Comparison of decision tree and stepwise regression methods in classification of FDG-PET brain data using SSM/PCA features

Mudali, Deborah; Roerdink, Jos B.T.M.; Teune, Laura K.; Leenders, Klaus L.; Renken, Remco J.

Published in:
Proceedings of the 8th International Conference on Advanced Computational Intelligence, ICACI 2016

DOI:
10.1109/ICACI.2016.7449841

8th International Conference on Advanced Computational Intelligence
Chiang Mai, Thailand; February 14-16, 2016

Comparison of Decision Tree and Stepwise Regression Methods in Classification of FDG-PET Brain Data using SSM/PCA Features

Deborah Mudali
and Jos B.T.M. Roerdink
Johann Bernoulli Institute for
Mathematics and Computer Science
University of Groningen
The Netherlands
Email: d.mudali@rug.nl

Laura K. Teune
and Klaus L. Leenders
Department of Neurology
University Medical Center Groningen
University of Groningen
The Netherlands
Email:l.k.teune@umcg.nl,
k.l.leenders@umcg.nl

Remco J. Renken
and Jos B.T.M. Roerdink
University of Groningen
University Medical Center Groningen
Neuroimaging Center, Groningen
The Netherlands
Email: r.j.renken@rug.nl,
j.b.t.m.roerdink@rug.nl

***Abstract*—Objective: To compare the stepwise regression (SR) method and the decision tree (DT) method for classification of parkinsonian syndromes.**

Method: We applied the scaled subprofile model/principal component analysis (SSM/PCA) method to FDG-PET brain image data to obtain covariance patterns and the corresponding subject scores. The subject scores formed the input to the C4.5 decision tree algorithm to classify the subject brain images. For the SR method, scatter plots and receiver operating characteristic (ROC) curves indicate the subject classifications. We then compare the decision tree classifier results with those of the SR method.

Results: We found out that the SR method performs slightly better than the DT method. We attribute this to the fact that the SR method uses a linear combination of the best features to form one robust feature, unlike the DT method. However, when the same robust feature is used as the input for the DT classifier, the performance is as high as that of the SR method.

Conclusion: Even though the SR method performs better than the DT method, including the SR procedure in the DT classification yields a better performance. Additionally, the decision tree approach is more suitable for human interpretation and exploration than the SR method.

***Keywords*—Parkinsonian syndromes; FDG-PET; scaled subprofile model; principal component analysis; decision tree classification; stepwise regression**

I. Introduction

Parkinsonian syndromes, like other neurodegenerative diseases, are not easy to diagnose and distinguish at an early stage [1], [2]. With the purpose of classifying these syndromes, the scaled subprofile model/principal component analysis (SSM/PCA) method as explained by Moeller phet al. [3] is used to extract disease-related metabolic brain patterns in the form of principal component images from subject brain images. Then these individual subject images are projected onto the patterns to obtain their corresponding scores. These scores depict the network expression of individual subjects on the pattern [4].

The SSM/PCA method has been used in several studies to extract disease-related patterns from imaging data. In Moeller phet al. [5] the SSM method is applied to regional metabolic rates for glucose data to identify specific age-related disease profiles. Similarly, in Spetsieris phet al. [1] the SSM/PCA method is used to derive disease-related spatial covariance patterns which are represented as spatial weighted images. In the study by Spetsieris and Eidelberg [6] the methodological questions that arise regarding the use of the SSM method are addressed. In addition, the SSM/PCA method together with several versions of the Statistical Parametric Mapping (SPM) software were applied by Peng phet al. [7] to obtain disease-specific patterns. Therefore, from the aforementioned studies we can say that the SSM/PCA method application is quite broad and effective at identifying brain patterns. These patterns can act as biomarkers for predicting parkinsonian disorders and neurodegenerative diseases in general.

This paper, which is an extended version of [10], presents a comparison between the stepwise regression (SR) method [8] and the decision tree (DT) method in the classification of parkinsonian syndromes [9]. In both methods we apply the SSM/PCA method to the brain data to obtain subject scores, which are used as features in the subsequent classification process. Specifically, we use the C4.5 machine learning algorithm in this study to build the DT classifiers [11], [12], [13]. The SR method uses a mechanism of choosing one model or a few models (here known as components) from a larger set of models [14], [15]. Further, the components are chosen based on how well they separate subject image groups using the Akaike Information Criterion (AIC) [16].

There are three approaches we use in this study:

1) the stepwise regression method;
2) decision tree classification with all features, and a reduced set of features, respectively;
3) decision tree classification using the set of features obtained from the stepwise regression procedure.

