The Receiver Operational Characteristic for Binary Classification with Multiple Indices and Its Application to the Neuroimaging Study of Alzheimer's Disease

Xia Wu,
State Key Laboratory of Cognitive Neuroscience and Learning, College of Information Science and Technology, Beijing Normal University, Beijing 100875, P.R. China. wuxia@bnu.edu.cn

Juan Li,
State Key Laboratory of Cognitive Neuroscience and Learning, College of Information Science and Technology, Beijing Normal University, Beijing 100875, P.R. China. lijuan1109@126.com

Napatkamon Ayutyanont,
Banner Alzheimer's Institute (BAI) and Banner Good Samaritan PET Center, Phoenix, AZ, and Arizona Alzheimer's Consortium, Phoenix, AZ. Napatkamon.Ayutyanont@bannerhealth.com

Hillary Protas,
Banner Alzheimer's Institute (BAI) and Banner Good Samaritan PET Center, Phoenix, AZ, and Arizona Alzheimer's Consortium, Phoenix, AZ. Hillary.Protas@bannerhealth.com

William Jagust,
School of Public Health and Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA. jagust@berkeley.edu.

Adam Fleisher,
Banner Alzheimer's Institute (BAI) and Banner Good Samaritan PET Center, Phoenix, AZ, and Arizona Alzheimer's Consortium, Phoenix, AZ. Adam.Fleisher@bannerhealth.com

Eric Reiman,
Banner Alzheimer's Institute (BAI) and Banner Good Samaritan PET Center, Phoenix, AZ, and Arizona Alzheimer's Consortium, Phoenix, AZ. Eric.Reiman@bannerhealth.com

Li Yao, and
State Key Laboratory of Cognitive Neuroscience and Learning, College of Information Science and Technology, Beijing Normal University, Beijing 100875, P.R. China. yaoli@bnu.edu.cn

Kewei Chen
Banner Alzheimer's Institute (BAI) and Banner Good Samaritan PET Center, Phoenix, AZ, and Arizona Alzheimer's Consortium, Phoenix, AZ. Kewei.Chen@bannerhealth.com

for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Given a single index, the receiver operational characteristic (ROC) curve analysis is routinely utilized for characterizing performances in distinguishing two conditions/groups in terms of sensitivity and specificity. Given the availability of multiple data sources (referred to as multi-

indices), such as multimodal neuroimaging data sets, cognitive tests, and clinical ratings and genomic data in Alzheimer's disease (AD) studies, the single-index-based ROC underutilizes all available information. For a long time, a number of algorithmic/analytic approaches combining multiple indices have been widely used to simultaneously incorporate multiple sources. In this study, we propose an alternative for combining multiple indices using logical operations, such as "AND," "OR," and "at least n" (where n is an integer), to construct multivariate ROC (multiV-ROC) and characterize the sensitivity and specificity statistically associated with the use of multiple indices. With and without the "leave-one-out" cross-validation, we used two data sets from AD studies to showcase the potentially increased sensitivity/specificity of the multiV-ROC in comparison to the single-index ROC and linear discriminant analysis (an analytic way of combining multi-indices). We conclude that, for the data sets we investigated, the proposed multiV-ROC approach is capable of providing a natural and practical alternative with improved classification accuracy as compared to univariate ROC and linear discriminant analysis.

Keywords

Alzheimer's dementia (AD); multiple indices; multiV-ROC; receiver operational characteristic (ROC)

1 Introduction

Cʟᴀꜱꜱɪꜰɪᴄᴀᴛɪᴏɴ of healthy and unhealthy individuals for a given disease or condition is a common practice. In clinical diagnosis, possible treatment effectiveness evaluation, and other settings, such classification can also be challenging. Without exception, classification is particularly challenging in developing biomarkers for studies of neurodegenerative diseases, such as Alzheimer's disease (AD). In routine practice, such classification can be carried out either using measures from a single index or, as is becoming more frequent every day, based on measurements from multiple indices. With advanced data acquisition/collection techniques, it is no longer unusual to have data from multiple sources (each source is associated with one index) available to medical professionals. For example, a number of biomarkers, including those based on neuroimaging techniques, body fluid, or neuropsychiatric (NP) measures, have been carefully examined and proposed for their potential uses in the study of AD. The widely used NP measures for AD include the AD assessment scale-cognitive subscale (ADAS-Cog) [1], auditory verbal learning test (AVLT) total, and long-term memory (LTM) scores [2], CDR global and sum of boxes (SB) [3], and mini-mental state exam (MMSE) [4]. Among the neuroimaging techniques used in the study of AD are volumetric magnetic resonance imaging (MRI), which can be used to measure the volume of a specific brain region (e.g., hippocampus) [5]; regional gray matter volume or cortical thinning [6]; functional MRI [7], [8], which can be used to examine the default mode network changes under resting state; fluorine-18 deoxyglucose (FDG) positron emission tomography (PET) [9], which can be used to evaluate glucose hypometabolism; and the Carbon-11 Pittsburgh Compound-B (PiB)-PET, florbetapir-PET and other amyloid PET techniques, which can be used to quantify the beta-amyloid deposition [10], [11].

