LVQ and SVM Classification of FDG-PET Brain Data

Mudali, Deborah; Biehl, Michael; Leenders, KlausL.; Roerdink, Jos

Published in:
Advances in Self-Organizing Maps and Learning Vector Quantization

DOI:
10.1007/978-3-319-28518-4_18

Publication date:
2016

LVQ and SVM Classification of FDG-PET Brain Data

Deborah Mudali[1], Michael Biehl[1], Klaus L. Leenders[2], and Jos B.T.M. Roerdink[1,3]

[1] Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands,
d.mudali@rug.nl, m.biehl@rug.nl, j.b.t.m.roerdink@rug.nl,
WWW home page: http://www.cs.rug.nl/svcg

[2] Department of Neurology, University Medical Center Groningen, University of Groningen, The Netherlands,
k.l.leenders@umcg.nl

[3] University of Groningen, University Medical Center Groningen, Neuroimaging Center, Groningen, The Netherlands

Abstract. We apply Generalized Matrix Learning Vector Quantization (GMLVQ) and Support Vector Machine (SVM) classifiers to fluorodeoxyglucose positron emission tomography (FDG-PET) brain data in the hope to achieve better classification accuracies for parkinsonian syndromes as compared to the decision tree method which was used in previous studies.
The classifiers are validated using the leave-one-out method. The obtained results show that GMLVQ performs better than the previously studied decision tree (DT) method in the binary classification of group comparisons. Additionally, GMLVQ achieves a superior performance over the DT method regarding multi-class classification. The performance of the considered SVM classifier is comparable with that of GMLVQ. However, in the binary classification, GMLVQ performs better in the separation of Parkinson's disease subjects from healthy controls. On the other hand, SVM achieves higher accuracy than the GMLVQ method in the binary classification of the other parkinsonian syndromes.

Keywords: Learning Vector Quantization, Support Vector Machine, Parkinsonian syndromes, Classification

1 Introduction

Diagnosis of neurodegenerative diseases (NDs), especially at an early stage, is very important to affect proper treatment [1], but it is still a challenge [19]. Nevertheless, some studies report considerable success in differentiating between some of these diseases [23]. In fact, promising classification performances were obtained for the multiple system atrophy (MSA) and progressive supranuclear palsy (PSP) groups versus the healthy control group in the study [16] where the decision tree (DT) method was used. The same study showed that discriminating the Parkinson's disease (PD) group from healthy controls (HC) on the basis of PET brain scan imaging data remains a challenge. Therefore, in this paper other classification methods are applied in the hope to improve classification of parkinsonian syndromes, in particular PD, MSA, and PSP. The classification methods used in this study are Generalized Matrix Learning Vector Quantization (GMLVQ) and Support Vector Machine (SVM).

LVQ is a method which uses prototypes assigned to each class. A new case is classified as belonging to the class of the closest prototype [12]. In the training phase, a set of appropriately chosen prototypes is computed from a given set of labeled example data. This training process can be based on a suitable cost function, as for instance in the so-called Generalized LVQ (GLVQ) introduced in [17]. The conceptional extension to matrix-based relevance learning was introduced in [18]; simpler feature weighting schemes had been considered earlier in [10]. Relevance learning provides insight into the data in terms of weighting features and combinations of features in the adaptive distance measure. Moreover, GMLVQ allows for the implementation of multi-class classification in a straightforward way.

The Support Vector Machine is a supervised learning method for classifying data by maximizing the margin between the defined classes, see for instance [4, 7]. The aim of SVM training is to minimize the classification error while maximizing the gap or margin between the classes by computing an optimally separating hyperplane. The training data points that lie closest to the hyperplane define the so-called support vectors [6, 25]. This method was originally designed for binary classification but has been extended to multi-class classification, see for instance [11] and references therein. Moreover, several studies including [9, 13] have used SVM to classify neurodegenerative diseases with high accuracy. Other examples of SVM applications like biological data mining are described in [7].

