# MultiCost: Multi-stage Cost-sensitive Classification of Alzheimer's Disease

Daoqiang Zhang[1,2] and Dinggang Shen[1]

[1] Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599
[2] Dept. of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
{zhangd,dgshen}@med.unc.edu

**Abstract.** Most traditional classification methods for Alzheimer's disease (AD) aim to obtain a high accuracy, or equivalently a low classification error rate, which implicitly assumes that the losses of all misclassifications are the same. However, in practical AD diagnosis, the losses of misclassifying healthy subjects and AD patients are usually very different. For example, it may be troublesome if a healthy subject is misclassified as AD, but it could result in a more serious consequence if an AD patient is misclassified as healthy subject. In this paper, we propose a multi-stage cost-sensitive approach for AD classification via multimodal imaging data and CSF biomarkers. Our approach contains three key components: (1) a cost-sensitive feature selection which can select more AD-related brain regions by using different costs for different misclassifications in the feature selection stage, (2) a multimodal data fusion which effectively fuses data from MRI, PET and CSF biomarkers based on multiple kernels combination, and (3) a cost-sensitive classifier construction which further reduces the overall misclassification loss through a threshold-moving strategy. Experimental results on ADNI dataset show that the proposed approach can significantly reduce the cost of misclassification and simultaneously improve the sensitivity, under the same or even higher classification accuracy compared with conventional methods.

**Keywords:** Cost-sensitive classification, cost-sensitive feature selection, MultiCost, multi-modality, Alzheimer's disease (AD).

## 1 Introduction

Alzheimer's disease (AD) is the most common form of dementia in elderly people worldwide. It is reported that the number of affected patients is expected to be doubled in the next 20 years [1]. Early diagnosis of AD and its prodromal stage, i.e., mild cognitive impairment (MCI), is very important for possible delay of the disease. At present, several biomarkers have been proved to be sensitive to AD, including brain atrophy measured by magnetic resonance imaging (MRI), hypometabolism measured by functional imaging (e.g., positron emission tomography (PET)), and quantification of specific proteins measured through cerebrospinal fluid (CSF) [2-3]. Over the past decades, a lot of AD classification methods have been developed based

K. Suzuki et al. (Eds.): MLMI 2011, LNCS 7009, pp. 344–351, 2011.

on one or two imaging modalities [4-6]. Recently, several studies have indicated that different biomarkers provide complementary information, which may be useful for diagnosis of AD when used together [2-3]. Motivated by this finding, some recent methods have been proposed to combine multimodal biomarkers to improve classification accuracy, which can generally do better than using only a single type of biomarkers [7-8].

Existing methods for AD classification aim to correctly assign subjects to one of several classes, typically two classes, i.e., AD patients and healthy subjects. Accordingly, most of the currently available algorithms for AD classification are designed to minimize the *zero-one loss* or *error rate*, i.e., the number of misclassified subjects. This implicitly assumes that all errors are equally costly, i.e., the loss of misclassifying an AD patient as healthy is the same as that of misclassifying a healthy subject as AD. However, this assumption rarely holds in real AD diagnosis. For example, it may be troublesome if a healthy subject is misclassified as AD, but it could result in a more serious consequence or even loss of life if an AD patient is misclassified as healthy. Apparently, the misclassification cost of AD is much higher than that of a healthy subject, and this important *a prior* knowledge should be used to guide AD classification. However, to the best of our knowledge, no previous studies explicitly consider different losses of misclassifying AD patients and healthy subjects in AD diagnosis.

On the other hand, cost-sensitive learning which can deal with the classification problem with unequal costs has been studied for years in the machine learning and data mining communities, and a lot of cost-sensitive learning algorithms have been proposed [9-10]. However, directly applying the existing cost-sensitive learning algorithms for AD classification may encounter several challenges. First, data in AD classification are usually available in multiple modalities, e.g., MRI, PET and CSF, etc., while most conventional cost-sensitive learning algorithms cannot be directly applied to the multimodal data. Second, because of the high dimensionality of neuroimaging data, the AD diagnosis method generally employs a feature selection stage before the classification stage; thus, it would be more helpful to use the cost information in both stages. However, the existing cost-sensitive learning algorithms are mainly used for classification rather than feature selection. To our best knowledge, this type of study on cost-sensitive feature selection was not done previously.

