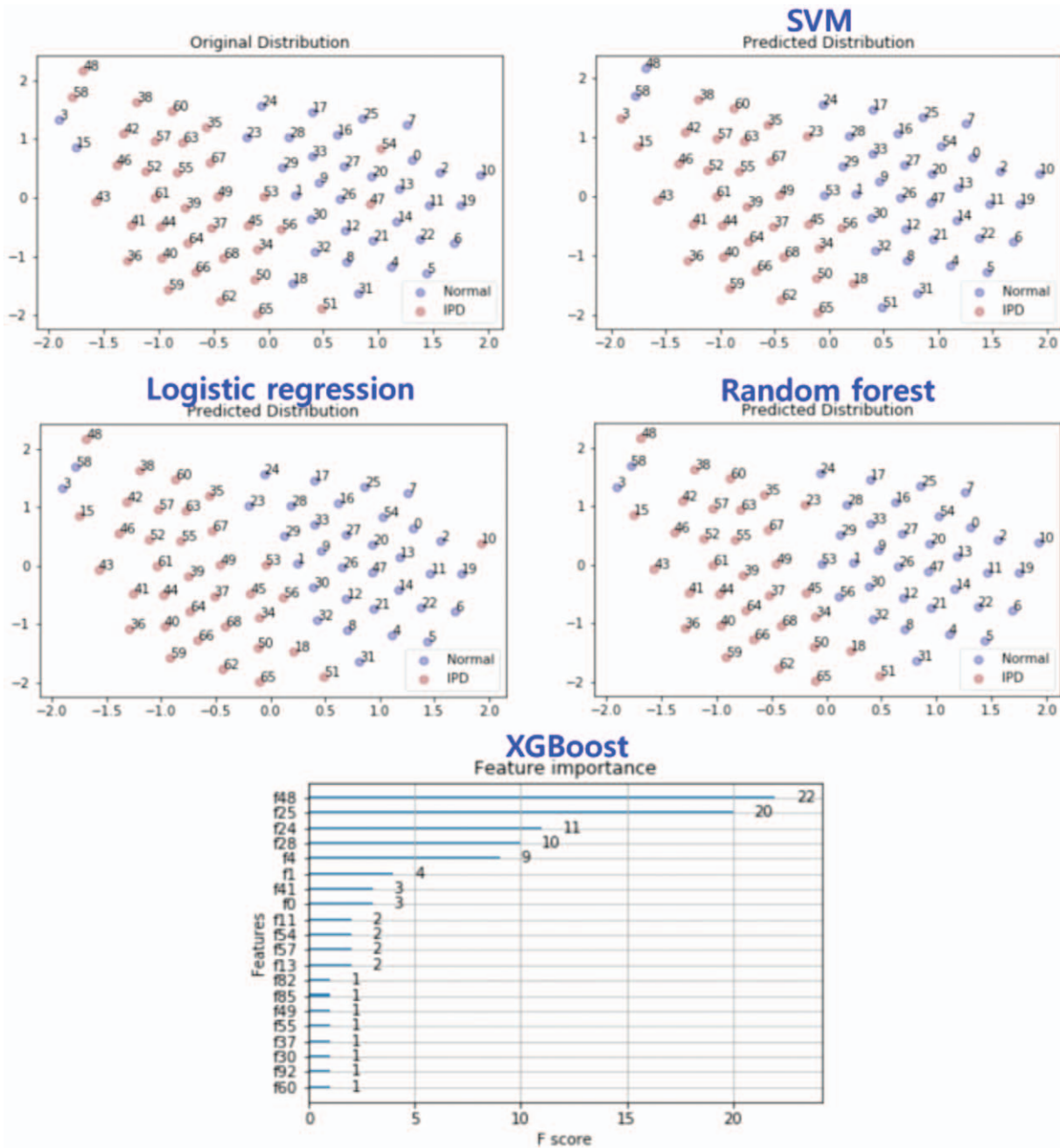HC = health controls, PD = Parkinson's disease.

Figure 8. Original distribution and predicted distribution by SVM, LR, RF, and XGBoost algorithms. The feature importance is displayed by XGBoost.

groups. The cut-off values at this time were 3.47 and 1906.44, respectively.

4. Discussion

Textural feature analysis showed that 71 heterogeneity indices could classify 2 groups, PD and NC, below $P < .05$. Textural feature analysis is quite useful because 71 out of 108 heterogeneity indices can distinguish between normal and PD. The distribution of the radio-pharmaceutical in F-18 FP-CIT brain PET image reflects the heterogeneity of the dopamine transporter. There must be at least 10 pixels to ensure the

accuracy of the calculation of the heterogeneity indices. The calculated values of the heterogeneity indices may not fully reflect their characteristics, because the distribution of the striatum region is not wide enough.

