Identifying dementia in MRI scans using machine learning

Victor Miller, Stephen Erlien, and Jeff Piersol

Abstract—A support vector machine and naive Bayes classifier are used to identify the presence of dementia in MRI brain scans. Features are computed for each scan: scans are processed via K-means to segment the image into different tissue types, from which different quantities are computed (e.g., total gray matter, a measure of symmetry); principal component analysis is used to reduce dimensionality of the images; and forward feature selection is used to identify the most important principal components, which are then included in the feature set. Using a naive Bayes classifier, we were able to obtain a classification accuracy of 87.73%, with precision of 87.95% and recall of 78.49%.

Keywords—dementia, eigenbrain, MRI

I. INTRODUCTION

DEMENTIA currently afflicts more than 36 million people globally, and this number is projected to grow to 100 million by the year 2050. The disease is tragically debilitating for the individual, and also takes a significant emotional toll on family and friends. Furthermore, caring for those who suffer from dementia currently costs the US healthcare system roughly $100 billion per year, and is forecast to cost $1 trillion per year by 2050 [1].

Diagnosis of dementia is typically done through a clinical examination of a patient's mental state, which can be a subjective measure of one's health, and often does not provide a concrete indicator of which type of treatment is most likely to produce favorable results. Neurological MRI imaging has provided physicians with better means to diagnose dementia, and physicians have discovered that dementia can exist in a variety of "mixed" forms [1]. Even so, neurological MRI scans are complex and often noisy, which make them a challenge to interpret even for a neurological radiologist with years of training and experience. In this work, we apply machine learning techniques to identify those MRI scans that indicate some form of dementia, with the goal of increasing the likelihood of early detection of dementia. Given the education, experience, and expertise of neurological radiologists, our goal is not to replace a radiologist, but to provide a tool to help prioritize MRI scans which display the most potential for disease diagnosis.
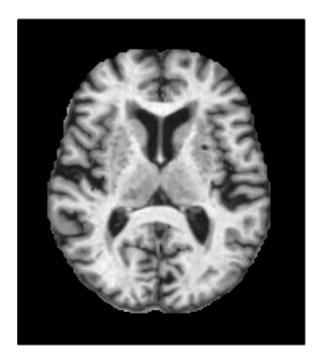

Fig. 1.   Typical axial scan from OASIS data collection

This report is structured as follows. First, we present the dataset of MRI scans that contains both healthy and diseased brains. Second, an initial set of features is motivated and the methods employed to extract these features are described. Third, initial results of machine learning algorithms applied to this small set of features are presented. Based on the results of this initial set of features, additional features are motivated and the methods to determine these features described. Lastly, the results from ML algorithms applied to this larger set of features are presented and discussed.

II. DATA

Our work utilizes a dataset containing 436 neurological MRI scans made available by the Open Access Series of Imaging Studies (OASIS) project [3]. The subjects' ages range from 18 to 96 with 100 subjects being clinically diagnosed with very mild to moderate Alzhiemer's disease. Each scan consists of 176x208x176 voxels which have been preprocessed to remove the skull, leaving only brain matter in the images [4]. Each scan is associated with additional information about the subject,
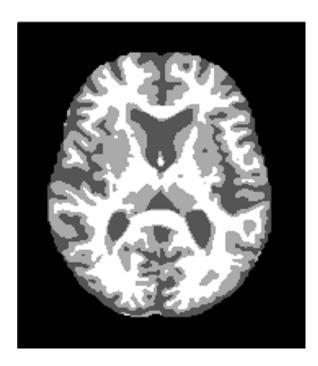
Fig. 2. Typical axial scan preprocessed into segmented colors

including age, sex, education level, socioeconomic status, intracranial volume, and normalized brain volume, and two measures of dementia: the clinical dementia rating (CDR) and mini-mental state exam (MMSE).

A single, axial slice of a typical scan is shown in Fig. 1. In addition to the preprocessing to remove the skull, the scans have been further processed by OASIS (using the k-means algorithm) to bin pixel intensities into one of four colors which correspond to the background (0), cerebrospinal fluid (CSF) (1), gray matter (2), and white matter (3). This segmented image is shown in Fig. 2.

III. FEATURE EXTRACTION

Each subject is labeled as demented or not-demented based on his MMSE score, where a subject receiving an MMSE score less than 30 is classified as demented (30 is the maximum possible score and corresponds to a healthy individual). For each subject, a set of features is created using the previously mentioned information included for that subject (e.g., age, gender) and additional features extracted from the subject's brain scan. It is known that the severity of dementia is correlated with volume of gray and white matter [2], and so the total amount of gray matter, white matter, and cerebrospinal fluid in each slice of the scan is computed. Asymmetry in brain structure can also be indicative of abnormality, so a measure of symmetry in each slice is computed by taking the correlation of a given slice with its flipped

image for zero lag, normalized by the integrated signal in the original slice:

$$s = (g \star g_f)(0) = \frac{\sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} g[x,y] g_f[x,y]}{\sum_{x=-\infty}^{\infty} g[x,y]} \quad (1)$$

Symmetry measures are computed for left-right and up-down symmetry about the middle of the image, and an example measure of left-right symmetry is displayed in Fig. 3. The average value of this symmetry is used in the feature set.

