Voxelwise Detection of Cerebral Microbleed in CADASIL Patients by Naive Bayesian Classifier

Fangzhou Bao¹, Meiling Shi² and Felix Macdonald³
¹School of Computer Science and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China
²School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100083, China
³Université de Sherbrooke, Sherbrooke, Québec, QC J1K 2R1, Canada

Abstract—It is important to detect cerebral microbleed voxels from the brain image of cerebral autosomal-dominant arteriopathy with subcortical infarcts and Leukoencephalopathy (CADASIL) patients. Methods developed by other researchers before have a high variability of intra-observer and inter-observer. In our study, we collect our dataset from the 20 brain volumetric images, 10 for CADASIL patients and 10 for healthy controls. And we used naive bayesian classifier to get the results. We use cross validation to improve the performance of naive Bayesian classifier. The results show that the average sensitivity is 74.53±0.96%, the average specificity is 74.51±1.05%, and the average accuracy is 74.52±1.00%.

Keywords—CADASIL; voxel; naive Bayesian classifier; cross validation

I. INTRODUCTION

The cerebral autosomal-dominant arteriopathy with subcortical infarcts and Leukoencephalopathy (CADASIL) syndrome is the most common hereditary stroke diseases which can never be ignored.

The subclinical sign of cerebral microbleed (CMB) [1] of CADASIL patients can be detected years prior to clinical manifestation by magnetic resonance imaging (MRI). Susceptibility weighted imaging (SWI) is a 3D flow-compensated T2* imaging technique. Compared to standard MRI protocol, the SWI scans the patient at high resolution (less than 1 mm). It uses the phase image to enhance the contrast, providing 2 to 6 times increased sensitivity in CMB detection.


Though methods mentioned above are very important attempts in the past, they are not acceptable for 2 reasons: (1) The complexity of these methods are very high. (2) Their detection do not reach voxel-wise resolution.

Our method used naive Bayesian theory to perform the classification task. For general classification tasks, if we have enough training data and the features are well designed, then the accuracy will be high enough to accept. It estimates the posterior probability by priori probability collected for training. And the mathematical foundations for Bayesian theory provides solid explain ability without much complexity. Besides, our purpose it to identify CMB at voxel resolution. In all, our methods is a new computer vision [7-10] and image processing [11-14] approaches.

II. METHODOLOGY

Naive Bayesian classifiers are classifiers based on statistical analysis [15]. They can predict class membership probabilities i.e. the probability that the sample with given values of its features belongs to a specific category [16]. Naive Bayesian classifiers assume that the effect of an attribute value on a given category is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computations involved and, in this sense, is considered “naive.”

Naive Bayesian classifier can present a comparable performance to state-of-the-art classifiers, such as support vector machine [17-21], shallow neural network [22-27], etc.

Let \( X \) be a sample which is described by measurements made on a set of \( n \) attributes. Let \( H \) be some hypothesis such as that the data sample \( X \) belongs to a specific category \( C \). For classification problems, we want to determine \( P(H|X) \), the probability that the hypothesis \( H \) holds given the observed sample \( X \). In other words, we are looking for the probability that sample \( X \) belongs to category \( C \), given that we know the attribute description of \( X \).

\( P(H|X) \) is the posterior probability of \( H \) conditioned on \( X \). \( P(H) \) is the prior probability. Similarly, \( P(X|H) \) is the priori probability of \( X \) conditioned on \( H \), and \( P(X) \) is the prior probability of \( X \). Bayes’ theorem is useful in that it provides a way to calculate the posterior probability, \( P(H|X) \), from \( P(H) \), \( P(X|H) \), and \( P(X) \). Bayes’ theorem is [28]:

\[
P(H|X) = \frac{P(X|H)P(H)}{P(X)}
\] (1)

Let \( D \) be a training set of samples and their associated class labels, each sample is represented by an \( n \)-dimensional attribute vector, \( X = (x_1, x_2, ... , x_n) \), depicting \( n \) measurements made on the sample from \( n \) attributes, respectively, \( A_1, A_2, ..., A_n \).
Suppose that there are \( m \) categories, \( C_1, C_2, \ldots, C_m \). Given a sample \( X \), the classifier will predict that \( X \) belongs to the category having the highest posterior probability \([29]\), conditioned on \( X \). That is, the naive Bayesian classifier predicts that sample \( X \) belongs to the class \( C_i \) if and only if
\[
P(C_i|X) > P(C_j|X) \text{ for } 1 \leq j \leq m, j \neq i. \tag{2}
\]

By Bayes’ theorem, we have
\[
P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}. \tag{3}
\]

Since \( P(X) \) is a constant for all categories, only \( P(X|C_i)*P(C_i) \) needs to be determined. Note that the class prior probabilities may be estimated by
\[
P(C_i) = \frac{|C_i,D|}{|D|} \tag{4}
\]
where \( |C_i,D| \) is the number of training samples of class \( C_i \) in \( D \) and \( |D| \) is the total number of training samples of \( D \).