In the context of the clinical diagnosis of AD, the receiver operational characteristic (ROC) curve analysis [12], [13] is routinely utilized to characterize the performance of a given index in distinguishing AD patients from normal healthy controls (or from another disease) in terms of sensitivity and specificity. The sensitivity and specificity analysis of a routine ROC analysis, however, is only applicable for a single index (e.g., measurements from only one region of interest (ROI) for one neuroimaging data set or scores of one NP test).

Given the availability of multiple indices, such as the large amounts of information for 3D neuroimaging data sets, the ROC analysis of only a single index is clearly inadequate and an inefficient use or underutilization of available information. To take advantage of the availability of multidata sources, Xiong and Ye applied them to combine multiple diagnostic tests to distinguish AD from normal controls [14], [15], [16]. In fact, it is natural for people to reach conclusions, be them medical or otherwise, based on multiple features. Similarly, for information-rich neuroimaging data, observations from multiple brain locations (i.e., from multi-ROIs) over the images are simultaneously taken into consideration. This common practice and the previous studies such as [14], [15], [16] raise the need to categorize subjects in a binary manner, mainly healthy or unhealthy, using multiple indices or integrating information from multiple indices for better classification.

For a long time, a number of algorithmic/analytic approaches combining multiple indices have been proposed, introduced and widely used to incorporate the simultaneous use of multiple sources. Among these approaches are logistic regression (including the relatively recently introduced sparse logistic regression [17], [18], linear classification/discriminant analysis [19], [20] (such as Fisher classification [21]), the support vector machine (SVM) [22] and others. For example, the linear Fisher discriminant estimates the parameters of a linear classifier that combines multiple variables into a single expression based on the assumptions of Gaussian conditional density models [21]. We categorize all these approaches as algorithmic or analytic, as they use arithmetic, mathematical, or analytic expressions to combine multiple indices into a single index. Alternatively, to the analytic/arithmetic approaches, one may also consider combining multiple indices in a logical manner to construct multivariate ROC (multiV-ROC) and to characterize the sensitivity and specificity statistically associated with the use of multiple indices. This logical combination is the focus of this study.

It is worth noting that, in comparison to the algorithmic approaches, such logical ways of combining multiple features are instinctual and intuitive to day-to-day human cognition. For the neuroimaging AD studies, the simultaneous use of information from multiple brain regions (multi-indices) to increase statistical power has been investigated and documented extensively and is, therefore, not new [11], [23]. The information integration from multiple ROIs or from voxels over the whole brain volume is most commonly executed using multivariate approaches, such as partial least square (PLS) [24], principal component analysis (PCA)-based scaled subprofile modeling (SSM) [25], independent component analysis (ICA) [7], [26], and others. The outcomes of these multivariate approaches are the so-called subject scores that summarize arithmetically the measurements from multiple ROI/voxels (usually in the form of weighted sum, the linear combination). Therefore, the end product of any multivariate approach is a single index that can be fed into ROC analysis or

statistical power analysis in the same way as any other single univariate index. Alternatively, for this arithmetic algorithm to deal with multiple variables (measurements from multiple ROI), we propose a logical operation for incorporating these multiple variables by treating each variable as a random event. Therefore, the logical operations, such as union and intersection, are applicable. In conjunction with this newly proposed approach, there is the need to introduce the concept of multivariable ROC [27], referred to as multiV-ROC below.

The examination of multiple variables/indices in making a clinical decision is actually a very common practice in our daily life and in diagnostic settings. For example, in assessing the amyloid accumulation in AV-45 PET studies for Alzheimer' dementia, six cortical ROIs are investigated visually or quantitatively [11]. The AV-45 amyloid positivity is defined as "any one of these 6 is positive," the logical OR (union). Equivalently, the combination of information from these six cortical ROIs is logical AND (intersection) for amyloid negativity. Should it be medically meaningful, one can clearly logically define positivity as "at least J measures are positive," where J is an integer less than the total number of indices (e.g., number of ROIs for imaging data). Clearly, given a logical way of combining multiple measures, the settings related to the optimal sensitivity and specificity need to be characterized, and their performances in comparison to each index separately or possible other manners of combination need to be examined. For this purpose, we propose a computational procedure to calculate ROC and its areas under curve (AUC) for this logical way of combining of multi-indices.