978-1-4673-7782-9/16/$31.00 ©2016 IEEE

With the SR method, one feature (subject z-score) is determined from a combination of components, while in the DT method several features (subject scores) are determined from individual components. In approach 3 we combine the SR procedure and DT method in two different ways. In the first approach, the best features obtained by the SR procedure are used as features for decision tree classification, that is, without linearly combining them. In the second approach, we use the exact same subject z-score (that is, a linear combination of best features) as obtained by the SR method (stepwise plus logistic regression procedure) and use it as a single feature for decision tree classification.

II. Method

A. Data acquisition and feature extraction

We used fluorodeoxyglucose positron emission tomography (FDG-PET) brain scans as described in the previous studies by Teune phet al. [17], [8]. The data set includes a total of 76 subject brain images, namely: 18 healthy controls (HC), 20 Parkinson's disease (PD), 21 multi-system atrophy (MSA), and 17 progressive supra-nuclear palsy (PSP). An implementation of the SSM/PCA method developed in Matlab was used, following the procedure as described by Eidelberg [18], Spetsieris phet al. [1], [19], and Spetsieris & Eidelberg [6].

The SSM/PCA method was applied to the FDG-PET data to obtain principal components (PCs) onto which original images were projected to obtain their weights on the PCs, known as subject scores. Thereafter, we used the subject scores as features for the decision tree method and the stepwise regression procedure to differentiate among the parkinsonian syndromes.

B. Classification

1) Stepwise regression method: Following Teune phet al. [8], the SR procedure is used to obtain a linear combination of PCs (combined pattern) that best discriminates groups. The SR method consists of the following steps:

- The principal components that make up 50% of the variance are considered in the stepwise regression procedure. This procedure retains only those components which best separate groups according to Akaike's information criterion (AIC) [16].
- By fitting the subject scores corresponding to the retained PCs to a logistic regression model, scaling factors (or weights) for all PCs are obtained. The combined pattern is a sum of PCs weighted by the scaling factors. Then the subject score on the combined pattern is determined by adding the retained subject scores multiplied by their corresponding scaling factors.
- Z-scores are calculated and displayed on scatter plots, and receiver operating characteristic (ROC) curves are determined. Then a subject is classified according to the z-score cut-off value, which corresponds to the z-score where the sum of sensitivity and specificity is maximised. A subject is diagnosed as belonging to the class of patients if the z-score value is higher than the cut-off value, and as a healthy control if it is lower than the cut-off value.

a) Leave one out cross validation (LOOCV): In the SSM/PCA-SR method, one subject (for testing) is removed from the training set at a time and the SSM/PCA method is applied to the remainder of the subjects. The stepwise regression procedure is followed to create a combined pattern. The left-out subject scores on the PCs that form the combined pattern are multiplied by the scaling factors to obtain a single subject score on the combined pattern. Each subject score is transformed into a z-score which then becomes the feature used to separate groups.

2) Decision tree method: This method builds a classifier from a set of training samples with a list of features and class labels. We used the C4.5 machine learning algorithm by Quinlan [12] to train classifiers based on the subject scores as features. As a result, a pruned decision tree showing classified subject images is generated. Pruning helps to obtain a tree which does not overfit. Note that with the decision tree method the principal components are not combined but instead used individually. Therefore, the DT method uses several features (subject scores on several PCs), unlike the SR method which uses only one feature (z-score).

a) Leave one out cross validation: We placed one subject into a test set and the rest of the subjects into a training set. Then the SSM/PCA method was applied to the training set to obtain subject scores. These subject scores were used to train the classifier, and subsequently the test subject was used to test the DT classifier performance. The procedure was repeated for each subject in the dataset. We used the AIC criterion in conjunction with the SR procedure to pre-select features for the DT method in order to improve the classifier performance. Furthermore, we provided the single combined feature from the SR method as input to the DT method.

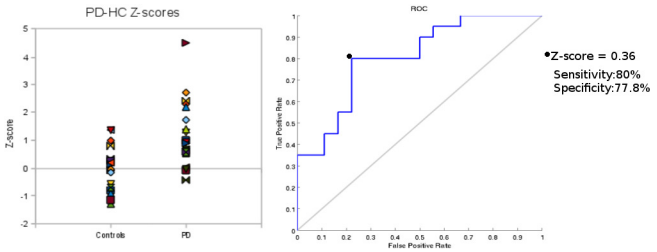
III. Results

A. Stepwise Regression Procedure

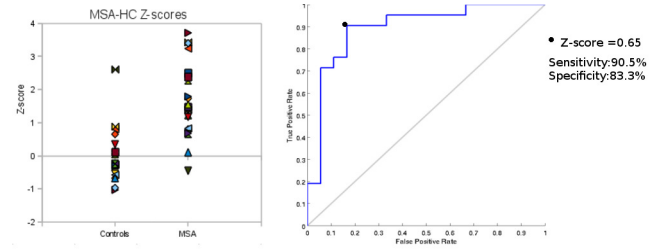
The z-score scatter plots of the combined pattern and the ROC curves are illustrated in Fig. 1. For the scatter plots, the groups are displayed on the X-axis and the z-scores on the Y-axis. On the ROC curves the bullet (•) represents the cut-point where the difference between true positive rate and false positive rate, also called the Youden index [20], is maximised. These results are similar to those in Teune phet al. [8]. The only difference is seen in Fig. 1(a), where the cut-off is 0.36 instead of 0.45. This can be explained by the fact that at both cut-off points the sensitivity and specificity are the same; in this case the value 0.36 is chosen, being the first z-score value in ascending order.