In this paper, we propose a **Multi**-stage **Cost**-sensitive approach (**MultiCost**) for AD (or MCI) classification. The MultiCost method contains three main components: (1) a cost-sensitive feature selection which can select more AD-related brain regions by using different costs of misclassification in the feature selection stage, (2) a multimodal data fusion which effectively fuses data from MRI, PET and CSF biomarkers based on multiple kernels combination, and (3) a cost-sensitive classifier construction which further reduces the overall misclassification loss through a threshold-moving strategy. Experimental results on ADNI dataset are presented to show the efficacy of the proposed approach.

The rest of this paper is organized as follows. In Section 2, we present the proposed MultiCost method for multimodal AD/MCI classification. Experimental results are given in Section 3. Finally, Section 4 concludes this paper and indicates points for future work.

## 2 MultiCost

In this section, we present a Multi-stage Cost-sensitive classification approach (MultiCost) for multimodal AD or MCI classification. We call it as MultiCost because the cost information is used in multiple stages, i.e., both feature selection and classification stages. There are three main steps in MultiCost, i.e., cost-sensitive feature selection on high-dimensional brain imaging data, multimodal data fusion from MRI, PET, and CSF biomarkers, and cost-sensitive classification.

### 2.1 Cost-sensitive Feature Selection

Because of the high dimensionality of brain imaging data, feature selection is usually required before classification. Like cost-sensitive classification, we can also explicitly exploit different costs of misclassifying diseased patients and healthy subjects for better feature selection, which is expected to select more 'good' features for predicting AD patients rather than healthy subjects. Here, we focus on binary classification, i.e., diseased ('D') or healthy ('H') class. Let $C_{DH}$ and $C_{HD}$ denote the cost of misclassifying a diseased patient (AD or MCI) as healthy and the cost of misclassifying a healthy subject as diseased, respectively. Without loss of generality, we assume $C_{DH} = C_{DH} / C_{HD}$ and $C_{HD} = 1$, as this will not change the final results. Typically, we require $C_{DH} > 1$ because the cost of misclassifying diseased patients as healthy is higher than that of misclassifying healthy subject as diseased.

In this section, we formally address the problem of cost-sensitive feature selection by explicitly incorporating different misclassification costs in the objective function of the feature selection algorithm. Specifically, we will extend a widely used filter-type feature selection algorithm, i.e., Variance Score (VS) [11], to the corresponding cost-sensitive version (CostVS). It is worth noting, however, similar idea can also be used to derive other cost-sensitive feature selection algorithms.

Let $f_{r,i}$ denote the $r$-th feature of the $i$-th example $\mathbf{x}_i$. Then, we can define the scoring function of CostVS as maximizing

$$CostVS_r = \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} cost(i,j)(f_{r,i} - f_{r,j})^2 \tag{1}$$

Where $cost(i,j)$ denotes the cost of misclassifying the $i$-th example as the $j$-th example. Because we use the class-dependent cost in this paper, those values can be easily gotten from $C_{DH}$ and $C_{HD}$. Specifically, we have

$$cost(i,j) = \begin{cases} C_{DH} & \text{if } y_i = \text{'D'} \text{ and } y_j = \text{'H'} \\ C_{HD} & \text{if } y_i = \text{'H'} \text{ and } y_j = \text{'D'} \\ 0 & \text{elsewhere} \end{cases} \tag{2}$$

Where $y_i$ and $y_j$ are the class labels ('D' or 'H') of the $i$-th and $j$-th examples, respectively. Intuitively, in Eq. 1, we want to increase the contribution from those examples with higher misclassification cost, i.e., diseased patients with AD or MCI, such that the features which are more related to AD or MCI diseases can be selected.

### 2.2 Multimodal Data Fusion

To effectively fuse data from MRI, PET and CSF modalities, we adopt a multiple-kernels combination scheme in this paper. Assume we have $m$ training examples, and each example has $M$ modalities of data (i.e., $M$=3 in this paper), represented as $\mathbf{x}_i=\{\mathbf{x}_i^{(1)},\ldots \mathbf{x}_i^{(p)},\ldots, \mathbf{x}_i^{(M)}\}$, $i$=1,…,$m$. Let $k^{(p)}(.,.)$ denote the Mercer kernel function on the $p$-th modality, then we can define the kernel function $k(.,.)$ on two data $\mathbf{x}$ and $\mathbf{z}$ as

$$k(\mathbf{x},\mathbf{z})=\sum_{p=1}^{M}\beta_p k^{(p)}(\mathbf{x}^{(p)},\mathbf{z}^{(p)}) \tag{3}$$

Where $\beta_p$s are the nonnegative weighting parameters used to balance the contributions of different modalities. All $\beta_p$s are constrained by $\sum_p \beta_p = 1$.