There are numerous methodological investigations, discussions, and introductions on the use of the ROC for single-index measurement found in literature, on the internet, and in common statistical text books. In fact, these approaches are almost common knowledge among people in the statistics community. Conversely, multiV-ROC, excluding a single instance, has not been thoroughly studied or reported on [27]. In this study, we will briefly introduce the numerical procedures for determining the optimal threshold combination and estimating its corresponding sensitivity and specificity. Then, two data sets are used to illustrate the potential added value of this approach and its performance compared to linear discriminant analysis. We will also briefly discuss some challenging questions associated with this approach but will not attempt to address them in this study.

2 Methods

2.1 Data Set 1

This data set was from the AD neuroimaging initiative (ADNI) project. (The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and nonprofit organizations as a $60 million, 5-year public-private partnership. The primary goal of ADNI is to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. The determination of sensitive and specific markers for very early AD progression is intended to aid researchers and clinicians in developing new treatments and monitoring their effectiveness and lessening the time and cost of clinical trials. The principal investigator of

this initiative is Michael W. Weiner, MD, from the VA Medical Center and University of California, San Francisco. ADNI is the result of the efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 adults, aged 55 to 90 years, to participate in the study (approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years). For up-to-date information, see www.adni-info.org. For 74 AD patients who had both baseline and 12-month measurements from an MMSE, we investigated the extent to which the baseline measurements could be used to distinguish AD patients who demonstrated subsequent 12-month MMSE declines of at least three points from those who did not. The 74 subjects included 44 nondecliners and 30 decliners. Three baseline indices: the FDG PET-based hypometabolic convergence index (HCI), the total-score of AVLT, and scores from AD ADAS-cog, were each examined for their differentiating power, separately or combined. HCI was intended to characterize the extent to which the magnitude and pattern of cerebral glucose hypometabolism in a person's FDG PET image corresponded to that in a group of AD patients [28].

2.2 Data Set 2

This data set consisted of resting state functional magnetic resonance imaging (fMRI) measurements extracted from multiple ROI in our previous study with 15 AD patients and 16 normal controls [26]. These ROI data are the spatial extent (ROI volume) and averaged voxel intensities of the core regions in three resting state networks: the default-mode network (DMN), dorsal attention network (DAN), and ventral attention network (VAN). These core regions are the posterior cingulate cortex (PCC), medial prefrontal cortex, left inferior parietal cortex, right inferior parietal cortex, left inferior temporal cortex (LITC) and right inferior temporal cortex (RITC) from DMN, left superior/inferior parietal lobule, right superior/inferior parietal lobule, left middle/inferior frontal gyrus, right middle frontal gyrus from DAN, temporal-parietal junction (TPJ), and ventral frontal cortex from the VAN. We originally examined the abnormalities of attention-related functional networks in AD and evaluated the sensitivity and specificity of these networks as potential biomarkers compared to the DMN. Univariate ROC curve analysis was performed for activity in core regions within each of these networks. The results of these analyses suggested that activity in the left intraparietal sulcus and left frontal eye field (LFEF) from DAN as well as the posterior cingulate cortex from DMN could serve as sensitive and specific biomarkers, distinguishing AD from NC separately as a single index. To focus our discussion more on the methodology and simplify the question discussed, data from only four ROIs were included in this study: the averaged voxel intensities of PCC, LFEF, LITC, and TPJ.

2.3 Constructing a Multivariable ROC Curve

As in the univariate ROC case, constructing a multivariable ROC curve is numerical. There are n continuous indices, where a high value of each is indicative of unhealthiness (otherwise negate the index).

2.4 The Procedure to Determine the Optimal Threshold Combination

Define case positivity (unhealthiness). For example, a case is positive if any one of the n indices is above its threshold/cut-off value. This is the operation of logic, OR (at least one is positive). Other operations of logic, such as the intersection, "at least k," and logical precedent, can also be used to define the case positivity. Note that, for a real-world problem, the value of k to determine positivity is usually unknown for a new application. The determination of k can be obtained either numerically with an exhaustive search, as is the case with the optimization procedure, or with the use of the clinical knowledge (a priori) from medical professionals (e.g., if one brain region is found to have amyloid, then the brain has amyloid, [$k = 1$ in this case]) or a combination of both. Different from univariate ROC, we need to distinguish case positivity from index positivity. For multiV-ROIC, an index can be positive, but the case may not be (e.g., if the logical operation is ALL indices must be positive as defined by the logical intersection operation). Note that for each index, the positivity threshold is not defined independently. Instead, it is defined in the context of determining the optimal combined thresholds for all indices together given a logical operation.

Divide the measurements for index i into k_i subintervals, and take the $k_i - 1$ end points of the subintervals as cut-off values for index i. For simplicity, we let the subintervals have equal length and $k_i = k$ be common for all indices $i = 1, 2, \ldots, n$). When $k = 3$, for example, there are two cut-off points for each index. The total possible combinations is $(k - 1)^n$.

For each of ALL, the possible cut-off value combinations ($(k - 1)^n$ grows very rapidly with k, especially as n increases) and each of the randomly selected N combinations (e.g., $N = 50,000$), calculate the corresponding sensitivity and specificity. Save the setting and corresponding sensitivity and specificity.