B. Decision tree classifiers for disease groups versus the healthy group

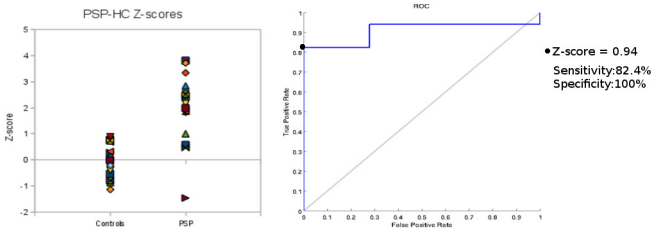
The decision tree classifiers are built from the disease datasets (PD, PSP, MSA), all compared to the healthy control (HC) group of 18 subjects. Fig. 2 and 3 show the decision tree diagrams and corresponding scatter plots. The internal



(a)



(b)



(c)

Fig. 1: Scatter plots and ROC curves for subject z-scores. (a): PD vs HC; (b): MSA vs HC; (c): PSP vs HC.

tree nodes are drawn as oval shapes, corresponding to the attributes (subject scores) on which decisions are made. The threshold values for splitting the dataset are indicated next to the lines connecting two internal nodes. The actual class labels are represented by the rectangles (leaves), where "1" is the label for the disease group (PD, PSP, or MSA) and "0" the label for the healthy group (HC). In addition, the number in the brackets within a rectangle shows the total number of subjects that are classified at that leaf; the number after the slash (if present) represents the number of misclassified cases at that leaf.

1) PD Group: The output of the decision tree method applied to the PD-HC dataset (18 healthy and 20 PD) is illustrated in Fig. 2. The attributes are subject scores derived from 38 principal components.

As can be seen in Fig. 2, the classifier chooses the subject score based on component number 5 (SSPC5) to make the first split of the dataset. As a result, nine PD subjects (feature value $>$ 254.14) are identified. Next, the classifier uses component number 26 to separate the rest of the subjects, where nine subjects (feature value $<=$ -32.241) are identified as HC; etc. Only one PD subject is misclassified as HC. Looking at the

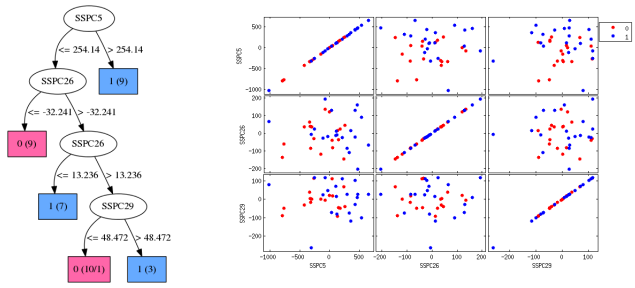


Fig. 2: The decision tree diagram and the scatter plot showing the distribution of the subject scores of the chosen PCs by the decision tree classifier, without feature pre-selection.

scatter plots on the right of Fig. 2, we can clearly see that for the chosen PCs there is no clear separation between PD and healthy controls.

2) MSA and PSP Groups: Fig. 3 shows the decision trees and the distribution of subject scores displayed on scatter plots for the MSA-HC (18 HC and 21 MSA) and PSP-HC (18 HC and 17 PSP) datasets. The attributes are subject scores derived from 39 and 35 principal components for MSA and PSP, respectively. For the MSA group, one HC subject is misclassified, whereas no subject is misclassified for the PSP group. Also, note that for the PSP group the classifier chooses only 2 out of the available 35 PCs, i.e., SSPC1 and SSPC12, as illustrated in the scatter plot of Figure 3(b). Moreover, it uses SSPC1 repeatedly to classify the subjects. Indeed, the C4.5 decision tree inducer can use a feature more than once for making splits, as long as it maximizes the information gain.

C. Decision trees with reduced number of features

In Section III-B we noticed an overlapping distribution of subject scores of the chosen PCs by the classifier, with no clear cut between the PD and HC. To improve robustness, we considered using only the first two components obtained from the PCA process since they depict the highest variance. Fig. 4(a) is an example of one of the 38 classifiers for the PD vs HC group generated during the LOOCV process, that is the classifier constructed after removing one subject from the training set, which is thereafter used for testing the left-out subject. For the purpose of comparing with the SR method, we reproduce some of the LOOCV results from the previous study by Mudali phet al. [9], as shown in Table I.