Once we have defined the kernel function $k(.,.)$ on multimodal data, the $m$ by $m$ kernel matrix $\mathbf{K}$ on the training multimodal data can be straightforwardly obtained as $\mathbf{K}=\{k(\mathbf{x}_i, \mathbf{x}_j)\}$. Then, the subsequent classifier can be directly built with the kernel matrix $\mathbf{K}$.

### 2.3 Cost-sensitive Classification

In this paper, we adopt support vector machine (SVM) implemented in LibSVM [12] with a threshold-moving strategy [9-10] as the cost-sensitive classifier, denoted as CostSVM in the rest of paper. Without loss of generality, we assume in this paper that the diseased patients are in the positive class (+1) while the healthy subjects are in the negative class (-1). Threshold-moving strategy moves the output threshold toward inexpensive class such that examples with higher costs become harder to be misclassified. Specifically, in the case of SVM classification, a positive constant is added to the threshold of SVM decision function such that the function value increases towards the positive class. It is worth noting that the threshold-moving is a post-processing strategy which introduces cost-sensitivity at test stage.

## 3 Experiments

In this section, we evaluate the effectiveness of the proposed cost-sensitive feature selection method (CostVS) and multimodal classification method (MultiCost) on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

### 3.1 Experimental Settings

The ADNI database (www.loni.ucla.edu/ADNI) contains approximately 200 cognitively normal elderly subjects to be followed for 3 years, 400 subjects with MCI to be followed for 3 years, and 200 subjects with early AD to be followed for 2 years. In this paper, we focus on multimodal classification of AD and MCI converters who convert to AD after some years. Accordingly, corresponding subjects with all MRI, PET, and CSF data at baseline are included. This yields a total of 146 subjects, including 51 AD patients, 43 MCI patients who had converted to AD within 18 months, and 52 healthy controls (HCs). Standard image pre-processing is performed

for all MRI and PET images, including anterior commissure (AC) - posterior commissure (PC) correction, skull-stripping, removal of cerebellum, and segmentation of structural MR images into three different tissues: grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). With atlas warping, we can partition each subject image into 93 regions of interests (ROIs) [13]. For each of the 93 ROIs, we compute the GM tissue volume from the subject's MRI image. For PET image, we first rigidly align it with its respective MRI image of the same subject, and then compute the average value of PET signals in each ROI. Therefore, for each subject, we can finally obtain totally 93 features from MRI image, other 93 features from PET image, and 3 features ($A\beta_{42}$, t-tau, and p-tau) from CSF biomarkers.

To evaluate the performances of different algorithms, we use 10-fold cross-validation strategy to compute *the total cost* (cost) in misclassifying subjects, *the classification accuracy* (ACC) for measuring the proportion of subjects correctly classified among the whole population, *the sensitivity* (SEN) for measuring the proportion of AD or MCI patients correctly classified, and *the specificity* (SPE) for measuring the proportion of healthy subjects correctly classified. To compute the total cost, we use the fixed costs of $C_{DH}$ =20 and $C_{HD}$ =1.

In our experiments, we compare four algorithms for multimodal classification, which are built based on a similar flowchart as MultiCost; and the differences lie in that they use different feature selection and classification methods:

- **VS_SVM** -- VS feature selection plus SVM;
- **VS_CostSVM** -- VS feature selection plus cost-sensitive SVM (CostSVM);
- **CostVS_SVM** -- CostVS feature selection plus SVM;
- **MultiCost** (CostVS_CostSVM) -- CostVS feature selection plus CostSVM.

All algorithms use a linear kernel to compute the kernels, and the values of the weighting parameters $\beta_m$s are gotten through cross-validation on training data using the grid search. For VS_CostSVM and MultiCost, the optimal values of the positive constant are tuned on the training data to achieve the highest sensitivity while maintaining similar accuracy as the corresponding cost-blind classifier.

### 3.2 Results

Fig. 1 shows the performances (cost, accuracy, and sensitivity) of the four algorithms under different number of features selected for AD classification. As can be seen from Fig. 1, compared with VS_SVM and VS_CostSVM, both MultiCost and CostVS_SVM greatly reduce the cost and at the same time improve the accuracy, as well as sensitivity and specificity, especially when a small number of selected features is used. Fig. 1 also indicates that by using the threshold-moving strategy, MultiCost further improves the sensitivity while keeping a similar accuracy as CostVS_SVM.