Eliminate any possible duplicates and form the ROC curve (which requires only the knowledge of the sensitivity and specificity). The AUC is computed using the Trapezoidal numerical integration method.

Find the best sensitivity/specificity as the point on the ROC curve closest to (0 1) and identify the corresponding cut-off combination.

2.5 Estimation of the Sensitivity and Specificity Using Jackknife Leave-One-Out Procedure

The use of the overall data set to determine the threshold combination and to calculate the corresponding sensitivity and specificity will be informative in comparing multiVROC and univariate ROC. For objective assessment, however, the data used for the determination of the threshold combination should not be used for the evaluation of the performance. Instead, the performance of the determined threshold should be examined using an independent data set. We adopted the leave-one-out strategy to investigate the performance of the proposed multiV-ROC in comparison to the univariate ROC for each variable, which uses a single observation from the original sample as the validation data and the remaining observations as the training data. The training data set was used to construct the ROC curve and to determine the cut-offs combinations. The validation data is used to objectively evaluate sensitivity and specificity.

2.6 Comparison of MultiV-ROC with Linear Discriminant Analysis

We compared the newly introduced logical combination of multiple variables to one of commonly used analytic combination approach: linear Fisher discriminant analysis. It is implemented in SPSS 16.0 (SPSS, Inc., Chicago, IL). This comparison is in terms of sensitivity/specificity and accuracy.

3 Results

3.1 Data Set 1

For the data set extracted from the ADNI AD decliner versus AD stable subjects, the performance of using baseline measurements for each variable separately to predict subsequent 12-month decline (classifying AD patients who declined from those who did not) is shown in Tables 1 and 2 in comparison to that of using multi-ROC approach. Comparing multiV-ROC with single-index ROC (Tables 1 and 2, and Fig. 1), we note a 9 percent increase of the accuracy in differentiating AD patients who experienced a decline from those who did not. In addition to these overall accuracy increases, more balanced sensitivity/specificity was observed when HCI alone or multiV-ROC was used. Finally, the multiV-ROC approach shows accuracy improvement compared to the linear Fisher discrimination analysis, logistic regression and support vector machine (see supplemental material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2012.141). The average computation time is 56.7 seconds, and 4,212.3 seconds when leave-oneout cross-validation was applied. See online supplementary material for more detailed result presentations.

3.2 Data Set 2

For the resting state fMRI data, including the averaged voxel intensities of the PCC, LFEF, LITC, and TPJ from the three resting-state networks, DMN, DAN, and VAN, the performance of each separate biomarker in distinguishing the AD from NC is shown in Table 3. Table 4 lists the performance of the multiV-ROC using the logical operation of "at least 2." Comparing these two tables, we observed about a 10 to 20 percent increase of the accuracy of distinguishing AD from NC (Tables 3 and 4, and Fig. 2). Table 4 also shows the classifying results of linear Fisher discrimination, which has both a specificity and sensitivity lower than those from the multiV-ROC analysis. The average computation time is 43.2 seconds and 1,286.4 seconds when leave-one-out cross-validation was applied.

4 Discussion

The use of multiple variables could potentially result in increased sensitivity and specificity when compared to the use of a single variable or the specific analytic combination technique, linear discriminant analysis. This increase was clearly demonstrated in our current study.

However, it is worth noting that there are situations where the proposed multiV-ROC may not provide additional benefits. One can think of instances where the use of multiV-ROC approach will not result in increased accuracy. For example, if the first variable is strongly

correlated with the second, then one should not expect any added benefit with the use of the two simultaneously. In general, however, multiV-ROC should be as good as using any of the variables alone. However, there is clearly an increased computational cost associated with the use of more than one variable. In addition, considering all possible cut-off value combinations is very time-consuming when the index is greater than 3. To address this, we randomly selected a subset of all possible combinations. Even with the possibility that the, thus, determined cut-off values could be suboptimal, the results were satisfactory and similar to the results achieved from running all combinations.

The approach that we proposed of combining multiple variables seems natural, considering human reasoning in evaluating multiple sources of evidence. We still do not have a good understanding of how the human brain combines these multiple sources of evidences together, but it may not be in any of the analytic manners. Using dual variables as an example, a natural decision-making process might be more of a logical combination.

In comparison to the specific algebraic combination, linear discriminant analysis, our results based on the two data sets support the claims that the logical combination carries some practical value in terms of increased accuracy. With ever-rapidly growing computing power, the difficulty associated with the establishments of the optimal threshold combination is no longer a concern. Moreover, we note that the applications of this approach with the established threshold combination are as simple and straightforward as the traditional single variable-based classification (and possess no further associated cost in terms of computation time). Therefore, we do not see it as computationally burdensome.