2 Method

The data used in this study is described in [22]. The brain data were obtained from 18 healthy controls (HC), 20 Parkinson's Disease (PD), 17 progressive supranuclear palsy (PSP) and 21 multi system atrophy (MSA) cases. We apply the scaled subprofile model with principal component analysis (SSM/PCA), based on the methods by Spetsieris et al. [21], to the datasets to extract features. The method was implemented in Matlab R2014a. The SSM/PCA method [14, 15, 20] starts by double centering the data matrix and then extracts metabolic brain patterns in the form of principal component images, also known as *group invariant subprofiles*. The original images are projected onto the extracted patterns to determine their weights, which are called *subject scores*. The subject scores then form the features that are input to the classifiers to classify the subject brain images. Because of the application of the PCA method, the computed subject scores are dependent on the whole input dataset, an unusual circumstance in the standard situation. This makes the number of features extracted equal to the number of samples in the dataset.

A leave-one-out cross validation (LOOCV) of the classifiers is performed to predict their performance on new subject cases. For each run, a subject (test sample) is left out, then the SSM/PCA process is performed on the rest of the subjects (training set) to obtain their scores on the principal components. These subject scores are then used to train the GMLVQ and the SVM classifiers. The test subject is projected onto the invariant profiles to obtain its scores on the extracted profiles. Then the test subject scores are used to evaluate the trained classifier. The sensitivity (true positive rate), specificity (true negative rate) and classifier accuracy are determined. Note that the test

subject is removed *before* the SSM/PCA process in order to deal with dependencies of the extracted features on both the training and test sets. In addition, the test set receiver operating characteristic (ROC) curve and Nearest Prototype Classifier (NPC) confusion matrix are computed for all the left-out subjects. The area under curve (AUC) of the ROC curve is a measure of the ability of the features (i.e., subject scores on the principal components) to separate the groups.

For both the SVM and GMLVQ classifiers, we do binary and multi-class classification. The binary classification involves comparing the distinct disease groups (PD, PSP, and MSA) with the healthy control group. The multi-class classification concerns the comparison of all the groups, i.e., HC versus PD versus PSP versus MSA (a total of 76 subjects), as well as only the disease groups, i.e., PD versus PSP versus MSA (a total of 58 subjects). The goal is to determine the class membership (healthy or diseased) of a new subject of unknown diagnosis and also determine the type of parkinsonian syndrome.

For SVM training and testing, we use the Matlab R2014a functions "fitcsvm" and "predict", respectively, with default parameters and a linear kernel, representing a large margin linear separation in the original feature space. Also, all features are centered at their mean in the dataset and scaled to have unit standard deviation. The "fitcsvm" returns an SVM classifier which can be used for classification of new data samples. It also provides class likelihoods which can be thresholded for an ROC analysis. For the SVM multi-class classification we use the LIBSVM library [5] with the one-against-one method, since the previously mentioned Matlab functions support only binary classification. The one-against-one method has a shorter training time than the one-against-all, as reported in [11].

As for GMLVQ, we employ it in its simplest setting with one prototype $\boldsymbol{w_k}$ per class. A global quadratic distance measure of the form $d(\boldsymbol{w_k}, \boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{w_k})^T \Lambda (\boldsymbol{x} - \boldsymbol{w_k})$ is used to quantify the dissimilarity of an input vector $\boldsymbol{x}$ and the prototypes. The measure is parameterized in terms of the positive semi-definite relevance matrix Λ [18]. Both, prototypes and relevance matrix are optimized in the training process which is guided by a suitable cost function [18]. We employed the gmlvq-toolbox [2], which performs a batch gradient descent minimization with automated step size control, see [2] for details. All the results presented here were obtained using the default parameter settings of [2]. After 100 gradient steps, the training errors and cost function appeared to have converged in all considered classification problems.