The scatter plot in Fig. 4(a) shows that there is no clear cut for the classifier to separate the PD and HC groups. This is because the subject scores for both PD and HC are overlapping. As seen from the tree diagram, the classifier chooses one threshold for each of the two given PCs to correctly classify all PD subjects (100% sensitivity), but misclassifies 7 out of 18 HC subjects as well as the test subject (22.2% specificity). That is to say, the decision boundaries found by the classifier were not successful at separating the two groups. In this case, even classifiers which use non-axis aligned decision boundaries may not perform well. Accordingly, there is a need to rescale or

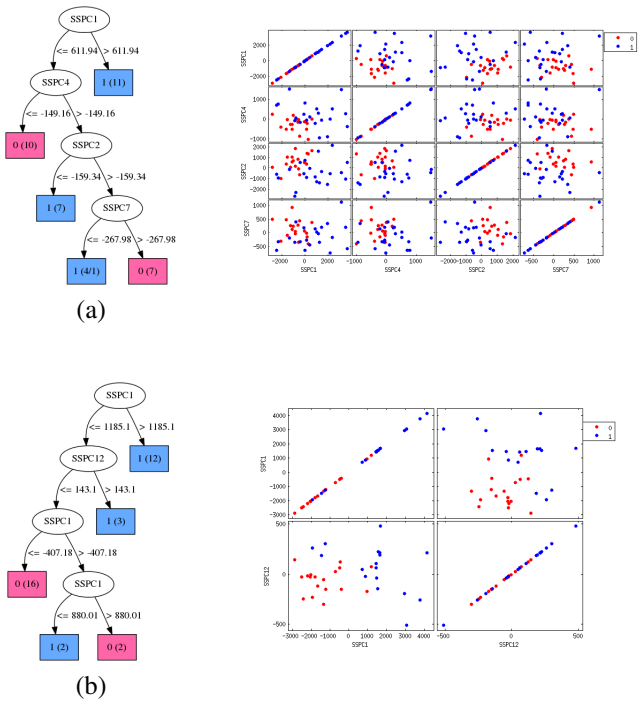


Fig. 3: Decision tree diagrams and scatter plots showing the distribution of subject scores for the PCs chosen by the classifier. No pre-selection of features is applied. (a): MSA vs HC; (b): PSP vs HC (Note: For the PSP group only two PCs [SSPC1 & SSPC12] were used in the classification).

TABLE I: CLASSIFIER LOOCV PERFORMANCE FOR REDUCED NUMBER OF FEATURES, I.E., THE FIRST TWO COMPONENTS ACCORDING TO THE HIGHEST AMOUNT OF VARIANCE. THE COLUMN PERF. INDICATES THE PERCENTAGE OF SUBJECT CASES CORRECTLY CLASSIFIED; SENSITIVITY THE PERCENTAGE OF CORRECTLY CLASSIFIED PATIENTS; AND SPECIFICITY THE PERCENTAGE OF CORRECTLY CLASSIFIED HEALTHY CONTROLS

Group	Perf.	Sensitivity	Specificity
PD (38)	63.2	100	22.2
MSA (39)	74.3	83.3	76.2
PSP (35)	80	70.6	88.9

modify the subject scores (like for the SR method) so that the classifier can find better decision boundaries to efficiently separate the groups.

Unlike the PD-HC group, the MSA-HC group as illustrated in Fig. 4(b) has a better separation with the two decision boundaries chosen by the classifier. Only 6 out of 39 subjects are misclassified. Note that for the PSP-HC group the classifier uses only one feature, i.e., SSPC1, out of the available two features to separate the two groups; 5 out of 17 PSP subjects are misclassified.