To make a further comparison among the four algorithms, we averaged the performances under different numbers of selected features (from 10 features to all features) and the results are given in Table 1. Here, we do not consider the cases of using less than 10 features, because all algorithms achieve bad performance. Moreover, we perform significance test between MultiCost (or CostVS_SVM) and other method on all performance measures, for both AD and MCI classifications. It can be seen from Table 1 that MultiCost always achieves significantly better results

than other methods on cost and sensitivity measures, while keeping accuracy similar to CostVS_SVM but better than VS_SVM and VS_CostSVM. On the other hand, Table 1 shows that both MultiCost and CostVS_SVM have much lower deviation than VS_SVM and VS_CostSVM, and thus are more robust to the variations on the number of selected features.

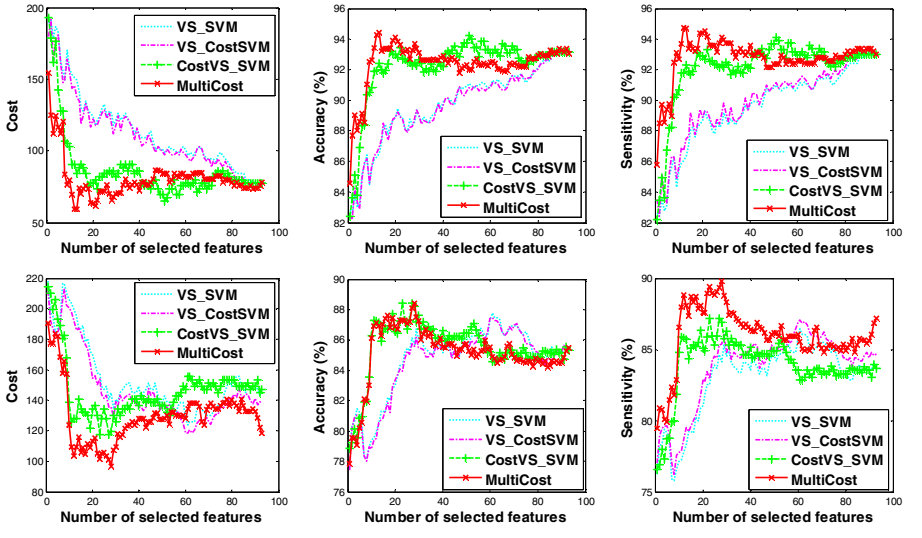


**Fig. 1.** Comparison of performance of four different methods on multimodal-data based AD (top) and MCI (bottom) classification

**Table 1.** Comparison of the averaged performance of different algorithms on multimodal data. The values in brackets are the standard deviation. The symbol * (or √) denotes that the difference between MultiCost (or CostVS_SVM) and other method is significant (at 95% significance level using paired t-test). The best value in each column is bolded.

| Methods | AD cost | ACC (%) | SEN (%) | SPE (%) | MCI cost | ACC (%) | SEN (%) | SPE (%) |
|---|---|---|---|---|---|---|---|---|
| VS_SVM | 107.2 (19.0) *√ | 90.4 (1.8) *√ | 90.2 (1.8) *√ | 90.5 (1.8) *√ | 148.5 (17.9) *√ | 85.2 (1.8) √ | 83.6 (2.0) *√ | 86.5 (1.6) *√ |
| VS_CostSVM | 104.3 (18.1) *√ | 90.3 (1.8) *√ | 90.5 (1.7) *√ | 90.0 (1.9) *√ | 142.2 (17.7) * | 85.0 (1.8) *√ | 84.4 (2.0) * | 85.6 (1.7) *√ |
| CostVS_SVM | 80.1 (6.4) * | **92.9 (0.6)** | 92.7 (0.6) * | **93.0 (0.6)** * | 140.9 (9.6) * | **86.0 (1.0)** | 84.5 (1.1) * | **87.3 (0.9)** * |
| MultiCost | **76.8 (6.3)** | 92.8 (0.6) | **93.1 (0.6)** | 92.5 (0.6) | **124.9 (11.7)** | 85.7 (1.1) | **86.5 (1.3)** | 84.9 (0.9) |

Fig. 1 and Table 1 show that VS_CostSVM achieves a similar performance as VS_SVM. Under the constraint of keeping a similar accuracy as VS_SVM, VS_CostSVM slightly improves the sensitivity at the price of a slight reduction in specificity, resulting in a slight improvement on the total cost. This implies that VS_CostSVM alone which incorporates the cost information only at the classification stage is not sufficient, due to the complexity and high-dimensionality of brain imaging data. In contrast, the MultiCost method exploits the cost information at both feature selection and classification stages, and thus significantly improves the performance.

