TIPTEKNO'18
TIP TEKNOLOJİLERİ KONGRESİ
8-10 Kasım 2018 / Gazi Magosa / KKTC

Performance Analysis of Different Feature Selection Methods on Parkinson's Disease Diagnosis

Ozkan Cigdem*, Faezeh Yeganli†, Hasan Demirel‡
*ÖZHAK Engineering Ltd., Izmir, Turkey
†Department of Electrical and Electronics Engineering, Izmir University of Economics, Izmir, Turkey
‡Department of Electrical and Electronics Engineering, Eastern Mediterranean University, Famagusta, Via Mersin 10, Turkey
*ozkancigdem@ieee.org,†faezeh.yeganli@ieu.edu.tr,‡hasan.demirel@emu.edu.tr

***Abstract*—The neurodegenerative disorders are characterized by the progressive deterioration of the brain neurons. Parkinson's disease is the second common encountered disease which affects 1% of people over 60 years old. In this paper, identification of Parkinson's disease using structural magnetic resonance imaging together with computer-aided technology has been investigated. In order to compare morphological alterations between grey matter and white matter tissue volumes of Parkinson's disease and healthy controls, a voxel-based morphometry technique is used. The 3D masks for both tissue volumes are obtained separately by using the local differences between the groups and using two sample t-test method. The performances of five different feature ranking methods are compared and a feature-level fusion technique using a majority voting method is studied. In order to select the numbers of top-ranked features, an adaptive Fisher stopping criteria based feature selection method is used. Support vector machine with Gaussian kernel is utilized for classification of 40 Parkinson's disease patients and 40 healthy controls. The highest classification accuracy of 87.50% is obtained when the LLCFS feature ranking method is considered. However, using feature-level fusion improves the classification accuracy to 90.00%.**

Keywords*—*Parkinson's disease, SPM12, CAT12, Feature-level fusion.

I. Introduction

Neurodegenerative diseases affect the neurons in the human brain [1], [2]. Parkinson's disease (PD) is the second most common neurodegenerative disorders following Alzheimer's Disease (AD). Since the early diagnosis of the diseases is helpful to take precautions and develop required treatments, a Computer-Aided Diagnosis (CAD) method has been progressively used in neurodegenerative disorder imaging [1], [3], [4]. In neurodisease classification, in order to detect the disorder by using only one Magnetic Resonance Imaging (MRI) scan, there is a need to have a model generated from a large collection of diseased and healthy controls (HCs) datasets [4]. In literature, different neuroimaging methods, namely structural MRI (sMRI) [1], functional MRI (fMRI) [5], positron emission tomography (PET) [6], and single photon emission computed tomography (SPECT) [7] are studied. Since MRI has high spatial resolution neuroanatomy, high availability, good contrast, and no a requirement for any pharmaceutical injections, it has been widely used in literaure. While in fMRI, hemodynamic responses of the brain regarding to neural activities are considered, in sMRI, the atrophies and physical differences between the brain segments are taken into account. In this paper, the sMRI data of 40PD and 40HC are used. A Voxel-Based Morphometry (VBM) technique is utilized to capture the 3D Volumes of Interests (VOIs) which are obtained by analyzing group-wise comparisons of cross-sectional sMRI scans [8]. In order to extract the 3D VOIs, a statistical analysis using total intracranial volume (TIV) as a covariate and f-contrast for model building is used.

