Exploitation of 3D Stereotactic Surface Projection for Automated Classification of Alzheimer's Disease according to Dementia Levels

Murat Seckin Ayhan[1]   Ryan G. Benton[2]
[1,2,3]Center for Advanced Computer Studies
University of Louisiana at Lafayette
Lafayette, LA, USA
{[1]msa4307, [2]rbenton, [3]vijay}@cacs.louisiana.edu

Vijay V. Raghavan[3]   Suresh Choubey[4]
[4]Applied Science Laboratory
GE Healthcare
Waukesha, WI, USA
suresh.choubey@med.ge.com

Abstract—**Alzheimer's disease (AD) is one major cause of dementia. Previous studies have indicated that the use of features derived from Positron Emission Tomography (PET) scans lead to more accurate and earlier diagnosis of AD, compared to the traditional approach used for determining dementia ratings, which uses a combination of clinical assessments such as memory tests. In this study, we compare Naïve Bayes (NB), a probabilistic learner, with variations of Support Vector Machines (SVMs), a geometric learner, for the automatic diagnosis of Alzheimer's disease. 3D Stereotactic Surface Projection (3D-SSP) is utilized to extract features from PET scans. At the most detailed level, the dimensionality of the feature space is very high, resulting in 15964 features. Since classifier performance can degrade in the presence of a high number of features, we evaluate the benefits of a correlation-based feature selection method to find a small number of highly relevant features.**

Keywords: Alzheimer's Disease, Stereotactic Surface Projection, Naïve Bayes, Support Vector Machine

I. INTRODUCTION

Dementia is an umbrella term used to refer to the deterioration of cognitive functions, such as memory loss, speech impairment, disorientation and poor judgment. AD is one major cause of dementia. One promising source of information for the early diagnosis of AD is PET scans, which captures the brain's current metabolic activity.

Kloppel et al. [7] compared the accuracy of dementia diagnosis provided by radiologists to that of computer-based diagnostic methods and concluded that the accuracy of computerized diagnosis is equal to or better than that of radiologists. Wen et al. [11] investigated the potential of parametric images derived from PET scans in order to discriminate among three categories; AD, Frontotemporal Dementia (FTD) and Normal. To cope with the high number of features, Principal Component Analysis (PCA) was adopted. Xia et al. [12] utilized genetic algorithm for the selection of those components. Minoshima et al. [8] pioneered the utility of 3D-SSP in AD diagnosis and extracted the metabolic activity scores based on the PET-scans. In [10], Sadeghi et al. demonstrated the utility of decision trees in distinguishing AD from FTD.

In our work, we mine the imagery data supplied by Alzheimer's Disease Neuroimaging Initiative (ADNI). More specifically, 3D-PET scans of human brains compose our data collection; the goal is to discern the class label of an instance in regards to Clinical Dementia Ratings (CDRs). We provide an empirical comparison of two popular learning algorithms with different characteristics; NB and SVM. Moreover, we investigate a correlation-based technique for the dimensionality reduction. We also report some preliminary results regarding the impact of using a concept hierarchy, which is created according to a standard taxonomy of brain's anatomical regions.

The next section describes the data characteristics and preprocessing stages. Section III presents the feature selection method used. Section IV depicts the experimental methodology and Section V presents the results, while Section VI provides the conclusions.

II. DATA AND PROCESSIING

Table I describes the demographics of our data collection, which is composed of 394 PET scans. The images covered a period between October 25, 2005 and August 16, 2007. The metabolic activity of the cerebral cortex is extracted with respect to the 3D-SSP using a GE proprietary application known as Cortex ID. As a result, an ordered list of 15964 predefined points is obtained (Fig. 1). Each point (*voxel*) is assigned a *z-score*, which measures how many standard deviations the metabolic activity departs from a predefined control group [10].

Table II shows the clinical dementia ratings (CDRs) and the corresponding levels of dementia. Fig. 2 depicts the histogram of the data set with respect to CDR. No instance of CDR-3.0 is present in our collection. Our goal in classification is to agree with expert decisions since pathological confirmations of the CDRs are not present.