3 Results

3.1 Generalized Matrix Relevance LVQ (GMLVQ)

As mentioned earlier, in order to validate the classifiers the training process is repeated with one test subject removed from the training set before applying the SSM/PCA process. This section presents the LOOCV results for the distinct disease groups versus the healthy control group in the binary and multi-class classification. Important to note is that all the features (100%) as extracted from the brain image data using the SSM/PCA method are provided to the GMLVQ classifier. In the tables, sensitivity (%) is the per-

centage of correctly classified patients, specificity (%) the percentage of correctly classified healthy controls, and AUC is the area under the ROC curve. In addition, the corresponding results are visualized in terms of projections on the leading two eigenvectors of the relevance matrix. This exploits the fact that GMLVQ displays a tendency to yield low-rank matrices which correspond to an intrinsically low-dimensional representation of the feature space [3, 18]. Additionally, we include the corresponding plots showing diagonal and off-diagonal matrix elements for one LOOCV iteration as an example illustration.

Binary Classification The objective here is to separate the individual disease groups from the healthy control group. The GMLVQ results are shown in Table 1.

Table 1: GMLVQ Classifier performance in LOOCV for the different data sets (patients vs healthy controls, number of cases in brackets). The column Perf.(%) indicates the percentage of subject cases correctly classified per group. Perf. as well as Sensitivity and Specificity correspond to the Nearest Prototype Classifier (NPC).

Feature set(size)	Perf. (%)	Sensitivity (%)	Specificity (%)	AUC
PD-HC (38)	81.6	75	88.9	0.84
MSA-HC (39)	92.3	90.5	94.4	0.99
PSP-HC (35)	88.6	82.4	94.4	0.97

The results in Table 1 are much better than those of the decision tree as reported in [16]. In fact a tremendous improvement can be seen in the PD vs HC group, whose LOOCV performance has increased from 63.2% (decision trees) to 81.6% (GMLVQ). The use of the relevance matrix to weight features according to their relevance appears to boost performance. An illustration is shown in Fig. 1 where the training data points are displayed in a feature space of the two leading eigenvectors of the relevance matrix. Observe that the subject scores do not overlap after the GMLVQ classifier training phase, which corresponds to error-free classification of the training set. Further, the resulting AUC measures (for the different groups) are relatively high. This means that the GMLVQ weighted features are very suitable for separating the groups.

As observed in Fig. 1, the PD vs HC comparison shows a clear separation between the PD group and the healthy group. Apart from a few outliers, most of the data points cluster around the specific prototypes, i.e., the two bigger circles that each represent a class. Further, the relevance matrix histogram shows the features and their diagonal weights as used in the classification process. For example, in the PD vs HC group feature 1 was weighted the highest, implying that feature 1 carries relevant information required to separate the two groups. As a matter of fact, the highly weighted feature should be given more attention, i.e., critically analyze the principal component image corresponding to this feature to gain insights from the clinical perspective.

Multi-class classification Here we show the results for the LOOCV of the GMLVQ classifier on the multi-class datasets, i.e., the classification of all the four classes, and the three disease classes, respectively. The latter is considered separately, because the