Of the total of 49 heterogeneity indices for SUV statistics, 24 indices were available for classification. SUV statistics could be classified with relatively high classification performance, even in the case of the SUV statistics global indices. This is an evaluation of the SUV characteristics of F-18 FP-CIT PET images, which means that they can be classified with relatively high accuracy. It has been shown that classification using heterogeneity indices reflecting texture features can be separated using more indices

Table 3

The feature importance are listed by using an XGBoost calculation.

No.	Feature name	Importance
1	SUV statistics, SUV bias-corrected kurtosis	22
2	Intensity-size-zone, size-zone variability	20
3	Intensity-size-zone, intensity variability	11
4	Intensity-size-zone, high-intensity zone emphasis	10
5	Co-occurrence-dissimilarity	9
6	Co-occurrence-contrast	4
7	Co-occurrence-second angular moment	3
8	SUV statistics, maximum SUV	3
9	Voxel-alignment, low-intensity short-run emphasis	2
10	SUV statistics, surface mean SUV 1	2
11	SUV statistics, asphericity	2
12	Voxel-alignment, low-intensity run emphasis	2
13	SUV statistics, surface SUV variance 2	1
14	Co-occurrence-correlation	1
15	SUV statistics, surface SUV NSR 4	1
16	SUV statistics, surface total SUV 3	1
17	Intensity-size-zone, zone percentage	1
18	SUV statistics, surface SUV NSR 2	1
19	Texture spectrum, max spectrum	1
20	SUV statistics, surface SUV NSR 3	1

NSR = radial noise-to-signal, SUV = standardized uptake value.

rather than just SUV indices. Therefore, it is suggested that diagnosis considering heterogeneity indices can improve the accuracy of classification between the normal group and the PD group.

Heterogeneity indices of co-occurrence and normalized co-occurrence feature parents could be classified with better performance than other indices. All 6 heterogeneity indices of normalized co-occurrence feature parents were statistically able to classify the 2 groups. Contrast and dissimilarity, derived from

co-occurrence feature parents, were particularly well classified. Using these heterogeneity indices, constructing a 2- or 3-dimensional phase space quantifier will enable efficient and accurate diagnosis. Figure 3 shows a histogram and a scatter plot of the heterogeneity indices, showing high accuracy diagnostic possibilities. The pattern of the figure and the P value allow one to choose the heterogeneity indices to use in 2 or 3 dimensions. The accuracy of the classification can vary, depending on the heterogeneity indices chosen and the shape of the classification. The limitation of this method is that one needs to configure the quantifier in every case and check the results, so as to determine which one performs the best. Of the 108 heterogeneity indices, 73 were found to be effective for classification. The disadvantage of classifying using a 3-dimensional phase-space quantifier is that only 3 indices should be used among the 73 valid indices. Despite some limitations, 3-dimensional phase space quantifiers have been observed to exhibit fairly good classification performance.