Given data sets with many attributes, it would be extremely computationally expensive to compute \( P(X|C_i) \). To reduce computation in evaluating \( P(X|C_i) \), the naive assumption of class-conditional independence is made. This presumes that the attributes’ values are conditionally independent of one another, given the class label of the sample (i.e., that there are no dependence relationships among the attributes). Thus, we can obtain \( P(X|C_i) \) by following formula:
\[
P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i). \tag{5}
\]

To predict the class label of \( X \), \( P(X|C_i)*P(C_i) \) is evaluated for each class \( C_i \). The classifier predicts that the class label of sample \( X \) is the class \( C_i \) by formula (2) mentioned above.

![Figure I. Illustration of Naïve Bayesian Classifier](image)

Figure 1 shows the illustration of naive Bayesian classifier. We did not use deep learning methods, because this dataset is relatively small, and the convolution operation \([30-35]\) is difficult to handle on an 7x7-size input.

### III. Material and Preprocessing

Figure 2 shows a portion of generated dataset. It can be regarded as a matrix with width of 50 and height of 50. Each row of this matrix represents a sample, the first 49 columns represent input data and the 50-th column represents the target. Input data were generated by vectorizing sliding neighborhood with size of 7x7.

![Figure II. Generated Dataset (Column 1-49 are Inputs, Column 50 is the Target)](image)

All the pixels in the 7x7 neighbor were regarded as features, and submitted to the naïve Bayesian classifier. We collected in total 20 subjects, and obtained 69,356 CMB voxels, and 124,063,981 non-CMB voxels. For class imbalance problem, we selected randomly 69,327 non-CMB voxels from those 124,063,981 samples. Now we have a dataset as listed in Table 1.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMB</td>
<td>69,365</td>
</tr>
<tr>
<td>Non-CMB</td>
<td>69,327</td>
</tr>
<tr>
<td>Total</td>
<td>138,692</td>
</tr>
</tbody>
</table>

A 10-fold cross validation was used to segment the dataset into 10 folds, and report the out-of-sample error. For avoiding the effect of randomness, we run the 10-fold cross validation ten times, and report the average and standard deviation of sensitivity, specificity, and accuracy. The definition of those three measures are depicted by following three equations:
\[
\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}
\]
\[
\text{Specificity} = \frac{TN}{FP + TN} \tag{7}
\]
\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{8}
\]

where TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative.

Figure 3 shows the illustration of one run of cross validation. Figure 3(a) shows the index, where the x-axis is the combination of both CMB and non-CMB voxels, in total 138,692. Figure 3(b) shows the legend.
IV. RESULTS

The sensitivities, specificities, and accuracies of this 10×10-fold cross validation were shown below in Table 2. The average sensitivity is 74.53±0.96%, the average specificity is 74.51±0.05%, and the average accuracy is 74.52±1.00%. Figure 4 shows the box plot of our algorithm, from which we can observe the three measures can similar results, indicating the effectiveness of our technique to handle the imbalance problem.