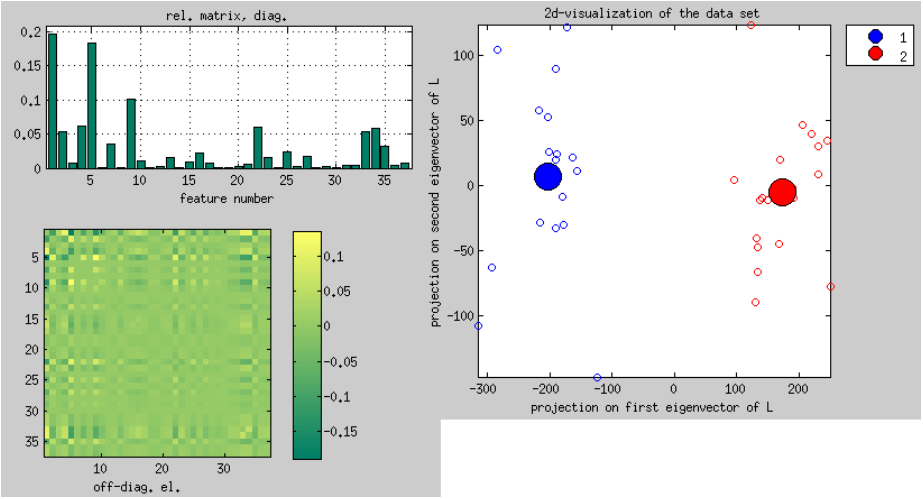


Fig. 1: Illustrations of the results of a single GMLVQ training process in the LOOCV of the PD vs HC two class-problem, 1 = HC, 2 = disease group. Graphs show diagonal relevances (upper left), and off-diagonal relevance matrix elements (lower left). The visualization of the training data in terms of their projection on the two leading eigenvectors of the relevance matrix is displayed on the right.

main task in clinical practice is to distinguish the three parkinsonian syndromes. Additionally, for the four-class comparison, we include the HC group because we want to build a classifier which can also distinguish a healthy subject from the parkinsonian groups. The results are shown in Tables 2 and 3 for four-class comparison and three disease groups, respectively. Also included are the scatter plots showing the distribution of training data points in the two-dimensional projection of the feature space in a single run of the training process.

Four-class comparison. From the results in Table 2, we notice that most of the misclassified HC subjects are classified as PD and *vice versa*. As already observed in [16], the PD and HC subjects have a closely related metabolic pattern. Likewise, the PSP and MSA groups display a similarity, in view of the fact that four (majority of the misclassification) MSA subjects are misclassified as PSP.

Three-class comparison. The classifier results show that the PD group is clearly separable from the other two disease groups. On the other hand, the PSP and MSA groups seem to overlap more strongly. We observe that the majority of the misclassification for both the PSP and MSA belong to either classes, which shows that these two groups are quite similar. In fact, it is known that PSP and MSA are hard to distinguish because the patients with either disorders show similar reduction in striatal and brain stem volumes [8].

Table 2: Four-class problem: The table shows the number of subject images correctly classified for each class in bold and the overall performance in percentage as obtained in the LOOCV.

GMLVQ classification	HC	PD	PSP	MSA
HC(18)	**14**	3	1	0
PD(20)	5	**13**	1	1
PSP(17)	2	2	**11**	2
MSA(21)	0	1	4	**16**
Class accuracy (%)	77.8	65	64.7	76.2
Overall performance (%)	71.1			

Table 3: Three-class problem: The table shows the number of subject images correctly classified for each class in bold with the overall LOOCV performance in percentage.

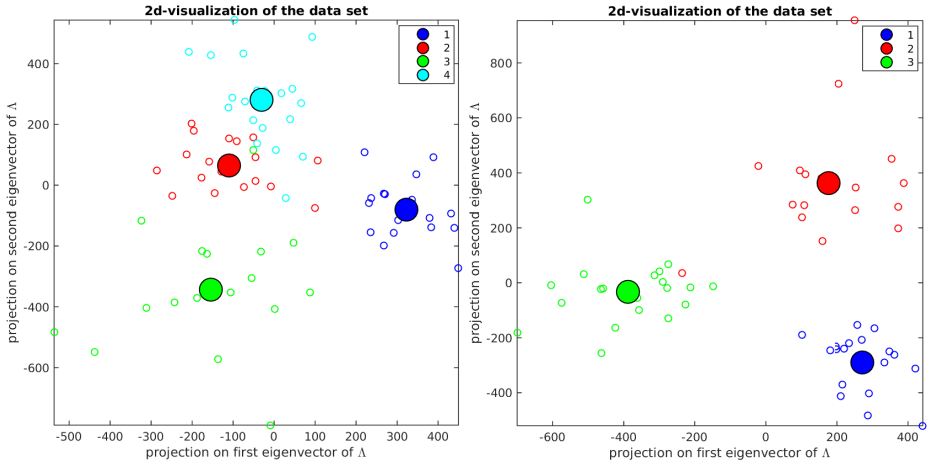
GMLVQ classification	PD	PSP	MSA
PD(20))	**19**	0	1
PSP(17)	2	**12**	3
MSA(21)	2	3	**16**
Class accuracy (%)	95	70.6	76.2
Overall performance (%)	81.03		