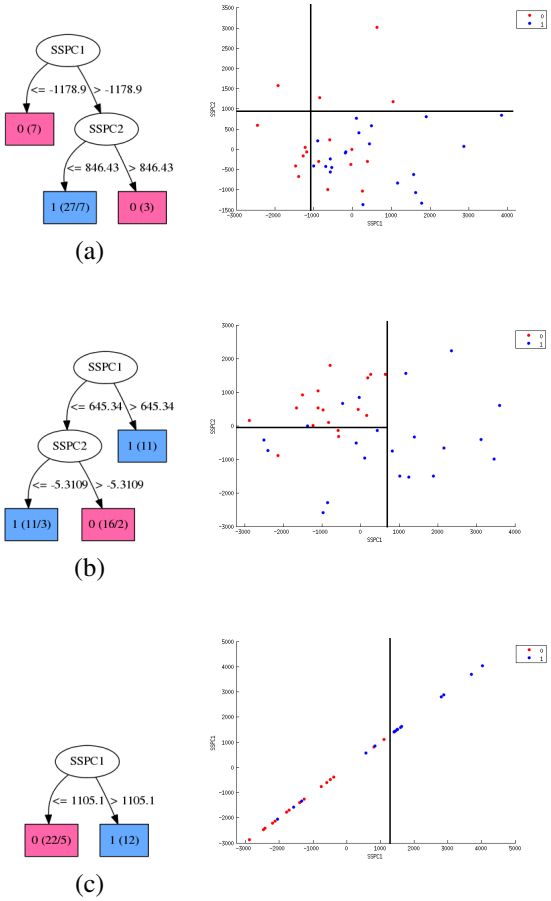


Fig. 4: The decision tree diagrams and scatter plots showing the distribution of subject scores for the first two features obtained from the LOOCV process. (a): PD vs HC; (b): MSA vs HC; (c): PSP vs HC.

D. Decision trees with the subject z-score on a combined pattern as the single feature

In the next experiment, the subject z-score determined by the SR method in the study by Teune phet al. [8] is used as a single feature for the decision tree classification. This feature is the result of a linear combination of the best PCs according to AIC (for details see Section III-A). Note that we used only this single feature (the subject z-score) as input to the decision tree classifier to separate the patient group from the healthy controls. The results are shown in Fig. 5 and Table II.

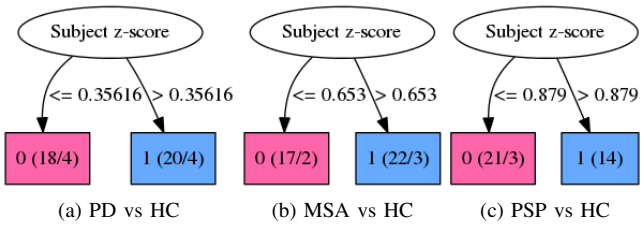


Fig. 5: The trees obtained after using the subject z-score on the combined pattern as the single feature for classification.

TABLE II: SUMMARY OF THE DECISION TREE CLASSIFICATION WITH THE Z-SCORE ON THE COMBINED PATTERN AS THE SINGLE FEATURE

Group	Perf.	Sensitivity	Specificity
PD (38)	79	80	77.8
MSA (39)	87.2	90.5	83.3
PSP (35)	91.4	82.4	100

In Fig. 5a the tree chooses a cutoff value of 0.36 as the threshold for the single z-score feature to split the dataset, with 14 out of 18 healthy controls and 16 out of 20 PD subjects correctly classified. These results correspond to the 80% sensitivity and 77.8% specificity at the z-score cutoff value of 0.45 as reported in the study by Teune phet al. [8]. That the cutoff values are not identical can be explained as follows. Since the z-scores take a discrete number of values, there can be a small interval of cut-off values which lead to the same sensitivity and specificity (for both the SR and DT methods). The decision tree method uses a mechanism called phinformation gain to sort the thresholds in ascending order and then chooses the first threshold. For example, the cut-off interval for the PD group was [0.36,0.45] (with the same sensitivity and specificity), and the decision tree method chose the first value, which is 0.36. For testing new data samples, a mid-value threshold should be considered to avoid a reduction in specificity.

Interestingly, the DT method produces exactly the same values for sensitivity and specificity as the SR method of Teune phet al. [8], although with small differences in z-score cut-off values. That is to say, at thresholds 0.65 and 0.88 the DT results correspond to the 90.5% sensitivity, 83.3% specificity results for the MSA group, and the 82.4% sensitivity, 100% specificity results for the PSP group, respectively, in the same study [8]. Therefore, with the same single feature (z-score) obtained from a linear combination of the best PCs, the decision tree method is as capable as the SR method (with optimal cut-point value determined from the ROC curve) to obtain high classification performance.

E. Information gain versus Youden index

Here we illustrate in more detail that maximising the information gain by the DT method and maximising the Youden index [20], [21] in the SR method lead to identical results. We conjecture that this identity holds in more generality, although we do not have a proof at this point.

Let us consider a data set with healthy and non-healthy cases and compute the optimal split of this data set based on a single attribute according to two different criteria: information gain (as used in decision tree classifiers) and the Youden index. We will illustrate by an example that these two different measures give identical results.

a) Computing the information gain: Let T be a set of cases, where each case belongs to one of k classes $C_1, C_2, \ldots, C_k$ (for example, $k = 2$, i.e., healthy and diseased). Let $freq(C_j, T)$ be the number of cases belonging to class C_j.