In neurodegenerative disorder diagnosis, mostly the differences in either GM or WM tissue map using VBM have been considered separately [1]–[3]. In this paper, the preprocessing of 40PD and 40HC datasets has been performed for both the GM and WM tissue volumes and the 3D masks are obtained for GM and WM, separately. To analyze the effects of GM and WM tissue volumes together, the 3D masked GM and WM data are combined by concatenating them in a single vector. Combining the two tissue volumes reduced the dimension of the data to a lower VOIs level. However, owing to the fact that the number of samples is more than the number of observations, a feature selection (FS) method might be used to remove the irrelevant, redundant, and noisy information from the data [9]. In this paper, five different filter based feature ranking approaches including both the supervised and the unsupervised learning have been investigated for the preprocessed, 3D masked, and the extracted GM and WM data sets. These approaches are Relief-F, Laplacian score (LS), unsupervised feature selection for multi-cluster data (MCFS), unsupervised discriminative feature selection (UDFS), and feature selection and kernel learning for local learning-based clustering (LLCFS) [9], [10]. A feature-level fusion technique which uses a majority voting method to combine the ranking order results of the five different feature ranking approaches. After combining the five outputs, a new ranked data is obtained and in order to determine the optimum number of top-ranked features, a Fisher Criterion (FC) stopping method which maximizes the class separation between the PD and HC is used [9]. The aim of using FC is to determine the optimal number of top-ranked features adaptively based on training data in each fold instead of identifying a fixed number of features. It also aims to determine a discriminative feature subset with high performance by using training data in each fold of classification algorithm [9]. A Support Vector Machine (SVM) method with Gaussian kernel and 10-fold cross validation is used as a classification approach. The proposed feature-level fusion approach has been performed for the combination of GM as well as WM datasets. The combination of the GM as well WM datasets, a source fusion technique, significantly increases the classification performances of using the GM and the WM datasets, separately [8], [9]. Furthermore, the experimental results indicate that among all five feature ranking techniques, using LLCFS with FC method provides the highest classification accuracy, since it considers feature selection as a regularization problem in which the features are nodes in a graph and a selection is a path through them [10]. However, the obtained results indicate that the classification accuracy of using proposed feature-level fusion technique outperforms that of using each feature ranking method individually. By using the feature-level fusion technique, a higher classification result is achieved for PD detection.

The remainder of this paper is arranged as follows: Section II provides statistics of the data namely materials used in the work and Section III describes the methodology used to design an automatic CAD tool. In Section IV, experimental results and the discussions for the proposed method are introduced. In Section V, the conclusion is drawn.

II. Materials

A. MRI Acquisition

Data used in this research are obtained from the Parkinson's Progression Markers Initiative (PPMI) data set (www.ppmi-info.org/data). The protocol included T1-weighted MRI images based on a scanner by Siemens with acquisition plane=sagittal, acquisition type=3D, coil=Body, flip angle=9.0 degrees, matrix X/Y/Z=240.0/256/176 pixel, mfg model=TrioTim, pixel spacing X/Y=1.0/1.0 mm, pulse sequence=GR/IR, slice thickness=1 mm, and TE/TI/TR=2.98/900/2300 ms.

B. Subjects

In this study, 40 PD patients (mean age$\pm$standard deviation (SD)= 62.34$\pm$8.28 years, range: 38.5–77.3 years, gender: 25M-15F) and 40 healthy subjects (mean age $\pm$ standard deviation (SD)= 61.91 $\pm$ 9.45 years, range: 39.7–78.9 years, gender: 27M-13F) are taken into account.

III. PD Classification Method

A. MRI Data Pre-processing and Statistical Analysis

The 3D T1-weighted sMRI images are pr-processed through taking advantage of statistical parameter mapping (SPM12) package (Welcome Trust Centre for Neuroimaging, London, UK; available at {http://www.fil.ion.ucl.ac.uk/spm} and its extension called as computational anatomy toolbox(CAT12) {http://www.neuro.uni-jena.de/cat/} implemented in Matlab 2017b. The 3D image data are downloaded in DICOM formats. Through using SPM12, all DICOM files are converted into NifTi formats. First of all, the anterior commissural (AC) point of all subject is coregistered to the central point space to have every image with the same central locations. VBM technique is an automated method to analyze tissue volumes between different subject groups. It takes into account the whole brain structure through comparing voxel-by-voxel, and discriminates the degenerated tissue concentrations by referencing the brain of HCs as a template. Then the data are segmented into six tissue probability maps regarding to the existing templates for each of six modalities. After segmentation of all data, DARTEL approach gives better results over VBM due to the fact that it has more precise inter-subject alignment of MRI images [4]. In order to create deformation field of every data, instead of creating template for studied data, existed DARTEL template created from 555 healthy controls are used. All data are registered to standard Montreal Neurological Institute (MNI) space that includes affine transformation and nonlinear deformation done by DARTEL normalization. Later on, the segmented, normalized images are modulated through preserving amount which preserves total amounts of tissue corrected for individual differences in brain size, and spatially smoothed with an 8 mm full-width-half-maximum (FWHM) Gaussian kernel. All other parameters kept as default. Finally, a GLM is described for the segmented, DARTEL-warped, modulated, and smoothed images by using statistical analysis for GM and WM modalities individually. The voxel-wise two sample t-test is applied to the 3D GM and WM data sets separately. The design parameters are estimated and the inference on these estimated parameters are handled by using SPM12. This is done in order to find volume changes of each GM and WM data sets separately. The whole brain analysis is experimented with a threshold of uncorrected $p<.001$ and none extend threshold voxels. The framework is given in Fig. 1.