Visualization of the data points. The scatter plots show the training data points with respect to their projections on the two leading eigenvectors of the relevance matrix. It can be observed in Fig. 2(a) that the PSP and healthy groups are clearly separable from the rest of the groups. But a small overlap exists between the PD and MSA groups even in the training set. Meanwhile, the three-class comparison in Fig. 2(b) shows a clear separation among the disease groups. This is encouraging since we are generally interested in distinguishing between the parkinsonian syndromes.

3.2 Support Vector Machine (SVM)

Next we show the results of the leave-one-out cross validation of the SVM classifier for the different groups, both in a binary and multi-class comparison. Note that, as before, a subject is left out before the SSM/PCA process.

Binary Classification Here, the classifier was used to separate each disease group from the healthy control group to determine its classification performance. As seen in Table 4, apart from the PD vs HC comparison, the other groups' performances improve in comparison to GMLVQ (cf. Table 1). However, the AUC measures for MSA and PSP are lower than those of GMLVQ, indicating that it outperforms the SVM when choosing an appropriate class bias to modify the nearest prototype classification. In comparison to the linear SVM in [16], the results differ because different features have



(a) Four class problem; 1=HC, 2=PD, 3=PSP, 4=MSA (b) Three class problem; 1=PD, 2=PSP, 3=MSA

Fig. 2: The visualization of the training data with respect to their projections on the two leading eigenvectors of the relevance matrix as observed in a single run of GMLVQ training.

Table 4: SVM classifier LOOCV performance for the different data sets (patients vs healthy controls, number of cases in brackets). The column Perf.(%) indicates the percentage of subject cases correctly classified per group, Sensitivity (%) the percentage of correctly classified patients, and Specificity (%) the percentage of correctly classified healthy controls.

Feature set(size)	Perf. (%)	Sensitivity (%)	Specificity (%)	AUC
PD-HC (38)	76.3	75	77.8	0.84
MSA-HC (39)	94.9	90.5	100	0.97
PSP-HC (35)	91.4	88.2	94.4	0.92

been used. Furthermore, here the LOOCV is done correctly by removing the test subject from the training set before applying the SSM/PCA method, whereas in [16] the SSM/PCA method was applied to all subjects to obtain the scores before the LOOCV was performed.

Multi-class Classification We also applied SVM to the multi-class datasets to determine its performance on larger datasets.

Four-class comparison. This involved the comparison of all the four groups, i.e., HC, PD, PSP, and MSA. In Table 5, the SVM four-group classification accuracy is slightly above chance level and lower than that of GMLVQ (see Table 2). But the classifier can separate the MSA group from the rest of the groups with an accuracy of 81%.

Table 5: Four-class problem: The confusion matrix and the overall performance of the SVM in the LOOCV scheme.

SVM classification	HC	PD	PSP	MSA
HC(18)	**12**	3	2	0
PD(20)	4	**12**	1	3
PSP(17)	1	2	**9**	5
MSA(21)	0	2	2	**17**
Class accuracy (%)	66.7	60	52.9	81.0
Overall performance (%)	65.8			

Three disease groups. This involved the comparison of only the disease groups, i.e., PD, PSP and MSA without the healthy group. The separation of the disease groups

Table 6: Three-class problem: The table shows the confusion matrix with the number of subject images correctly classified by the SVM for each class in bold and the overall LOOCV performance in percentage.