In order to incrementally build complexity into our work, we start by using age, sex, education level, socioeconomic status, intracranial volume, normalized brain volume, average measures of gray matter, white matter, CSF, and up-down and left-right symmetry as an initial feature set.

IV. INITIAL RESULTS

A Support Vector Machine (SVM) and a naive Bayes modelled with normally distributed features were used on the feature set described above. An SVM was chosen to exploit linearly separable patterns in the data, while naive Bayes was chosen to exploit the normal distribution of the feature data. Originally, only the features described in the previous section were used; raw pixel data was not used directly in the feature set. Moreover, each feature was normalized to values between 0 and 1, and performance was determined using 10-fold cross validation.

The learning curve for an SVM with a linear kernel is shown in Fig. 5. The performance with the full dataset is 84.7%; however, the data only has 166 examples of demented subjects and a trivial classifier could easily achieve a performance of 73% by always outputting "no dementia". For this reason, the precision and recall metrics are calculated and displayed as well in Fig. 5
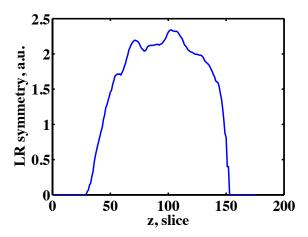

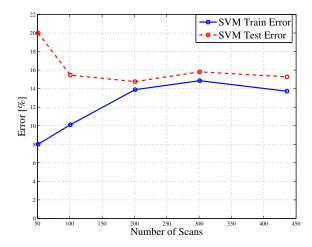Fig. 3. Example of LR-symmetry vs slice of a typical scan
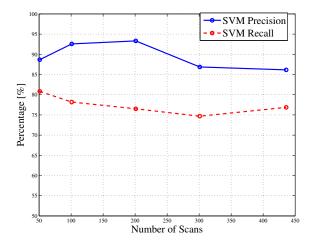
Fig. 4. SVM learning curve on initial feature set

Fig. 5. SVM precision and recall on initial feature set

The learning curve suggests we have high bias and therefore would benefit from additional features. To ensure that the SVM is not heavily reliant on non-medical features, we removed the education level and socioeconomic status and achieved a nearly identical performance.

Preliminary Naive Bayes classification yielded results similar to that of an SVM.

V. ADDITIONAL FEATURES

The initial results indicate high error rates in both the training and the test data, implying high bias in the model. These results suggest additional features could be used to improve performance. A close examination of scans from demented and non-demented subjects, like those illustrated in Fig. 6, provides insight into additional features that may be useful in identifying demented subjects. As indicated, the demented subject shows cerebrospinal fluid in the areas which are occupied by brain
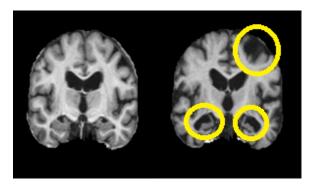

Fig. 6. Healthy brain (left) versus demented brain (right) with structural differences highlighted

matter in the non-demented subject. Principal component analysis (PCA) is used to capture these structural features from the voxel data in the scans. Running PCA on the entire scan is computationally burdensome, and instead coronal slice 110 (shown in Fig. 6) is used since it captures structures indicative of dementia.

The mean-centered and unit-variance scaled 110th coronal slice from each subject was represented as a vector, $x^{(i)}$. PCA on this set of vectors produced a set of eigenbrains, each represented as a vector, $v^{(i)}$. In class, we discussed computing the eigenvectors of the empirical covariance matrix:

$$\frac{1}{m}\sum_{i=1}^{m} x^{(i)}x^{(i)T} \qquad (2)$$

as a method for determining the principal components. In practice, explicit computation of this covariance matrix is computationally burdensome for the 176x208 pixels in the slice. Methodologies that use SVD decomposition provide much faster methods for computing the principal components. Moreover, MATLAB has a built-in function, *princomp*(), which computes the PCA decomposition directly from the given data, and we leveraged this capability in our work. The results of running PCA on this slice for all subjects is a set of eigenbrains, some of which appear to capture structures found in demented subjects as seen in Fig. 7 while others appear to capture general brain structure as seen in Fig. 8. Using all of these eigenbrains as additional features causes our learning algorithms to overfit the training examples and subsequently perform poorly on the test subjects. Therefore, forward feature selection was used to select only the eigenbrains that appear to capture structures useful in classifying demented subjects. Eigenbrains were added as features in this manner until the increases in performance were minimal. This method resulted in about five eigenbrains being used as additional features.