Figures 4 and 5 show that heterogeneity indices of intensity-size-zone and texture feature coding co-occurrence feature parents can be classified with good performance. As shown in Figure 5 and Table 1, all 8 of the heterogeneity indices of texture feature coding co-occurrence feature parents were statistically classified into 2 groups. Inverse different moment and homogeneity of texture feature coding co-occurrence feature parents showed excellent classification performance, even with a single parameter. These 2 indices show good performance in classification, even when the 2 are combined to form a 2-dimensional quantifier. The intensity variability and zone percentage variability derived from the intensity-size-zone feature parent were not very good when classified as a single parameter. Intensity variability and zone percentage variability of intensity-size-zone feature parents show better classification performance when using 2-dimensional quantifiers. This means that a more accurate classification can be achieved by combining the various indices appropriately. The visually distinguishable maximum is a
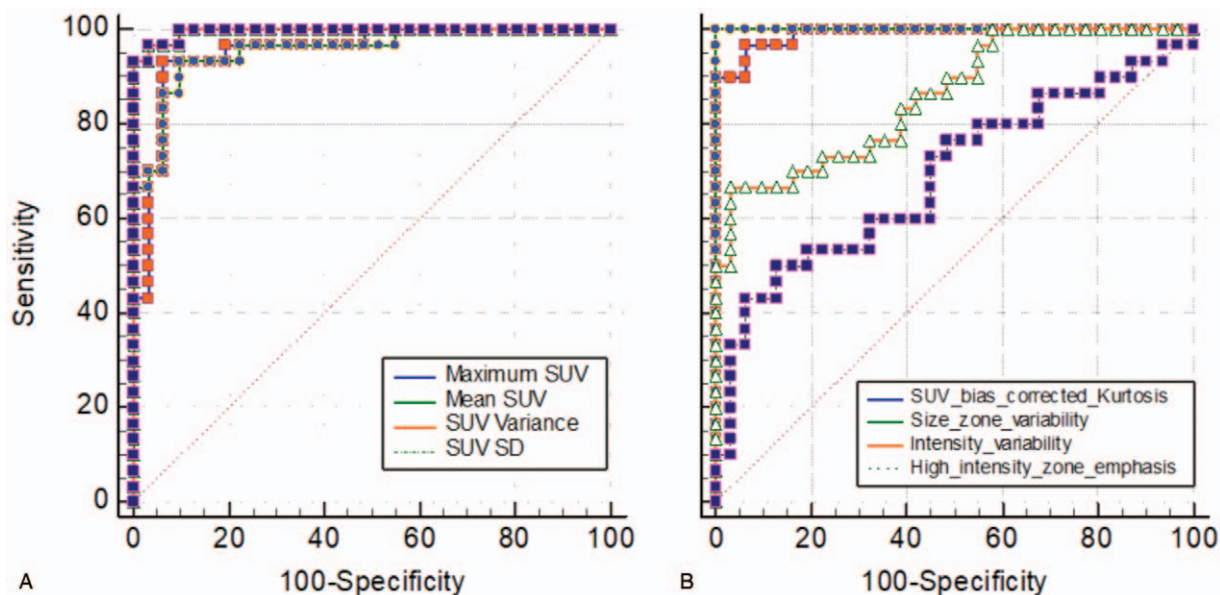

Figure 9. ROC curve (A) of SUV-related parameters (SUVmax, SUVmean, SUV variance, SUV SD) for classification of HC and IPD and ROC curve (B) of 4 major heterogeneity parameters (SUV bias-corrected kurtosis, size-zone variability, intensity variability, high intensity zone emphasis) according to feature importance in Table 3.

Table 4

AUC, sensitivity, specificity, cut-off value, and *P* value based on SUV-related items and 4 major heterogeneity parameters in Table 3 for classification of HC and IPD.

Region	AUC	Sensitivity	Specificity	Cut-off value	*P*
Maximum SUV	0.954	93.33	93.55	13.42	<.0001
Mean SUV	0.956	93.33	90.32	4.46	<.0001
SUV variance	0.996	96.67	96.77	5.31	<.0001
SUV SD	0.996	96.67	96.77	2.3	<.0001
SUV bias corrected kurtosis	0.99	96.67	93.55	3.47	<.0001
Size-zone variability	1	100	100	1906.44	<.0001
Intensity variability	0.859	66.67	96.77	129.21	<.0001
High intensity zone emphasis	0.689	50	87.1	800.29	.0066

SD = standard deviation, SUV = standardized uptake value.

3-dimensional phase space quantifier created by combining 3 indices. Three-dimensional quantifiers can be used to provide the best performance for classification, but the optimum-selection of 3 heterogeneity indices is still a problem to be solved.

The heterogeneity indices with good classification were selected and are shown in Figure 6. The selection condition was to select the low *P* value first, and then randomly select those that stably separate the normal and PD groups in the scatter plot. It is well-separated in most quantifiers, when it is classified using a 2-dimensional scatter plot. Among the heterogeneity parameter pairs that make up the 2-dimensional quantifier, classification was clearly possible with: (code similarity, SUV SD), (inverse different moment, homogeneity), and (code similarity, contrast). The construction of 3-dimensional quantifiers allowed a clearer classification. Figure 6 shows that the 3-dimensional quantifier, consisting of heterogeneity indices of size-zone variability, code similarity, and contrast, classified the normal and PD groups most clearly.

In Figure 7, a combination of heterogeneity indices was constructed to observe the effect on classification. The combinations of 3-dimensional heterogeneity indices were: (normalized contrast, code similarity, contrast), (SUV SD, code similarity, contrast), (dissimilarity, code similarity, contrast), and (dissimilarity, contrast, SUV SD). The classification of the 3-dimensional quantifier in Figure 7 was clear, although there were some differences in the distributions of the data. To evaluate new data, diagnosis can be made by verifying the location on a 3-dimensional quantifier. However, as the number of data is not yet sufficient, there is a possibility that the boundary can be changed when new data are introduced. If more data are used, it is expected that the boundary will be fixed stably and the classification will be clear.