Our results demonstrated that combining information from multiple ROIs can improve the ability of differentiating AD patients under normal control or AD patients who declined from those who did not. For data set 2, several studies suggested these ROIs play different roles in the resting state networks and that the changed pattern of each ROI may represent different aspects of the disease. By combining these together, the disease is better characterized and yields improved classification results.

For its use in data set 1, we note that HCI was first introduced as a single variable based on information-rich 3D FDG-PET data, its superior performance in assessing the risk of conversion to AD among MCI patients was reported in our earlier study [24]. In assessing the possible disease prognosis (in terms of staying stable or not over the next 12-month period) as a single variable, its overall accuracy seemed inferior to the ADAS-Cog (0.72 versus 0.76 in AUC measures). With more careful examination, however, one finds that ADA-cog possessed very unbalanced sensitivity and specificity (87 percent versus 54 percent in the overall test or 80 percent versus 50 percent in the leave-one-out test). The HCI, on the other hand, provided overall a relatively balanced sensitivity and specificity.

Our results demonstrated the performance of the proposed multiV-ROC approach using an overall data set. We also evaluated its performance again in comparison to single-variable ROC and linear Fisher discriminant analysis using the Jackknife leave-one-out cross validation procedure. The same increase in accuracy was also observed with the use of the procedure for both data set 1 and data set 2. We also studied the robustness of the multiV-

ROC approach with respect to the number of subinterval k_i and combination N. We showed that the outcomes of numerical iterative procedure, the optimal threshold combination were very robust when these two parameters are greater than a certain threshold ($k_i = 200$, $N = 50,000$), while the lower values of these two parameters might slightly affect the accuracy of our method. Detailed results of the cross-validation [30] and robustness were not included in this report because of concerns about limitations on the article length. See online supplementary material for more detailed result presentations.

In this preliminary study, we only proposed and introduced the approach and numerically validated its application values. Although not included in this study, some in-depth theoretical discussions about this approach are nevertheless necessary. Among the issues that require better understandings are the monotonicity of the ROC curve, the statistical inference to examine AUC (providing for the type I error assessment) (note that we did not use conventional AUC statistical test in our study), parametric or nonparametric ways to assess whether one threshold combination is statistically significantly better than another. Another theoretical issue we did not discuss is how the covariance structure among the multiple indices will influence the performance. No attempt was made in this study to evaluate the degree of such influence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Biography

Xia Wu received the PhD degree from Beijing Normal University (BNU). She is an associate professor at the College of Information Science and Technology of BNU, China. Her main research interests include intelligent signal processing, especially neuroimaging data processing.

Juan Li received the master degree from Beijing Normal University. Her main research interests include intelligent signal processing, especially neuroimaging data processing.

Napatkamon Ayutyanont received the PhD degree from Arizona State University. She is a biomathematician at Banner Alzheimer's Institute's computational image analysis program. Her research focuses on the study of the risk modification, early detection, diagnosis, tracking, and treatment evaluation of Alzheimer's disease and normal aging in older adults using statistical, mathematical, and data mining methodologies on cognitive ratings, neuroimaging and other biomarker data.

Hillary Protas received the PhD degree in biomathematics at the University of California, Los Angeles. She is a postdoctoral fellow in computational image analysis at Banner Alzheimer's Institute. Her main interests include the images analysis of the Alzheimer's disease.

William Jagust received the MD degree from Stony Brook University, New York, following which he was a resident in neurology at Boston University and a postdoctoral

fellow at Lawrence Berkeley National Laboratory. He is a professor of public health and neuroscience at the University of California, Berkeley, and faculty senior scientist at Lawrence Berkeley National Laboratory. His research is focused on the use of positron emission tomography and magnetic resonance imaging in understanding brain aging and dementia.

Adam Fleisher received the medical degree from the University of Rochester School of Medicine, New York, and the general neurology training at The Johns Hopkins Hospital in Baltimore, Maryland. He then completed a clinical and research dementia fellowship at the University of California, San Diego (UCSD), as well as a master's degree of advanced studies in clinical research. He is currently serving as the director of brain imaging at Banner Alzheimer's Institute. He is an associate professor in the Department of Neurosciences at UCSD, where he is the medical director of the Alzheimer's Disease Cooperative Study. He is an expert in the field of imaging for studying the earliest evidence of Alzheimer's pathology in the brain.

Eric Reiman is an executive director of the Banner Alzheimer's Institute, chief executive officer for Banner Research, clinical director of the Neurogenomics Division at the Translational Genomics Research Institute, professor of psychiatry at the University of Arizona, and the director of the Arizona Alzheimer's Consortium. His research interests include brain imaging, genomics, and their use in the unusually early detection and tracking of Alzheimer's disease, the evaluation of genetic and nongenetic risk factors, and the accelerated evaluation of treatments to prevent Alzheimer's disease.