The phinformation of T is:

$$info(T) = -\sum_{j=1}^{k} \frac{freq(C_j,T)}{|T|} \log_2 \left(\frac{freq(C_j,T)}{|T|} \right) \quad (1)$$

When T is split into subsets $T_1, T_2, \ldots, T_n$ by some attribute X which has n outcomes, the phexpected information of T with respect to X is:

$$info_X(T) = \sum_{i=1}^{n} \frac{|T_i|}{|T|} info(T_i) \quad (2)$$

Now consider the complete data set T, with attributes $X_1, X_2, \ldots$, and two classes, healthy and diseased. The proportion of healthy cases is p_H, the proportion of disease cases is p_D. Let $info(T)$ be the phinformation (entropy) of T. (For a pure set, for example if there are only healthy cases, $info(T) = 0$.) Consider an attribute X and a phsplit value V of this attribute. Split the data set T into two subsets T_1 and T_2:

$$\begin{cases} T_1 &= \text{all cases from } T \text{ where } X \leqslant V \\ T_2 &= \text{all cases from } T \text{ where } X > V \end{cases} \quad (3)$$

The expected phinformation of this partition of T is denoted by $info_X^{(V)}(T)$.

The information gain is: $gain_X^{(V)}(T) = info(T) - info_X^{(V)}(T)$. In order to find the optimal split of the data set, one computes $gain^{(V)}(X)$ for all attributes X and all split values V. Then the attribute X which maximizes $gain_X^{(V)}$ is chosen as the first node of the tree, with V the corresponding split value.

b) Youden index: For distinguishing between individuals with and without a disease, the phYouden index [20], [21] is often used as a measure of overall diagnostic effectiveness. This index is defined by $J = TPR - FPR$, with TPR the true positive rate (fraction of true positives out of all positives), and FPR the false positive rate (1-fraction of true negatives out of all negatives). In other words, J is the maximum vertical distance between the ROC curve and the diagonal or chance line. Note that TPR equals sensitivity and FPR equals 1-specificity, so that J is equal to sensitivity+specificity-1.

c) Example: Consider now an example data set T with six cases, two healthy (labeled H) and four diseased (labeled D). Let us consider the disease cases as positives and the healthy cases as negatives. We now consider all possible choices for the split point; let us indicate the cases by 0,1,2,...,6. This leads to the seven pictures in Fig. 6.

TABLE III: THE YOUDEN INDEX AND INFORMATION GAIN COMPUTED FOR ALL THE SEVEN CASES IN FIG. 6

Case No	0	1	2	3	4	5	6
Youden index	0	0.5	0.25	0.75	0.5	0.25	0
Information gain	0	0.32	0.05	0.46	0.25	0.11	0

For all these cases we have computed the Youden index and the information gain $gain_X^{(V)}(T) = info(T) - info_X^{(V)}(T)$, where V refers to the possible cases 0,1,2,...,6 for choosing the split value. Table III shows the results.

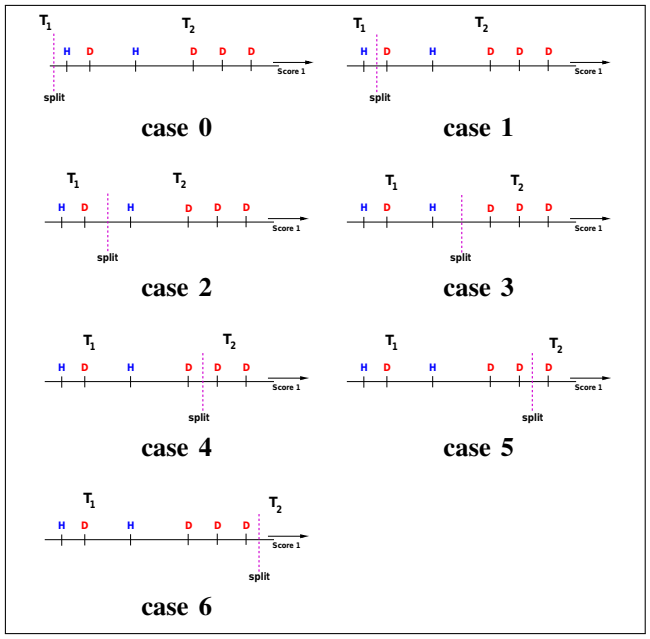


Fig. 6: All possible cases for the split point

As can be seen in Table III, Case 3 has both the highest Youden index J and the highest information gain $gain_X^{(V)}(T)$. This illustrates the relationship between the information gain (the mechanism used in the C4.5 decision tree inducer to determine thresholds) and the Youden index used to determine the best cut-off point (best combination of sensitivity and specificity) on the ROC curve, as used in the SR method [8].

F. Pairwise disease-group comparisons

In pairwise binary classification we do direct comparisons of each disease against another (that is, excluding the healthy group).

In this experiment, the LOOCV procedure is carried out in the usual way; however, we combine the SR procedure and decision tree method in two different ways. In the first approach, the subject scores of the best components obtained by the SR procedure and AIC (that is, without linearly combining the best components) are used as features for training the decision tree classifier. In the second approach, we linearly combine the best components to form a single pattern and the subject score on the combined pattern is used for training, as in Section III-D. The left-out subject is then tested on the best components (approach 1), or the combined pattern (approach 2), respectively.