Using a recent ML technique, the PD family was classified in a way that all 108 indices were considered. This approach is intended to maximize the advantage of considering 108 heterogeneity indices. Using dimension reduction eliminates unnecessary indices automatically, so that the most important indices have a great influence on the classification. As shown in Figure 7, classification was performed using SVM, LR, RF and XGBoost algorithms. The most accurate algorithm is 95.7% accurate, using the XGBoost algorithm, as shown in Table 2. The XGBoost algorithm shows which heterogeneity indices play a key role in classification by displaying feature importance, as shown in Table 3. It is very useful to classify a small number of data because the ML algorithm is efficient and the reason for classification can be explained.

As a result of performing the *t* test, an index indicating the difference between the 2 groups of HC and IPD, it was shown that, when using the heterogeneity parameters with high *P* value, classification of HC and IPD was possible as with SUV-related parameters. The PET images of the brains of patients who reached a stable state after administration of F-18 FP-CIT radiopharmaceuticals were relatively clearly distinguished by comparison of images by the intensity of SUV. When employing classification using heterogeneity parameters with a high *P* value through *t* test, the AUC in ROC could be increased. The limitation is that the number of data is 61, so there is a possibility of over-fitting because the relatively small number being used. When acquiring more data, errors due to over fitting can be reduced. Quantitative analysis is possible by using the cut-off presented in Table 4. If the classification result changes when using cut-off in quantitative diagnosis, the AUC can be analyzed by placing weights in the sequence of higher order.

5. Conclusion

Heterogeneity indices of the striatum region of F-18 FP-CIT brain PET images were calculated, and normal and PD groups were successfully separated by statistical comparison and scatter plot comparison. One hundred eight heterogeneity indices were calculated, of which 71 were found to be very useful for classification. The classification was attempted by constructing a scatter plot in 2-dimensional and 3-dimensional phase spaces, and the 2 groups were separated very successfully. When selecting the best indices for classification, based on low *P* values, and generating 3-dimensional quantifiers, PD patients were successfully separated from normal groups. ML was introduced to classify all 108 heterogeneity indices, and the separation was very successful with the XGBoost algorithm. The method of extracting heterogeneity indices and using ML is very useful because it can provide a reason for classification through feature importance. The classification by artificial intelligence analysis has the characteristic that the accuracy of learning increases in proportion to the number of data, while the possibility of overfitting exists due to the small number of data used in this study. Quantitative classification of HC and IPD is possible by using the heterogeneity parameter, and it can be used for quantitative analysis of new images by presenting the cut-off value.

Author contributions

Conceptualization: Do-Young Kang, Young-Jin Jeong, Ji-Eun Jeong.

Data curation: Kook Cho, Hyun Jin Yoon.

Formal analysis: Woong Gon Kim.

Funding acquisition: Do-Young Kang.

Investigation: Kook Cho.

Methodology: Kook Cho, Hyun Jin Yoon, Woong Gon Kim.

Project administration: Do-Young Kang.

Resources: Do-Young Kang.

Software: Kook Cho, Woong Gon Kim.

Statistical processing: Woong Gon Kim, Kook Cho

Supervision: Young-Jin Jeong, Ji-Eun Jeong.

Visualization: Woong Gon Kim.

Writing – original draft: Hyun Jin Yoon, Do-Young Kang.

Writing – review & editing: Young-Jin Jeong, Ji-Eun Jeong, Do-Young Kang.

References

[1] Oh M, Kim JS, Kim JY, et al. Subregional patterns of preferential striatal dopamine transporter loss differ in Parkinson disease, progressive supranuclear palsy, and multiple-system atrophy. J Nucl Med 2012;53:399–406.

[2] Brooks DJ. Functional imaging in relation to parkinsonian syndromes. J Neurol Sci 1993;115:1–17.

[3] Huang W, Lee M, Lin J, et al. Usefulness of brain 99m Tc-TRODAT-1 SPET for the evaluation of Parkinson's disease. Eur J Nucl Med Mol Imaging 2004;31:155–61.

[4] Ma Y, Dhawan V, Mentis M, Chaly T, Spetsieris P, Eidelberg D. Parametric mapping of [18F] FPCIT binding in early stage Parkinson's disease: a PET study. Synapse 2002;45:125–33.