TABLE I. DEMOGRAPHIC DATA OF SCANS

Avg. Birth Year	Gen.		Ethnicity		
	M	F	Hispanic/ Latino	Not Hispanic	Not Prov.
1930	134	69	4	193	6

TABLE II. LEVELS OF DEMENTIA (CLASS LABELS)

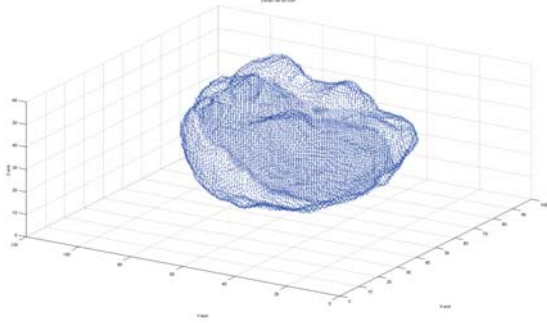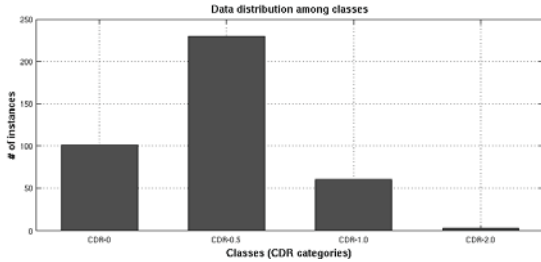| Classes	CDR Value	Degree of Disease
Class 1	0.0	Normal (Negative)
Class 2	0.5	Questionable
Class 3	1.0	Dementia
Class 4	2.0	Severe Dementia

Figure 1. Cortex extracted via 3D-SSP

Figure 2. Data distribution (101, 230, 60, 3, respectively)

A. Region Taxonomy

Voxels are grouped according to a taxonomy of anatomical regions, which conforms to the Talairach-Tourneau atlas (Fig. 3 and Table III). This scheme enables us to compute averages by region.
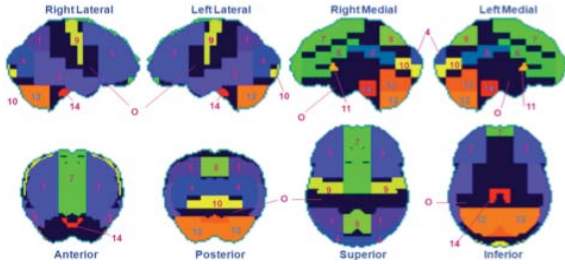

Figure 3. Taxonomy of cortical regions

TABLE III. REGION MAPPING TABLE

Region ID	Anatomic Region
0	Other
1	Parietal Association Cortex
2	Temporal Association Cortex
3	Frontal Association Cortex
4	Occipital Association Cortex
5	Post Cingulate Cortex
6	Anterior Cingulate Cortex
7	Medial Frontal Cortex
8	Medial Parietal Cortex
9	Primary Sensorimotor Cortex
10	Visual Cortex
11	Caudate Nucleus
12	Cerebellum
13	Vermis
14	Pons

B. Averaging Strategies

We apply a simple averaging filter to obtain the two higher-level representations of the voxel-level features. The number of features in higher-levels is dramatically smaller; the reduction is from 15964 to 30, representing the 15 regions on left and right hemispheres, and then to 15 (Fig. 4), due to the union of left and right hemispheres.
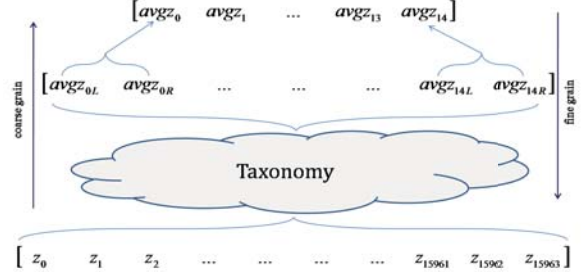

Figure 4. Feature hierarchy

III. CORRELATION-BASED FEATURE SELECTION

CFS [3] is a filter-based subset evaluator. It dramatically reduces the number of features. At the core of the algorithm is the function, given by (1), which measures the quality of the subset. The assumption is the relevant features are highly correlated with the class label, and the inter-correlation among these features is low.

$$Merit_s = \frac{kr_{cf}}{\sqrt{k + k(k-1)r_{ff}}}, \qquad (1)$$

where $Merit_s$ is the quality of the subset S with k features, r_{cf} the average feature-class correlation of the k-features, and r_{ff} the selected features' average feature-feature correlation.