For the PD vs MSA comparison the performance is better than for the other comparisons, both for individual and combined PCs. Similarly, the PD vs PSP classification performs well, especially when the PCs are combined. Note that the performance is lowest for the PSP group versus the MSA group. This can be attributed to the fact that PSP and MSA have a quite similar disease pattern [22]. As can be seen, combining PCs to form a single pattern is always better than using individual PCs for the pairwise comparisons.

TABLE IV: PAIRWISE DISEASE-GROUP COMPARISONS: CLASSIFIER LOOCV PERFORMANCE FOR (1) SUBJECT SCORES ON PCS SELECTED BY THE SR PROCEDURE AND AIC; AND (2) SUBJECT SCORES ON THE COMBINED PATTERN. FOR EACH PAIR OF DISEASE GROUPS A AND B, SENSITIVITY SS THE PERCENTAGE OF CORRECTLY CLASSIFIED SUBJECTS OF GROUP A, AND SPECIFICITY THE PERCENTAGE OF CORRECTLY CLASSIFIED SUBJECTS OF GROUP B

	Subject scores on individual PCs			Subject scores on combined pattern		
Group	Perf.	Sensitivity	Specificity	Perf.	Sensitivity	Specificity
PD vs MSA (41)	75.6	70	81	90.2	90	90.5
PD vs PSP (37)	70.3	85	52.9	81.1	80	82.4
MSA vs PSP (38)	63.2	66.7	58.8	65.8	61.9	70.6

IV. DISCUSSION

The SR method was found to work better than the DT method, especially when considering all or a few features for the DT method. In most cases a major difference was notable in the performance of the PD vs HC classification, which can be attributed to the fact that the PD-related pattern is very similar to the healthy pattern. Additionally, in the PD vs HC comparison, the principal components generated have less variance. Hence, a combination of several best components yields better results, which is exactly what the SR method does.

Furthermore, when the same single z-score feature corresponding to the combined pattern in the SR method is used in the DT classification (see Section III-D), the performance is as high as that of the stepwise regression method [8]. The pairwise disease comparisons yielded quite an impressive performance, especially for the PD vs MSA group, when compared to those in Mudali phet al. [9]. Combining the SR procedure with the DT method improved performance in the separation of some disease groups. Therefore, the robust feature obtained using the SR procedure could be used in the DT method to improve classification.

In a previous study [9], we compared the performances of several classification methods, i.e., random forest, nearest neighbors, classification and regression trees (CART), linear support vector machine (SVM), linear discriminant analysis (LDA), and naive Bayes in the separation of parkinsonian syndromes. Indeed some of these classifiers, like linear SVM and nearest neighbors, performed better than the decision tree method. However, our specific purpose in this paper was to compare the classification performances of the SR and DT methods, also to show that the classification performance improves when both methods are combined.

V. CONCLUSION

Covariance patterns were extracted from four distinct groups of FDG-PET data using the SSM/PCA method. The subject scores served as the feature set and input to the C4.5 decision tree classification algorithm. Classifiers were constructed from distinct groups for future prediction of new unlabeled subject images. Validation of classifiers was performed using the

leave-one-out method. The decision tree results were compared to the scatter plots and receiver operating characteristic (ROC) curves obtained in the stepwise regression method.

In some instances, the DT results are still not competitive with the SR method. To maximise classifier performance the decision tree method would require several horizontal and vertical decision boundaries to separate the dataset (especially for the PD group), since the subject scores overlap in the feature space. But this could lead to a high generalization error. Hence, it is preferable to combine the features to form one robust feature (subject z-score) which is capable of separating the groups while minimizing the generalization error. In fact, when we included the z-score feature (as determined by Teune phet al. [8]) in the DT classification, we obtained identical results for the C4.5 algorithm and the SR methods. Therefore, we can improve the DT method by using the linearly combined features obtained by the SR procedure. It would be interesting to study the performance of a multi-class classification of all parkinsonian syndromes, i.e., PD vs MSA vs PSP using SR feature(s) in the DT classification. Unfortunately, with the SR method in its current form only two groups can be compared.

Nevertheless, given the small size of the current datasets the decision tree method is highly promising. In addition it provides a visual understanding of the classification results and accommodates multi-class classification, as reported in Mudali phet al. [9]. In the long run, we need to devise means of obtaining a more diverse set of features and / or a larger set of training data for the decision tree to perform even better.

REFERENCES

[1] P. G. Spetsieris, Y. Ma, V. Dhawan, and D. Eidelberg, "Differential diagnosis of parkinsonian syndromes using PCA-based functional imaging features," NeuroImage, vol. 45, no. 4, pp. 1241–1252, 2009.