[5] Mozley PD, Acton PD, Barraclough ED, et al. Effects of age on dopamine transporters in healthy humans. J Nucl Med 1999;40:1812–7.

[6] Lavalaye J, Booij J, Reneman L, Habraken JB, van Royen EA. Effect of age and gender on dopamine transporter imaging with [123 I] FP-CIT SPET in healthy volunteers. Eur J Nucl Med 2000;27:867–9.

[7] Whone AL, Watts RL, Stoessl AJ, et al. Slower progression of Parkinson's disease with ropinirole versus levodopa: the REAL-PET study. Ann Neurol 2003;54:93–101.

[8] Parkinson Study GroupLevodopa and the progression of Parkinson's disease. N Engl J Med 2004;351:2498–508.

[9] Meara J, Bhowmick BK, Hobson P. Accuracy of diagnosis in patients with presumed Parkinson's disease. Age Ageing 1999;28:99–102.

[10] Antonini A, Benti R, De Notaris R, et al. 123 I-Ioflupane/SPECT binding to striatal dopamine transporter (DAT) uptake in patients with Parkinson's disease, multiple system atrophy, and progressive supranuclear palsy. Neurol Sci 2003;24:149–50.

[11] Cheng N, Fang YD, Tsan D, Hsu C, Yen T. Respiration-averaged CT for attenuation correction of PET images – impact on PET texture features in non-small cell lung cancer patients. PLoS One 2016;11:e0150509.

[12] Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18 F-FDG PET/CT imaging by texture analysis. Eur J Nucl Med Mol Imaging 2013;40:133–40.

[13] Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images

[14] predicts response to concomitant radiochemotherapy in esophageal cancer. J Nucl Med 2011;52:369–78.

[14] Saini H, Sahni V. Region growing segmentation using de-noising algorithm for medical ultrasound images. 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT); 2017:1–5.

[15] Amadasun M, King R. Textural features corresponding to textural properties. IEEE Trans Syst Man Cybern 1989;19:1264–74.

[16] Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. IEEE Trans Biomed Eng 2013;61:630–7.

[17] He D, Wang L. Texture features based on texture spectrum. Pattern Recognit 1991;24:391–9.

[18] Horng M, Sun Y, Lin X. Texture feature coding method for classification of liver sonography. Comput Med Imaging Graphics 2002;26:33–42.

[19] Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. Comput Vis Graph Image Process 1983;23:341–52.

[20] Tixier F, Hatt M, Le Rest CC, Le Pogam A, Corcos L, Visvikis D. Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET. J Nucl Med 2012;53:693–700.

[21] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;610–21.

[22] Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48:441–6.

[23] Wang H, Zhou Z, Li Y, et al. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images. EJNMMI Res 2017;7:11.

[24] Zhang R, Chen Z, Xu L, Ou C. Meteorological drought forecasting based on a statistical model with machine learning techniques in Shaanxi province, China. Sci Total Environ 2019;665:338–46.

[25] Gao C, Sun H, Wang T, et al. Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. Sci Rep 2018;8:1–21.

[26] Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 1992;55:181–4.

[27] Kang Y, Na D, Hahn S. Seoul Neuropsychological Screening Battery. Incheon: Human Brain Research & Consulting Co; 2003.

[28] Jeong YJ, Jeong J, Cheon S, Yoon B, Kim JW, Kang D. Relationship between the washout rate of I-123 MIBG scans and autonomic function in Parkinson's disease. PLoS One 2020;15:e0229860.

[29] Yoon HJ, Jeong YJ, Son HJ, Kang D, Hyun K, Lee M. Optimization of the spatial resolution for the GE discovery PET/CT 710 by using NEMA NU 2-2007 standards. J Korean Phys Soc 2015;66:287–94.

[30] Son HJ, Jeong YJ, Yoon HJ, et al. Parkinson disease-related cortical and striatal cognitive patterns in dual time F-18 FP CIT: evidence for neural correlates between the caudate and the frontal lobe. Q J Nucl Med Mol Imaging 2019;63:379–86.

[31] PMOD. Available at: https://www.pmod.com, Accessed December 6, 2020.

[32] Fang YD, Lin C, Shih M, et al. Development and evaluation of an open-source software package "CGITA" for quantifying tumor heterogeneity with molecular images. BioMed Res Int 2014;2014:248505.

[33] Brooks FJ, Grigsby PW. Current measures of metabolic heterogeneity within cervical cancer do not predict disease outcome. Radiat Oncol 2011;6:69.

[34] Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. J Clin Epidemiol 2010;63:826–33.