B. Feature Extraction

The generated 3D masks for GM and WM tissues are multiplied with the DARTEL normalized, modulated, and smoothed GM and WM tissues, respectively. Therefore, only the voxels clustered in the VOIs are extracted from the whole GM and WM tissue maps. The 3D masks of the GM modality is given in Fig. 2a and the WM modality is given in Fig. 2b. The dimensions of the extracted GM and WM data are reduced significantly through feature extraction. In this study, GM and WM VOIs are concatenated to investigate the effects of both modalities.

C. Feature Ranking

In order to alleviate the effect of the curse of dimensionality, improve the performance of the designed model, reduce the learning process time, and enhance data understanding, an FS method is required. In this paper, five filter feature ranking methods, Relief-F, Laplacian score (LS), unsupervised feature selection for multi-cluster data (MCFS), unsupervised discriminative feature selection (UDFS), and LLCFS, including both supervised and unsupervised ones are studied for detection of PD. The details of Relief-F, LS, MCFS, and UDFS feature selection methods are provided in [9]. In LLCFS, the goal is to obtain an appropriate data representation through FS or kernel learning within the framework of the LLC approach which can outperform the global learning-based ones when dealing with the high-dimensional data lying on manifold. A weight is assigned to each feature or kernel as well as it is added into the built-in regularization of the LLC algorithm to consider the relevance of each feature or kernel for the clustering [10].

D. Feature-Level Fusion and Feature Selection

After features are ranked by using five ranking methods, the order number of each ranked feature is assigned as a weight to that feature and then it is multiplied by its weight. For instance, the total number of extracted feature is said to be 350. The feature of 250 ordered to be the most important

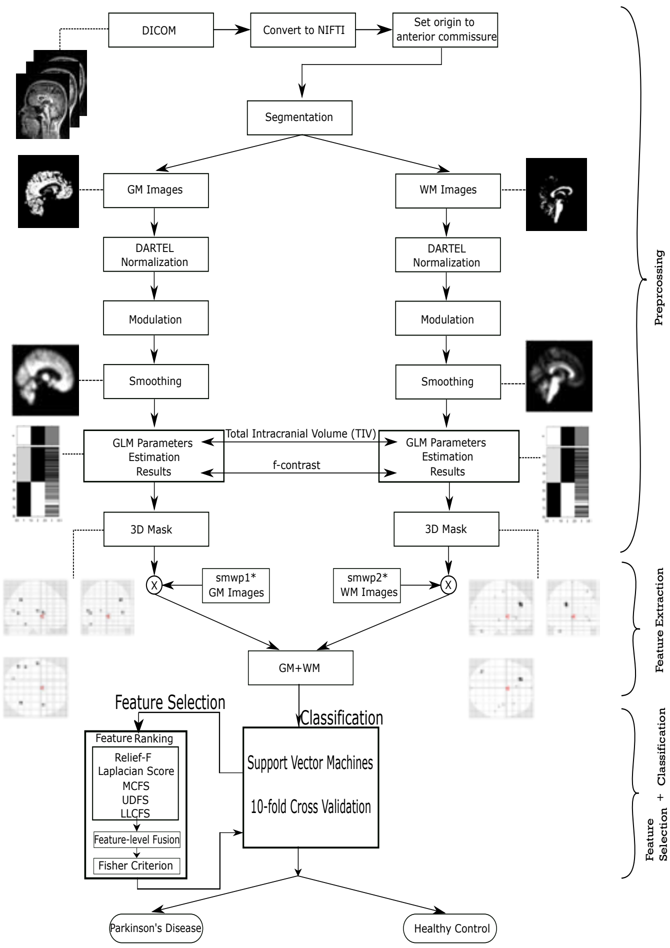

Figure 1: The framework of VBM plus DARTEL processing pipeline and classifying PD apart from HC.

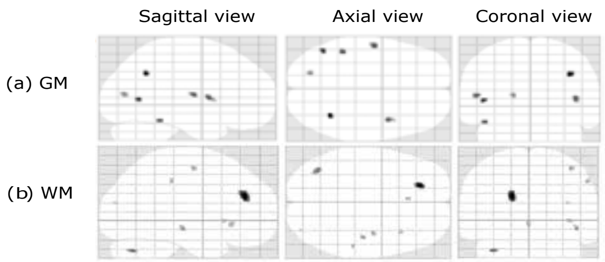

Figure 2: The 3D masks of GM and WM tissues with TIV as a covariate and f-contrast.