IV. EXPERIMENTAL METHODOLOGY

In our benchmarks, we utilized NB and SVM on 3 different feature levels. The voxel-level feature vectors belong to the lowest level (Fig. 4). In addition, another type of feature vector is also generated via a fusion approach, such that descriptors from the 3 levels are concatenated into a single feature vector. This is done to determine if the feature selection method would choose aggregated scores over the voxel-level ones. Experiments have been conducted using WEKA [4]. NB conforms to [6]. WEKA also supports LibSVM [1,2], which is used to create the SVMs.

We conducted 3 different experiments. In each, Class 4 (CDR-2.0) is disregarded since the dataset has only 3 such instances. The first one considers a 2-class problem in which normal patients are put against the demented ones. The second experiment studies a 3-class problem, which includes the category of CDR-0.5. In third experiment, the union of two classes of abnormality is tested against normal.

V. EXPERIMENTAL RESULTS

This section presents the results of 5x2CV tests, which are 5 iterations of 2-fold cross-validation. Results for each experiment are partitioned into two groups depending on whether dimensionality reduction is applied. The tables illustrate which method performs significantly better

(denoted by +) or worse (denoted by ~) than the method in first column. Our performance metric is classification accuracy and the results are averages of 10 classification tasks. Standard deviations are next to the corresponding entry. The observed differences are tested at the 5% significance level with respect to a paired two-sided t-test.

A. Normal vs. Dementia

Table IV presents the baseline results for each learning scheme. The two SVM configurations are compared to NB. On first and second levels, classifier performances are generally equal; two exceptions are the RBF kernel SVM (Table IV) and linear kernel SVM (Table V), which perform, respectively, better and worse than NB at corresponding levels. Also notable is the classification performance for each classifier on the 3rd level is nearly identical to the classifier's performance on the Fusion levels. If aggregated features are put against voxel-level features for the purposes of dimensionality reduction, voxel-level features are generally the winners. Considering the fusion level vectors, the number of high-level features selected is negligible compared to the number of total features. Since we have a small data set with too many features, higher-level features may have been swamped out by the quantitative dominance of voxel-level features; however, to decide this issue we may need a larger data set. The last two rows of Table IV indicate that both SVMs induce better classifiers than NB on many features. In the literature, it has been indicated that linear SVMs are usually more robust compared to RBF SVMs when the number of features greatly outnumber the number of examples [5,9]; nevertheless, Tables IV and V show that both SVMs perform equally in the presence of many features. However, in Table IV, RBF kernel gains a significant lead (denoted by "^") at the upper levels, potentially due to the decreasing ratio of feature counts over number of examples.

TABLE IV. RESULTS WITHOUT FEATURE SELECTION

Levels	Naïve Bayes	SVM with RBF Kernel	SVM with Linear Kernel
1st	82.11 (4.43)	84.97 (4.55) +	79.99 (3.22) ^
2nd	84.60 (2.98)	86.08 (5.78)	80.25 (4.72) ^
3rd	79.01 (5.19)	93.05 (3.89) +	92.18 (4.54) +
Fusion	79.13 (5.06)	93.05 (3.89) +	92.18 (4.54) +

TABLE V. RESULTS WITH FEATURE SELECTION

Levels	Naïve Bayes	SVM with RBF Kernel	SVM with Linear Kernel
1st	81.48 (3.27)	82.60 (4.31)	80.86 (4.47)
2nd	83.98 (3.96)	82.98 (5.41)	80.98 (4.72) ~
3rd	89.44 (2.99)	91.92 (2.95) +	92.29 (3.52) +
Fusion	89.44 (2.99)	91.80 (3.06) +	92.17 (3.63) +

NB assumes, given a class, features are uncorrelated. CFS provides NB classifier with a refined subset that is qualified for the assumption of independence. At the end, it gives rise to 10% boost in classification performance (Table V). For SVMs, while CFS does not significantly change the classification performance, it is beneficial, given the dramatic reduction of dimensionality (Table VI) at no significant compromise of accuracy (Table V).