Once the eigenbrains were computed, given a subject's
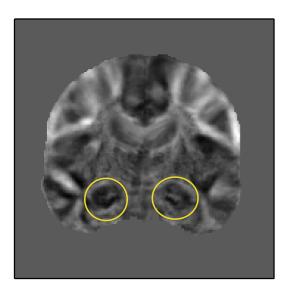
Fig. 7. Eigenbrain capturing dementia indicating structure as identified by forward feature selection
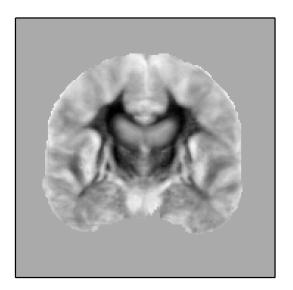

Fig. 8. Eigenbrain capturing general brain structure

110th axial scan and the ith eigenbrain, a feature, $y^{(i)}$, can be calculated as

$$y^{(i)} = v^{(i)T}x^{(i)} \tag{3}$$

This scalar can then be used as a feature to indicate the presence of the eigenbrain in the subject's scan.

Another method explored to extract features was region counting. As the cerebrospinal fluid seems to occupy larger contiguous regions in scans of demented subjects, the area of CSF seems like a natural feature to use in classification. However, our results indicate no substantial improvement when using such features.
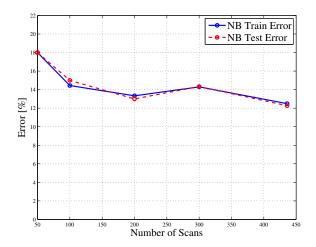

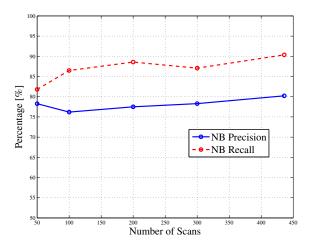Fig. 9. Naive Bayes classification error on final feature set

Fig. 10. Naive Bayes precision and recal on final feature setl

Region counting was implemented by adding the largest three regions of light matter, gray matter, and CSF in slice 110 as 9 new features to the original feature set, and no significant change in test error was observed. It is likely that the effects of region sizes are already captured in the features that give the total volume of light matter, gray matter, and CSF.

VI. FINAL RESULTS

After using forward feature selection to incorporate the best features, SVM classification yielded a test error of 12.70%, and naive Bayes yielded test error of 12.97%. Importantly, the recall rate using an SVM was 80.24%, meaning that 19.76% of dementia cases were not detected. Using naive Bayes, 12.65% of dementia cases were undetected.

It should be noted that age alone seems to be a strong indicator of the presence of dementia: a naive Bayes model with age as the only feature yielded a test error
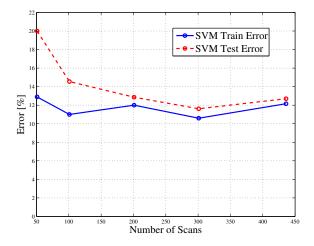
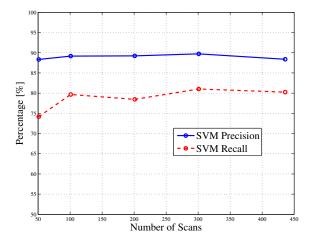Fig. 11. SVM classification error on final feature set

Fig. 12. SVM precision and recall on final feature set

of 13.64% and recall of 90.36%. However, an SVM with the same setup had nearly 60% error, and naive Bayes classification with the full feature set excluding age had 13.90% error. Therefore, the classifier is not merely keying to age data in order to produce an accurate result. The likely reason for the performance using only age is that there is a high concentration of data samples with age near 20, resulting in a non-dementia likelihood $\mathbb{P}(age|no\ dementia)$ that is heavily skewed to lower age values.

VII. CONCLUSION

By extracting features from brain scans, selecting relevant features, and classifying the scans based on their features, we were able to detect dementia with 87.73% accuracy. The most indicative features were found to be normalized brain volume, volume of gray matter, and volume of CSF. Though far from being able to replace a neurological radiologist, the classifier is of great benefit in prioritizing scans for radiologists to analyze.

VIII. FUTURE WORK

Running classification on a larger dataset that more accurately represents a cross-section of the population would help verify or refute the results presented. Particularly, the results of naive Bayes depend on $\mathbb{P}(age|no\ dementia)$, and the OASIS dataset used does not accurately reflect the true likelihood of age given no dementia.

Since the class labels for dementia were found by discretizing a relatively continuous scale, a helpful improvement may be to cast the problem as a multi-class classification problem with more possible output classes. This would add another layer of prioritization for doctors who will analyze these scans.

REFERENCES

[1] S. Hall. "The dementia plague." Internet: www.technologyreview.com/featuredstory/429494/the-dementia-plague/page/1/, October. 2012 [Dec. 2012].

[2] JC Stout. (1996, Aug.). "Association of Dementia Severity With Cortical Gray Matter and Abnormal White Matter Volumes in Dementia of the Alzheimer Type." *Archives of Neurology*. 53(8):742-749.

[3] Marcus, DS, Wang, TH, Parker, J, Csernansky, JG, Morris, JC, Buckner, RL. "Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults." *Journal of Cognitive Neuroscience*. 19:1498-1507.

[4] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, S.M. Smith. "FSL". *NeuroImage*. 62:782-90, 2012