Li Yao received the PhD degree from the Chinese Academy of Science. She is a professor and the vice dean at the College of Information Science and Technology of Beijing Normal University, China. Her main research interests include brain signal processing.

Kewei Chen received the master's degree in mathematics from Beijing Normal University, China, and the PhD degree in biomathematics in the area of neuroimaging at University of California, Los Angeles. He is a senior scientist and senior biomathematician at Banner Alzheimer's Institute. His primary research interests include the quantification and statistical analysis of positron emission tomography and magnetic resonance imaging data in the study of human brain functions and diseases, particularly Alzheimer's disease.

References

[1]. Rosen WG, Mohs RC, Davis KL. A New Rating Scale for Alzheimer's Disease. Am J. Psychiatry. 1984; 141:1356–1364. [PubMed: 6496779]

[2]. Rey A. L'examen Psychologique Dans Les Cas D'encephalopathie Traumatique. Archiv Psychologie. 1941; 28:286–340.

[3]. Morris JC. The Clinical Dementia Rating (CDR): Current Version and Scoring Rules. Neurology. 1993; 43:2412–2414. [PubMed: 8232972]

[4]. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State a Practical Method for Grading the Cognitive State of Patients for the Clinician. J. Psychiatric Research. 1975; 12(no. 3):189–198.

[5]. Karow DS, McEvoy LK, Fennema-Notestine C, Hagler DJ, Jennings RG, Brewer JB, Hoh CK, Dale AM. Relative Capability of MR Imaging and FDG PET to Depict Changes Associated with Prodromal and Early Alzheimer Disease. Radiology. 2010; 256:932–942. [PubMed: 20720076]

[6]. Whitwell JL, Shiung MM, Przybelski SA, Weigand SD, Knopman DS, Boeve BF, Petersen RC, Jack CR. MRI Patterns of Atrophy Associated with Progression to AD in Amnestic Mild Cognitive Impairment. Neurology. 2008; 70:512–520. [PubMed: 17898323]

[7]. Greicius MD, et al. Default-Mode Network Activity Distinguishes Alzheimer's Disease from Healthy Aging: Evidence from Functional MRI. Proc. Nat'l Academy of Sciences of the United States of USA. 2004; 101:4637–4642.

[8]. Wu X, et al. Altered Default Mode Network Connectivity in Alzheimer's Disease—A Resting Functional MRI and Bayesian Network Study. Human Brain Mapping. 2011; 32:1868–1881. [PubMed: 21259382]

[9]. Herholz K, et al. Discrimination between Alzheimer Dementia and Controls by Automated Analysis of Multicenter FDG PET. NeuroImage. 2002; 17:302–316. [PubMed: 12482085]

[10]. Thal DR, et al. Phases of a Beta-Deposition in the Human Brain and Its Relevance for the Development of AD. Neurology. 2002; 58:1791–1800. [PubMed: 12084879]

[11]. Fleisher AS, et al. Using Positron Emission Tomography and Florbetapir F 18 to Image Cortical Amyloid in Patients with Mild Cognitive Impairment or Dementia Due to Alzheimer Disease. Archives of Neurology. 2011; 68:1404–1411. [PubMed: 21747008]

[12]. Goodenough D, Rossman K, Lusted L. Radiographic Applications of Receiver Operating Characteristic (ROC) Curves. Radiology. 1974; 110:89–95. [PubMed: 4808546]

[13]. Swets J. ROC Curve Analysis Applied to the Evaluation of Medical Imaging Techniques. Invest Radiology. 1979; 14:109–121.

[14]. Xiong C, et al. Combining Correlated Diagnostic Tests: Application to Neuropathologic Diagnosis of Alzheimer's Disease. Medical Decision Making. 2004; 24(no. 6):659–669. [PubMed: 15534346]

[15]. Gao F, et al. Estimating Optimum Linear Combination of Multiple Correlated Diagnostic Tests at a Fixed Specificity with Receiver Operating Characteristic Curves. J. Data Science. 2008; 6:1–11.

[16]. Ye, J., et al. Heterogeneous Data Fusion for Alzheimer's Disease Study. Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD); 2008;

[17]. Balakrishnan, N. Handbook of the Logistic Distribution. Marcel Dekker, Inc.; 1991.

[18]. Krishnapuram B, et al. Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds. IEEE Trans Pattern Analysis and Machine Intelligence. Jun; 2005 27(no. 6):957–968.

[19]. Fisher RA. The Use of Multiple Measurements in Taxonomic Problems. Ann. of Eugenics. 1936; 7(no. 2):179–188.

[20]. Duda, RO.; Hart, PE.; Stork, DH. Pattern Classification. second ed. Wiley Interscience; 2000.

[21]. Fisher RA. The Statistical Utilization of Multiple Measurements. Ann. Eugenics. 1938; 8:376–386.

[22]. Vapnik, VN. The Nature of Statistical Learning Theory. Springer; 1995.