[2] P. Wu, J. Wang, S. Peng, Y. Ma, H. Zhang, Y. Guan, and C. Zuo, "Metabolic brain network in the Chinese patients with Parkinson's disease based on 18 F-FDG PET imaging," Parkinsonism & related disorders, vol. 19, no. 6, pp. 622–627, 2013.

[3] J. R. Moeller, S. C. Strother, J. J. Sidtis, and D. A. Rottenberg, "Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data," J Cereb Blood Flow Metab, vol. 7, no. 5, pp. 649–58, 1987.

[4] M. Fukunda, J. Mentis, M, Y. Ma, V. Dhawan, A. Antonini, E. Lang, A, M. Lozano, A, J. Hammerstad, K. Lyons, C. Koller, W, R. Moller, J, and D. Eidelberg, "Networks mediating the clinical effects of pallidal brain stimulation for Parkinson's disease. a PET study of resting-state glucose metabolism," BRAIN, vol. 124, pp. 1601–1609, 2001.

[5] J. R. Moeller, T. Ishikawa, V. Dhawan, P. Spetsieris, F. Mandel, G. E. Alexander, C. Grady, P. Pietrini, and D. Eidelberg, "The metabolic topography of normal aging.," J Cereb Blood Flow Metab, vol. 16, no. 3, pp. 385–98, 1996.

[6] P. G. Spetsieris and D. Eidelberg, "Scaled subprofile modeling of resting state imaging data in Parkinson's disease: Methodological issues," NeuroImage, vol. 54, no. 4, pp. 2899–2914, 2011.

[7] S. Peng, Y. Ma, P. G. Spetsieris, P. Mattis, A. Feigin, V. Dhawan, and D. Eidelberg, "Characterization of disease-related covariance topographies with SSMPCA toolbox: Effects of spatial normalization and PET scanners," Human brain mapping, vol. 35, no. 5, pp. 1801–1814, 2014.

[8] L. K. Teune, R. J. Renken, D. Mudali, B. M. D. Jong, R. A. Dierckx, J. B. T. M. Roerdink, and K. L. Leenders, "Validation of parkinsonian disease-related metabolic brain patterns," Movement Disorders, vol. 28, no. 4, pp. 547–551, 2013.

[9] D. Mudali, L. K. Teune, R. J. Renken, K. L. Leenders, and J. B. T. M. Roerdink, "Classification of Parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features," Computational and Mathematical Methods in Medicine, vol. Article ID 136921, pp. 1–10, 2015.

[10] D. Mudali, L. K. Teune, R. J. Renken, K. L. Leenders, and J. B. T. M. Roerdink, "Comparison of decison tree and stepwise regression methods in classification of FDG-PET data," in Third European Conference on Clinical Neuroimaging, March 31-April 1, Lille, France, p. 16, 2014. Abstract.

[11] J. R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, USA, 1993.

[12] J. R. Quinlan, "Learning decision tree classifiers," ACM Computing Surveys, vol. 28, no. 1, pp. 71–72, 1996.

[13] K. Polat and S. Güneş, "A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems," Expert Systems with Applications, vol. 36, no. 2, Part 1, pp. 1587–1592, 2009.

[14] T. Johnsson, "A procedure for stepwise regression analysis," Statistical Papers, vol. 33, no. 1, pp. 21–29, 1992.

[15] B. Thompson, "Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guidelines Editorial," Educational and Psychological Measurement, vol. 55, pp. 525–534, 1995.

[16] H. Akaike, "A new look at the statistical model identification," IEEE Transactions on Automatic Control, vol. 19, pp. 716–723, 1974.

[17] L. K. Teune, A. L. Bartels, B. M. de Jong, A. T. Willemsen, S. A. Eshuis, J. J. de Vries, J. C. van Oostrom, and K. L. Leenders, "Typical cerebral metabolic patterns in neurodegenerative brain diseases," Movement Disorders, vol. 25, no. 14, pp. 2395–2404, 2010.

[18] D. Eidelberg, "Metabolic brain networks in neurodegenerative disorders: a functional imaging approach," Trends in Neurosciences, vol. 32, no. 10, pp. 548–557, 2009.

[19] P. G. Spetsieris, V. Dhawan, and D. Eidelberg, "Three-fold cross-validation of parkinsonian brain patterns," in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 2906–2909, 2010.

[20] W. Youden, "An index for rating diagnostic tests," Cancer, vol. 3, pp. 32–35, 1950.

[21] E. F. Schisterman, N. J. Perkins, A. Liu, and H. Bondell, "Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples," Epidemiology, vol. 16, no. 1, pp. 73–81, 2005.

[22] T. Eckert, C. Tang, Y. Ma, N. Brown, T. Lin, S. Frucht, A. Feigin, and D. Eidelberg, "Abnormal metabolic networks in atypical parkinsonism," Movement Disorders, vol. 23, no. 5, pp. 727–733, 2008.