TABLE VI. EFFECT ON THE NUMBER OF FEATURES

Levels	Total # of features	# of selected features
1st	15	5.70 (1.16)
2nd	30	9.90 (1.91)
3rd	15964	137.20 (9.81)
Fusion	16009	136.00 (10.85)

B. Normal, Questionable and Dementia

This classification problem is the hardest of the three described as multiple decision boundaries are required to partition the feature space into multiple decision regions; this results in an increased computational complexity. In addition, the class of CDR-0.5 may present higher variance due to the non-uniform progression of AD. CDR-0.5 may be considered as a bridge between CDR-0 and CDR-1.0. Due to likely overlaps at two ends, the task of finding optimal decision surfaces is much harder. Comparing Table VII to Table IV, it is clear that overall performance is, expectedly, degraded. However, Table VII preserves the trend that RBF kernel SVM is the most accurate classifier throughout the hierarchy. It is also superior to linear kernel at 1st and 2nd levels. And NB is significantly worse than SVMs.

Once CFS is employed, Table VIII drifts from the Table V. Even though NB is augmented by CFS, none of the SVMs are favored. Hence, the CFS may not be selecting features that the SVMs consider useful. Table IX shows the average number of features selected at each level.

TABLE VII. RESULTS WITHOUT FEATURE SELECTION

Levels	Naïve Bayes	SVM with RBF Kernel	SVM with Linear Kernel
1st	59.54 (2.78)	65.52 (2.57) +	61.23 (2.54)
2nd	57.34 (2.15)	66.70 (3.10) +	61.02 (2.88) +
3rd	60.25 (2.34)	72.27 (2.83) +	71.87 (2.39) +
Fusion	60.25 (2.25)	72.22 (2.85) +	71.87 (2.39) +

TABLE VIII. RESULTS WITH FEATURE SELECTION

Levels	Naïve Bayes	SVM with RBF Kernel	SVM with Linear Kernel
1st	59.28 (2.44)	59.59 (1.93)	59.95 (1.10)
2nd	58.47 (2.10)	61.23 (3.09) +	59.44 (1.54)
3rd	65.06 (1.93)	66.75 (2.21) +	64.96 (2.86)
Fusion	64.96 (1.98)	66.24 (2.17)	65.37 (2.76)

TABLE IX. EFFECT ON THE NUMBER OF FEATURES

Levels	Total # of features	# of selected features
1st	15	5.00 (1.41)
2nd	30	8.90 (1.66)
3rd	15964	146.80 (11.37)
Fusion	16009	147.30 (11.17)

C. Normal vs. Abnormal

In this 2-class problem, two categories of abnormality (CDR 1 and 0.5) are combined with the intent to discern normal instances from abnormal ones. Since the problem is defined on two classes and it includes the category of CDR-0.5, its level of difficulty falls in between other two problems described earlier. Table X and Table XI show that overall classification performances, as expected, are better than the 3-class but worse than the normal versus dementia; yet, the general trend of SVMs dominating NB is preserved. SVMs, without feature selection, only differ at the 2nd level where the RBF kernel is significantly superior to linear kernel (Table X). CFS helps NB; however, once again, it hurts both SVMs. RBF kernel loses its advantage over linear SVM at the 2nd level; however, RBF kernel is more robust, when feature selection is employed, at lower levels (Table XI). Table XII shows the average number of features selected.

TABLE X. RESULTS WITHOUT FEATURE SELECTION

Levels	Naïve Bayes	SVM with RBF Kernel	SVM with Linear Kernel
1st	71.87 (2.73)	77.54 (3.72) +	76.83 (2.16) +
2nd	70.33 (2.68)	78.92 (3.71) +	75.76 (2.00) +
3rd	73.35 (4.37)	85.89 (2.85) +	85.32 (2.60) +
Fusion	73.35 (4.39)	85.89 (2.85) +	85.32 (2.60) +

TABLE XI. RESULTS WITH FEATURE SELECTION

Levels	Naïve Bayes	SVM with RBF Kernel	SVM with Linear Kernel
1st	72.68 (3.25)	74.73 (1.53)	74.68 (1.07)
2nd	72.17 (4.00)	75.91 (3.13)	75.40 (1.45) +
3rd	76.57 (3.27)	80.46 (3.16) +	77.81 (4.70)
Fusion	76.62 (3.03)	80.06 (3.10) +	77.75 (4.44)