Finally, we test the capability of the proposed CostVS feature selection method in selecting AD-related features, compared with standard VS method. In a typical experiment, we find, among its top 12 selected features, VS can only detect 4 AD-related features, computed from the ROIs such as 'supramarginal gyrus right', 'entorhinal cortex right', 'cingulate right', and 'temporal pole left', while CostVS can detect 7 AD-related features, computed from ROIs such as 'entorhinal cortex right', 'hippocampal formation right', 'amygdala right', 'parahippocampal gyrus right', 'parahippocampal gyrus left', 'hippocampal formation left', and 'amygdala left'. This shows that CostVS has more power than VS in detecting AD-related brain regions for guiding AD classification. Fig. 2 plots the top 12 features (brain regions) detected by CostVS.

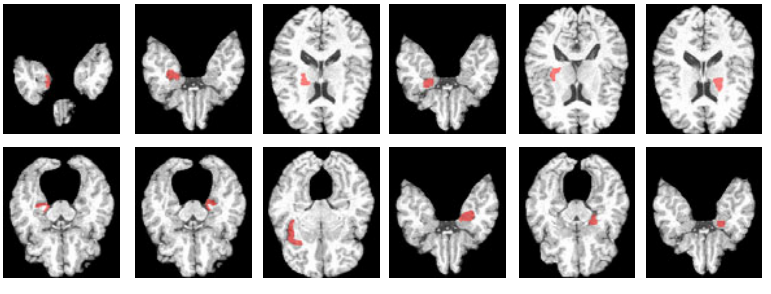

**Fig. 2.** Top 12 regions (highlighted) detected by CostVS. Each region is shown in a different view to enhance the visual quality.

## 4 Conclusion

This paper addresses the problem of exploiting the different misclassification cost in AD or MCI classification. Although using cost information for aiding classification is common in other areas such as machine learning and data mining, to our best knowledge, this issue is rarely investigated in AD-related studies. To effectively use the cost information for multimodal AD or MCI classification, in this paper we propose a Multi-stage Cost-sensitive classification method (MultiCost) to exploit cost information at both feature selection and classification stages. Experimental results on ADNI dataset validate the efficacy of the proposed method. In future work, we will extend our cost-sensitive feature selection to other feature selection methods and will also test other cost-sensitive classifiers for further improvement of classification performance.

**Acknowledgments.** This work was supported in part by NIH grants EB006733, EB008374, EB009634 and MH088520, and also by National Science Foundation of China under grant No. 60875030.

## References

1. Ron, B., Elizabeth, J., Kathryn, Z.G., Arrighi, H.M.: Forecasting the global burden of Alzheimer's disease. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 3, 186–191 (2007)
2. Walhovd, K.B., Fjell, A.M., Dale, A.M., McEvoy, L.K., Brewer, J., Karow, D.S., Salmon, D.P., Fennema-Notestine, C.: Multi-modal imaging predicts memory performance in normal aging and cognitive decline. Neurobiol. Aging 31, 1107–1121 (2010)
3. Vemuri, P., Wiste, H.J., Weigand, S.D., Shaw, L.M., Trojanowski, J.Q., Weiner, M.W., Knopman, D.S., Petersen, R.C., Jack Jr., C.R.: MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change. Neurology 73, 294–301 (2009)
4. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O.: Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database. Neuroimage (2010), doi:10.1016/j.neuroimage.2010.06.013
5. Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S.: Automatic classification of MR scans in Alzheimer's disease. Brain 131, 681–689 (2008)
6. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: COMPARE: Classification Of Morphological Patterns using Adaptive Regional Elements. IEEE Trans. Medical Imaging 26, 93–105 (2007)
7. Hinrichs, C., Singh, V., Xu, G., Johnson, S.: MKL for robust multi-modality AD classification. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 786–794. Springer, Heidelberg (2009)
8. Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N., Trojanowski, J.Q.: Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. Neurobiol. Aging (2010) (in press)
9. Elkan, C.: The foundations of cost-sensitive learning. In: The 17th International Joint Conference on Artificial Intelligence, pp. 973–978 (2001)
10. Zhou, Z.H., Liu, X.Y.: Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Trans. Knowledge and Data Engineering 18, 63–77 (2006)
11. Zhang, D., Chen, S., Zhou, Z.-H.: Constraint Score: A new filter method for feature selection with pairwise constraints. Pattern Recognition 41(5), 1440–1451 (2008)
12. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001)
13. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage 55(3), 856–867 (2011)