$6th, 16th, 32nd, 2nd, and 5th$ feature according to Relief-F, LS, MCFS, UDFS, and LLCFS method. Then the final value of feature 250 is $(6+16+32+2+5)/5$. Based on new values of the features, they are ranked again. After they are ranked, an FC technique is used to select the number of top-ranked attributes. The aim of FS is to find the feature subset of a certain size which causes the largest possible generalization or minimal risk [9]. The number of top-ranked features increases iteratively and the respective FC value is calculated for each iteration. The iterations are performed until a maximum FC value is obtained and the optimal number of top features is determined by taking the number of top-ranked features maximizing the FC.

E. The SVM Classifier

An SVM classification algorithm is used in order to classify PD patients apart from HCs. The main idea behind the SVM is to search for an optimal class-separation hyperplane in the maximal margin [11]. In this paper, the training is performed by using a classification learner application (https://www.mathworks.com/help/stats/classification-learner-app.html). An RBF-SVM classifier has been used. To evaluate the performance of the classifier, a procedure of two cross-validations (CVs) was combined with a grid search [8]. In this work, $K_1 = 10$ and $K_2 = 10$ were used.

IV. Experimental Results and Discussion

The experimental results obtained from 3D T1-weighted sMRI data by using SPM12 and CAT12 toolbox with DARTEL analysis. A GLM is configured by using two sample t-test on the SPM12 batch editor. The segmented, normalized, modulated and smoothed data are used in order to generate the 3D masks. The experiments are evaluated for the combination of GM and WM tissue volumes.

Table I: The SVM classification performance using five different feature ranking methods and an FC feature selecting approach for the combination of GM and WM datasets.

CM^*	FSM^*	ACC^*	SEN^*	SPE^*
SVM	Relief-F	86.25	82.50	90.00
	LS	83.75	77.50	90.00
	MCFS	83.75	80.00	87.50
	UDFS	86.25	82.50	90.00
	LLCFS	87.50	85.00	90.00
	Fusion All	**90.00**	**87.50**	**92.50**

* CM:classification method, FSM:feature selection method, ACC:accuracy(%), SEN:sensitivity(%), SPE:specificity(%).

The experimental results obtained from 3D T1-weighted sMRI data by using SPM12 and CAT12 toolbox with DARTEL analysis have been studied. The segmented, normalized, modulated and smoothed data are used in order to generate a 3D mask by using TIV as a covariate, f-contrast, and the combination of GM as well as WM tissue maps. The experimental results given in Table I indicate that among all five feature ranking methods, the LLCFS ranking with an FC stopping method has the highest classification performance, since it assigns a weight to each feature and incorporates it into the built-in regularization of the local learning-based clustering (LLC) algorithm to consider the relevance of each feature for the clustering. The Relief-f approach does not identify redundant attributes. The MCFS method performs especially well when the number of selected features is small [9]. However, in this paper, for some training folds,the number

Table II: Comparing classification performances of existing methods with applying these methods to 40 PD and 40 HC PPMI datasets used in this paper.

Research Work	Existing Methods			Applying Existing Methods		
	ACC^*	SEN^*	SPE^*	ACC^*	SEN^*	SPE^*
Salvatore et al. [3]	85.80	86.00	86.00	68.75	62.50	75.00
Babu et al. [1]	$67.44^\dagger$	$75.65^\dagger$	$58.00^\dagger$	62.50	57.00	68.00
Rana et al. [2]	88.33	90.00	86.67	78.75	85.00	72.50
Pahuja et al. [12]	$81.15^\dagger$	$90.34^\dagger$	$91.56^\dagger$	75.00	72.50	77.50
Proposed Method	-	-	-	**90.00**	**87.50**	**92.50**

* ACC:accuracy(%), SEN:sensitivity(%), SPE:specificity(%).

† SVM classification performances reported in [1] and [12].

of selected top-ranked features might be high. In this study, five different feature ranking methods following a feature-level fusion technique has been investigated. The feature-level fusion scheme which considers the ranked output of each ranking methods is used in order to improve the classification performance of PD detection. After the new ranked features are ordered based on their importance level, an adaptive FC stopping approach is take into account in order to select the optimal number of top-ranked features. The RBF-SVM classifier with 10-fold CV scheme is used for classification PD apart from HC. The classification accuracy of 90.00% given in Table I is achieved by using the proposed method.