TABLE XII. EFFECT ON THE NUMBER OF FEATURES

Levels	Total # of features	# of selected features
1st	15	5.60 (1.26)
2nd	30	9.10 (1.85)
3rd	15964	172.00 (17.66)
Fusion	16009	171.70 (17.41)

VI. CONCLUSION

This paper presents a comparison of a probabilistic learner, NB, and geometric learners, SVM, for the diagnostic purposes of AD. Earlier studies indicated that the use of features derived from PET scans lead to more accurate diagnosis of AD compared to the traditional approaches. In this study, 3D-SSP is utilized for feature extraction from PET scans; further, the benefits of utilizing a correlation-based feature selection method are assessed.

Results show that the feature selection procedure is beneficial for both NB and SVM in the sense that it dramatically reduces the dimensionality. We further show that the improvement in classification accuracy obtained for NB is statistically significant. However, the feature selection technique leads to slightly lower performance accuracy for SVMs' investigated; this may be due to the importance for SVM approach of retaining correlated features. Hence, we suggest that a SVM-oriented, rather than probabilistic, feature selection strategy be adopted.

Utilization of taxonomy information for creating aggregation levels poses the exploratory aspect of our study. While the higher level features smooth out the underlying levels' discriminative characteristics, we believe the idea of conceptual levels can be used to create a hierarchical feature selection method where the feature selection starts from the top and traverses the hierarchy in a top-down fashion.

REFERENCES

[1] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines", Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[2] Y. EL-Manzalawy, and V. Honavar, "WLSVM: Integrating LibSVM into Weka Environment" Software available at http://www.cs.iastate.edu/~yasser/wlsvm

[3] M. A. Hall, "Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning", Proceedings of the Seventeenth International Conference on Machine Learning, P. Langley, Ed. Morgan Kaufmann Publishers, San Francisco, CA, Jun. 29 – Jul. 02, 2000, pp. 359-366.

[4] M. Hall, E, Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, "The WEKA data mining software: an update", SIGKDD Explorations. Newsletter. Vol. 11, No. 1, ACM, New York, NY, USA, June 2009, pp. 10-18.

[5] C. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification", Technical Report (2009), Department of Computer Science National Taiwan University, http://www.csie.ntu.edu.tw/~cjlin

[6] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers", Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann, San Mateo, CA, USA, 1995, pp. 338-345.

[7] S. Kloppel, C. M. Stonnington, J. Barnes, F. Chen, et al. "Accuracy of dementia diagnosis - a direct comparison between radiologists and a computerized method", Brain: A Journal of Neurology, Vol. 131, No. 11, Oxford Journals, Oct. 2008, pp. 2969-2974.

[8] S. Minoshima, K. A. Frey, R. A. Koeppe, N. L.Foster, and D. E. Kuhl, "A Diagnostic Approach in Alzheimer's Disease Using Three-Dimensional Stereotactic Surface Projections of Fluorine-18-FDG PET", The Journal of Nuclear Medicine, Vol. 36, No. 7, July 1995, pp. 1238-1248.

[9] B. D. Ripley, Pattern Recognition and Neural Networks, 7th printing, Cambridge University Press, Cambridge, UK, 1996

[10] N. Sadeghi, N.L. Foster, A.Y. Wang, S. Minoshima, A.P. Lieberman and T. Tasdizen, "Automatic classification of Alzheimer's Disease vs. Frontotemporal dementia: A spatial decision tree approach with FDG-PET", 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, May 14 – 19, 2008, pp. 408 – 411

[11] L. Wen, M. Bewley, S. Eberl, M. Fulham and D. Feng, "Classification of dementia from FDG-PET parametric images using data mining", 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France, May 14 – 19, 2008, pp. 412 – 415.

[12] Y. Xia, L. Wen, S. Eberl, M. Fulham and D. Feng, "Genetic algorithm-based PCA eigenvector selection and weighting for automated identification of dementia using FDG-PET imaging", 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, 20-25 Aug. 2008, pp. 4812 – 4815.