SVM classification	PD	PSP	MSA
PD(20))	**17**	1	2
PSP(17)	2	**10**	5
MSA(21)	3	2	**16**
Class accuracy (%)	85	58.8	76.2
Overall performance (%)	74.1		

using SVM yields a better performance accuracy than the separation of the four groups (including the healthy group). Also, as in the GMLVQ classification, the PD group appears to be well separated from PSP and MSA.

4 Discussion and Conclusion

Both GMLVQ and SVM were studied and tested for the binary and multi-class problems. In the binary classification, GMLVQ performs better than SVM in the PD vs HC comparison (performance of 81.6%), but both achieve the same sensitivity of 75%. However, SVM performs better in the MSA vs HC and PSP vs HC comparisons. For the two-class problems we also considered the Area under Curve (AUC) of the ROC, as it does not depend on the choice of a particular working point (threshold, class bias) in the classifier. In terms of the AUC, GMLVQ was seen to outperform or equal the performance of the SVM classifier. Additionally, in the multi-class problems, GMLVQ achieves a better accuracy than SVM.

The GMLVQ relevance matrix, which makes use of an adaptive weighting of features according to their discriminative power, displayed overall superior classification

performance. In particular, for the PD vs HC comparison which has been challenging to discriminate using decision trees, GMLVQ was able to separate PD from HC with an accuracy of 81.6%, better than SVM by a margin of 5.3%. Although SVM classification performance for the MSA vs HC and PSP vs HC comparisons is better than GMLVQ, the AUC measures show that GMLVQ achieves superior binary classification of the distinct groups. Overall, GMLVQ also achieves a better accuracy for the multi-class classification. In addition, when it comes to explaining the results to the physicians, GMLVQ is more intuitive than SVM. The analysis of the resulting relevance matrix allows for the identification of particularly relevant features and combinations of features. These results should trigger further investigations from the clinical perspective.

Clearly, the number of cases in the available data set is fairly small and our findings could be partly skewed by the small sample size. For instance, leave-one-out validation schemes are known to frequently yield unreliable estimates of performance. It is also possible that the performance of decision trees in [16], which was found inferior to GMLVQ and SVM, might improve significantly for larger data sets (see comparable work in [24]). We intend to extend our work in this direction as more data become available in the future. Moreover, variants of the considered classifiers could be considered, e.g., SVM with more powerful kernels or LVQ systems with several prototypes per class or local distance matrices [18].

References

1. Appel, L., Jonasson, M., Danfors, T., Nyholm, D., Askmark, H., Lubberink, M., Sörensen, J.: Use of 11C-PE2I PET in differential diagnosis of parkinsonian disorders. Journal of Nuclear Medicine 56(2), 234–242 (2015), http://jnm.snmjournals.org/content/56/2/234.abstract
2. Biehl, M.: A no-nonsense Matlab (TM) toolbox for GMLVQ (2015), http://www.cs.rug.nl/biehl/gmlvq.html
3. Bunte, K., Schneider, P., Hammer, B., Schleif, F.M., Villmann, T., Biehl, M.: Limited rank matrix learning, discriminative dimension reduction and visualization. Neural Networks 26, 159–173 (2012)
4. Burges, C.J.: A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 2(2), 121–167 (1998)
5. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011), software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm
6. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995), http://dx.doi.org/10.1007/BF00994018
7. Cristianini, N., Shawe-Taylor, J.: An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press (2000)
8. Eckert, T., Sailer, M., Kaufmann, J., Schrader, C., Peschel, T., Bodammer, N., Heinze, H.J., Schoenfeld, M.A.: Differentiation of idiopathic Parkinson's disease, multiple system atrophy, progressive supranuclear palsy, and healthy controls using magnetization transfer imaging. Neuroimage 21(1), 229–235 (2004)
9. Haller, S., Badoud, S., Nguyen, D., Garibotto, V., Lovblad, K., Burkhard, P.: Individual detection of patients with Parkinson disease using support vector machine analysis of diffusion tensor imaging data: initial results. American Journal of Neuroradiology 33(11), 2123–2128 (2012)