In literature, the GM or WM tissue volumes of sMRI datasets are mostly used individually for the diagnosis of PD [1], [13], [14]. However, in this study, the performance of the proposed scheme is evaluated for the combination of the GM and WM tissue volumes. In order to compare the performance of the proposed method with the studies provided in Table II, the methods used in these studies are applied to the 40 PD and 40 HC data obtained from PPMI database and used in this paper. The detailed explanations of [1]–[3], [12], [13] are provided in [8], [9]. As seen from the Table II, the proposed method outperforms similar studies reported in the state-of-the-art by means of classification accuracies.

V. Conclusion

Detection and diagnosis of neurodegenerative diseases using the neuroimaging data have become an interesting research field, recently. In this study, classification of PD apart from HC has been analyzed by comparing the performances of five different feature ranking method, introducing a feature-level fusion followed with an adaptive Fisher criterion stopping method feature selection technique. The experiments are performed for the combination of GM and WM tissue volumes. Experimental results indicate that classification performance of fusion the ranking results of five feature ranking methods by using a majority voting technique enhances the classification performance of that using each feature ranking method individually. The highest classification accuracy of 90.00% is achieved when the proposed feature-fusion technique is taken into account.

References

[1] G. S. Babu, S. Suresh, and B. S. Mahanand, "A novel pbl-mcrbfn-rfe approach for identification of critical brain regions responsible for parkinson's disease," *Expert Systems with Applications*, vol. 41, no. 2, pp. 478 – 488, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413005605

[2] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, R. K. Agrawal, and M. Behari, "Voxel-based morphometry and minimum redundancy maximum relevance method for classification of parkinson's disease and controls from t1-weighted mri," in *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, ser. ICVGIP '16, 2016, pp. 22:1–22:6.

[3] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. Gilardi, and A. Quattrone, "Machine learning on brain mri data for differential diagnosis of parkinson's disease and progressive supranuclear palsy," *Journal of Neuroscience Methods*, vol. 222, pp. 230 – 237, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027013003993

[4] O. Cigdem, A. Yilmaz, I. Beheshti, and H. Demirel, "Comparing the performances of pdf and pca on parkinson's disease classification using structural mri images." 26th IEEE Signal Processing and Communication Applications (SIU) 2018, 5 2018.

[5] R. Schwartz, K. Rothermich, S. A. Kotz, and M. D. Pell, "Unaltered emotional experience in parkinson's disease: Pupillometry and behavioral evidence," *Journal of Clinical and Experimental Neuropsychology*, vol. 40, no. 3, pp. 303–316, 2018, pMID: 28669253. [Online]. Available: https://doi.org/10.1080/13803395.2017.1343802

[6] A. Pilotto, E. Premi, S. Paola Caminiti, L. Presotto, R. Turrone, A. Alberici, B. Paghera, B. Borroni, A. Padovani, and D. Perani, "Single-subject spm fdg-pet patterns predict risk of dementia progression in parkinson disease," *Neurology*, 2018.

[7] J. Pasquini, R. Ceravolo, Z. Qamhawi, J.-Y. Lee, G. Deuschl, D. J. Brooks, U. Bonuccelli, and N. Pavese, "Progression of tremor in early stages of parkinson's disease: a clinical and neuroimaging study," *Brain*, vol. 141, no. 3, pp. 811–821, 2018. [Online]. Available: +http://dx.doi.org/10.1093/brain/awx376

[8] O. Cigdem, I. Beheshti, and H. Demirel, "Effects of different covariates and contrasts on classification of parkinson's disease using structural mri," *Computers in Biology and Medicine*, vol. 99, pp. 173 – 181, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482518301136

[9] O. Cigdem and H. Demirel, "Performance analysis of different classification algorithms using different feature selection methods on parkinson's disease detection," *Journal of Neuroscience Methods*, vol. 309, pp. 81 – 90, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027018302528

[10] H. Zeng and Y. m. Cheung, "Feature selection and kernel learning for local learning-based clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1532–1547, Aug 2011.

[11] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[12] G. Pahuja and T. N. Nagabhushan, "A novel ga-elm approach for parkinson's disease detection using brain structural t1-weighted mri data," in *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*, Aug 2016, pp. 1–6.

[13] B. Rana, A. Juneja, M. Saxena, S. Gudwani, S. S. Kumaran, R. Agrawal, and M. Behari, "Regions-of-interest based automated diagnosis of parkinson's disease using t1-weighted mri," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4506 – 4516, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417415000858

[14] E. Adeli-Mosabbeb, C.-Y. Wee, L. An, F. Shi, and D. Shen, "Joint feature-sample selection and robust classification for parkinson's disease diagnosis," in *Medical Computer Vision: Algorithms for Big Data*. Springer International Publishing, 2016, pp. 127–136.