[23]. Landau SM, et al. Associations between Cognitive, Functional, and FDG-PET Measures of Decline in AD and MCI. Neurobiology of Aging. 2011; 32:1207–1218. [PubMed: 19660834]

[24]. Chen K, et al. Linking Functional and Structural Brain Images with Multivariate Network Analyses: A Novel Application of the Partial Least Square Method. NeuroImage. 2009; 47:602–610. [PubMed: 19393744]

[25]. Alexander GE, et al. Age-Related Regional Network of MRI Gray Matter in the Rhesus Macaque. J. Neuroscience. 2008; 28(no. 11):2710–2718.

[26]. Li R, et al. Large-Scale Directional Connections among Multi Resting-State Neural Networks in Human Brain: A Functional MRI and Bayesian Network Modeling Study. NeuroImage. 2011; 51:1035–1042. [PubMed: 21396456]

[27]. Shultz EK. Multivariate Receiver-Operating Characteristic Curve Analysis: Prostate Cancer Screening as an Example. Clinical Chemistry. 1995; 41:1248–1255. [PubMed: 7628115]

[28]. Chen K, et al. The Alzheimer's Disease Neuroimaging Initiative, Characterizing Alzheimer's Disease Using a Hypometabolic Convergence Index. NeuroImage. 2011; 56:52–60. [PubMed: 21276856]

[29]. Li R, et al. Attention-Related Networks in Alzheimer's Disease: A Resting Functional MRI Study. Human Brain Mapping. 2011; 33:1076–88. [PubMed: 21538702]

[30]. Payam, R.; Lei, T.; Huan, L. Cross Validation. In: Tamer Özsu, M.; Liu, L., editors. Encyclopedia of Database Systems. Springer; 2009.

Fig. 1.
Accuracy comparison of single-index, multiV-ROC classification and linear discrimination approaches in data set 1 without leave-one-out cross-validation.
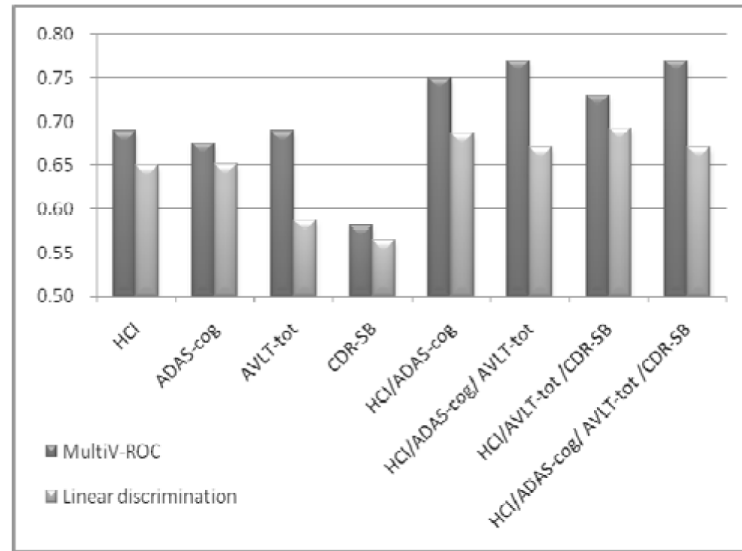
Fig. 2.
Accuracy comparison of single-index, multiV-ROC classification and linear discrimination approaches in data set 2 without leave-one-out cross-validation.
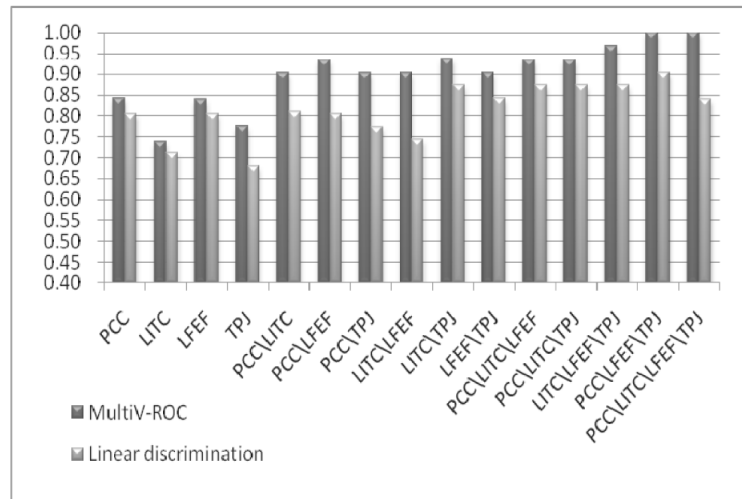
TABLE 1