10. Hammer, B., Villmann, T.: Generalized relevance learning vector quantization. Neural Networks 15(8–9), 1059–1068 (2002)
11. Hsu, C.W., Lin, C.J.: A comparison of methods for multiclass support vector machines. Neural Networks, IEEE Transactions on 13(2), 415–425 (2002)
12. Kohonen, T.: The self-organizing map. Neurocomputing 21(1), 1–6 (1998)
13. Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehéricy, S., Benali, H.: Support vector machine-based classification of Alzheimers disease from whole-brain anatomical MRI. Neuroradiology 51(2), 73–83 (2009)
14. Moeller, J.R., Strother, S.C.: A regional covariance approach to the analysis of functional patterns in positron emission tomographic data. J Cereb Blood Flow Metab 11(2), A121–135 (1991)
15. Moeller, J.R., Strother, S.C., Sidtis, J.J., Rottenberg, D.A.: Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. J Cereb Blood Flow Metab 7(5), 649–58 (1987)
16. Mudali, D., Teune, L.K., Renken, R.J., Leenders, K.L., Roerdink, J.B.T.M.: Classification of Parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features. Computational and Mathematical Methods in Medicine Article ID 136921, 1–10 (2015)
17. Sato, A., Yamada, K.: Generalized learning vector quantization. Advances in neural information processing systems pp. 423–429 (1996)
18. Schneider, P., Biehl, M., Hammer, B.: Adaptive relevance matrices in learning vector quantization. Neural Computation 21(12), 3532–3561 (2009)
19. Silverman, D.H.: Brain 18F-FDG PET in the diagnosis of neurodegenerative dementias: comparison with perfusion SPECT and with clinical evaluations lacking nuclear imaging. Journal of Nuclear Medicine 45(4), 594–607 (2004)
20. Spetsieris, P.G., Eidelberg, D.: Scaled subprofile modeling of resting state imaging data in Parkinson's disease: Methodological issues. NeuroImage 54(4), 2899–2914 (2011), http://www.sciencedirect.com/science/article/pii/S1053811910013170
21. Spetsieris, P.G., Ma, Y., Dhawan, V., Eidelberg, D.: Differential diagnosis of parkinsonian syndromes using PCA-based functional imaging features. NeuroImage 45(4), 1241–1252 (2009), http://www.sciencedirect.com/science/article/pii/S1053811908013335
22. Teune, L.K., Bartels, A.L., de Jong, B.M., Willemsen, A.T., Eshuis, S.A., de Vries, J.J., van Oostrom, J.C., Leenders, K.L.: Typical cerebral metabolic patterns in neurodegenerative brain diseases. Movement Disorders 25(14), 2395–2404 (2010)
23. Van Laere, K., Casteels, C., De Ceuninck, L., Vanbilloen, B., Maes, A., Mortelmans, L., Vandenberghe, W., Verbruggen, A., Dom, R.: Dual-tracer dopamine transporter and perfusion SPECT in differential diagnosis of parkinsonism using template-based discriminant analysis. Journal of Nuclear Medicine 47(3), 384–392 (2006)
24. Westenberg, M.A., Roerdink, J.B.T.M.: Mixed-method identifications. In: Du Buf, J.M.H., Bayer, M.M. (eds.) Automatic Diatom Identification, Series in Machine Perception and Artificial Intelligence, vol. 51, chap. 12, pp. 245–257. World Scientific Publishing Co., Singapore (2002)
25. Zhang, X.: Using class-center vectors to build support vector machines. In: Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop. pp. 3–11 (Aug 1999)