The Classifying Results of Single-Index from Data Set 1

variable	Cut-off	Sen	Spe	Accu-racy	AUC	95% C.I
HCI	11.75	0.70	0.68	0.69	0.72	[0.56, 0.84]
ADAS-cog	16.05	0.87	0.54	0.68	0.76	[0.65, 0.88]
AVLT-LTM	−1.98	0.87	0.18	0.46	0.28	[0.16, 0.39]
AVLT-tot	−4.95	0.43	0.87	0.69	0.63	[0.50, 0.76]
CDR-SB	3.08	0.97	0.32	0.58	0.65	[0.51, 0.78]
MMSE	−25.94	0.87	0.25	0.50	0.55	[0.41, 0.68]

Note: HCI, AVLT and ADAS-cog are three baseline indices, which were each examined for their differentiating power separately or combined (HCI: FDG PET-based hypometabolic convergence index; ADAS-cog: scores of Alzheimer's disease assessment scale-cognitive subscale). AVLT-LTM: long-term memory score of the Auditory Verbal Learning Test. AVLT-tot: Adult Verbal Learning Test total score cf learning trials. The sensitivity/specificity pair (Columns labeled as Sen and Spe) in the table are based on the entire dataset. The area under curve (AUG) and its 95% confidence interval (C.I) were obtained without the leave-one-out procedure.

TABLE 2

The Classifying Results of Multi-ROC and Linear Fisher Discrimination from Data Set 1

		MultiV-ROC			Linear discrimination		
ariable	Cut-off	Sen	Spe	Ac-cu-racy	Sen	Spe	Ac-cura-cy
HCI/A DAS-cog	(13.20, −2.85)	0.77	0.73	0.75	0.71	0.67	0.69
HCI/A DAS-cog/ AVLT-tot	(13.92, 22.61, − 1.35)	0.80	0.75	0.77	0.73	0.63	0.68
HCI/A VLT-tot /CDR-SB	(13.92, 21.79, 8.92)	0.73	0.73	0.73	0.68	0.70	0.69
HCI/A DAS-cog/ AVLT-tot /CDR-SB	(13.92, 26.43, − 1.95, 8.84)	0.77	0.77	0.77	0.72	0.63	0.68

The parameter K and N, representing the number sub-intervals and randomly selected combinations, respectively, are the same for all indices. Specifically, k=200, N=50000

TABLE 3

Biomarker in Distinguishing the AD from NC

varia-ble	Cut-off	Sen	Spe	Accu-racy	AUC	95% C.I
PCC	−4.01	0.74	0.94	0.84	0.82	[0.66 0.97]
LITC	−2.50	0.66	0.81	0.74	0.82	[0.67 0.97]
LFEF	−2.88	0.87	0.81	0.84	0.89	[0.76 1.00]
TPJ	−3.42	0.73	0.81	0.77	0.83	[0.67 0.98]

Note: PCC, LFEF, LITC and TPJ are core regions from three resting state networks, whose intensity could differentiate AD from NC. PCC: posterior cingulated cortex; LFEF: left frontal eye field; LITC: left inferior temporal cortex; TPJ: temporal-parietal junction. The sensitivity/ specificity pair (columns labeled as Sen and Spe) in the table is based on the entire dataset. The AUC and its 95% CI were obtained without the leave-one-out procedure.

TABLE 4

The Classifying Results of Multi-ROC and Linear Fisher Discrimination from Data Set 2

varia-ble	Cut-off	MultiV-ROC			Linear discrimina-tion		
		Sen	Spe	Ac-cura-cy	Sen	Spe	Ac-cura-cy
PCC/L ITC	(−4.01, −2.40)	0.87	0.94	0.90	0.75	0.87	0.81
PCC/L FEF	(−4.01, −2.71)	0.93	0.94	0.94	0.88	0.73	0.80
PCC/T PJ	(−4.01, −3.11)	0.87	0.94	0.90	0.81	0.73	0.77
LITC/ LFEF	(−2.40, −2.71)	0.80	1.00	0.90	0.69	0.80	0.74
LITC/ TPJ	(−2.40, −3.11)	0.87	1.00	0.93	0.81	0.93	0.87
LFEF/ TPJ	(−2.40, −3.11)	0.80	1.00	0.90	0.75	0.93	0.84
PCC/L ITC/L FEF	(−3.40, −2.41, −2.70)	0.93	0.94	0.94	0.81	0.93	0.87
PCC/L ITC/T PJ	(−3.40, −2.40, −3.09)	0.93	0.94	0.94	0.81	0.93	0.87
LITC/ LFEF/ TPJ	(−2.40, −2.70, −3.09)	0.93	1.00	0.97	0.81	0.93	0.87
PCC/L FEF/T PJ	(−3.56, −2.70, −3.09)	1.00	1.00	1.00	0.94	0.87	0.90
PCC/L ITC/L FEF/T PJ	(−3.56, −2.35, −2.65, −3.04)	1.00	1.00	1.00	0.81	0.87	0.84

The parameter K and N, representing the number sub-intervals and combinations respectively, are the same for all indices. Specifically, k=200, N=50000.