TABLE II. PERFORMANCE OF OUR NBC METHOD

<table>
<thead>
<tr>
<th></th>
<th>Sen</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>77.4</td>
<td>74.3</td>
<td>75.9</td>
<td>76.3</td>
<td>78.7</td>
<td>76.3</td>
<td>75.5</td>
<td>75.3</td>
<td>75.1</td>
<td>79.0</td>
<td>76.4</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>70.9</td>
<td>77.3</td>
<td>75.5</td>
<td>78.6</td>
<td>76.1</td>
<td>69.4</td>
<td>76.3</td>
<td>75.4</td>
<td>73.4</td>
<td>72.8</td>
<td>74.3</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>73.9</td>
<td>77.2</td>
<td>70.4</td>
<td>74.6</td>
<td>71.1</td>
<td>75.3</td>
<td>75.3</td>
<td>73.4</td>
<td>71.2</td>
<td>74.8</td>
<td>73.6</td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>73.8</td>
<td>73.1</td>
<td>75.6</td>
<td>74.5</td>
<td>73.1</td>
<td>72.0</td>
<td>73.8</td>
<td>70.7</td>
<td>72.4</td>
<td>71.2</td>
<td>73.0</td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>75.4</td>
<td>74.2</td>
<td>71.6</td>
<td>76.9</td>
<td>75.0</td>
<td>74.3</td>
<td>76.4</td>
<td>74.5</td>
<td>75.1</td>
<td>75.3</td>
<td>74.9</td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td>74.4</td>
<td>75.3</td>
<td>73.0</td>
<td>74.8</td>
<td>71.8</td>
<td>75.3</td>
<td>75.8</td>
<td>77.2</td>
<td>76.3</td>
<td>75.5</td>
<td>74.9</td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td>75.0</td>
<td>76.6</td>
<td>70.8</td>
<td>76.9</td>
<td>73.5</td>
<td>74.2</td>
<td>78.2</td>
<td>74.5</td>
<td>78.2</td>
<td>72.1</td>
<td>75.1</td>
<td></td>
</tr>
<tr>
<td>R8</td>
<td>71.6</td>
<td>73.8</td>
<td>74.1</td>
<td>72.2</td>
<td>74.4</td>
<td>72.4</td>
<td>76.2</td>
<td>71.8</td>
<td>73.6</td>
<td>75.0</td>
<td>73.5</td>
<td></td>
</tr>
<tr>
<td>R9</td>
<td>75.0</td>
<td>73.2</td>
<td>75.0</td>
<td>73.0</td>
<td>77.4</td>
<td>77.2</td>
<td>73.2</td>
<td>74.9</td>
<td>73.8</td>
<td>72.6</td>
<td>77.8</td>
<td></td>
</tr>
<tr>
<td>R10</td>
<td>73.6</td>
<td>79.6</td>
<td>76.1</td>
<td>73.3</td>
<td>73.8</td>
<td>75.9</td>
<td>72.2</td>
<td>76.8</td>
<td>76.5</td>
<td>74.5</td>
<td>74.4</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Spc</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
<th>F6</th>
<th>F7</th>
<th>F8</th>
<th>F9</th>
<th>F10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>75.3</td>
<td>76.6</td>
<td>75.8</td>
<td>76.6</td>
<td>78.7</td>
<td>77.4</td>
<td>76.2</td>
<td>74.8</td>
<td>74.6</td>
<td>76.4</td>
<td>76.2</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>75.3</td>
<td>74.6</td>
<td>75.4</td>
<td>72.6</td>
<td>74.9</td>
<td>74.9</td>
<td>74.7</td>
<td>75.2</td>
<td>75.3</td>
<td>70.1</td>
<td>74.3</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>74.0</td>
<td>72.5</td>
<td>74.3</td>
<td>73.7</td>
<td>73.0</td>
<td>69.8</td>
<td>71.2</td>
<td>78.2</td>
<td>71.9</td>
<td>74.6</td>
<td>73.3</td>
<td></td>
</tr>
<tr>
<td>R4</td>
<td>75.7</td>
<td>74.8</td>
<td>71.0</td>
<td>70.2</td>
<td>75.3</td>
<td>74.0</td>
<td>67.8</td>
<td>722</td>
<td>73.7</td>
<td>72.5</td>
<td>72.7</td>
<td></td>
</tr>
<tr>
<td>R5</td>
<td>76.4</td>
<td>77.4</td>
<td>72.2</td>
<td>74.6</td>
<td>76.7</td>
<td>76.8</td>
<td>72.9</td>
<td>74.7</td>
<td>72.4</td>
<td>77.6</td>
<td>75.2</td>
<td></td>
</tr>
<tr>
<td>R6</td>
<td>74.2</td>
<td>73.8</td>
<td>73.7</td>
<td>75.8</td>
<td>73.1</td>
<td>77.6</td>
<td>74.7</td>
<td>75.0</td>
<td>74.2</td>
<td>78.3</td>
<td>75.0</td>
<td></td>
</tr>
<tr>
<td>R7</td>
<td>73.0</td>
<td>76.1</td>
<td>75.0</td>
<td>75.8</td>
<td>76.6</td>
<td>77.3</td>
<td>70.5</td>
<td>75.8</td>
<td>75.2</td>
<td>73.7</td>
<td>75.0</td>
<td></td>
</tr>
</tbody>
</table>

In the next experiment, we compared NBC with decision tree using ID3 algorithm [36]. The results were shown in Figure 5. Here ID3 obtained a lower performance, with an overall
The results showed better result than ID3. System using naive baysian classifier for CADASIL patients. Can meet the requirement of real applications. Preprocessing, to see if we can achieve better accuracy which other advanced techniques for classification and image processing, to see if we can achieve better accuracy which can meet the requirement of real applications.

We shall test using optimization algorithm [37, 38] to improve the performance of our NBC classifier. In addition, we shall collect more data, and try deep learning methods.

**REFERENCES**


[32] Pan, C., Abnormal breast identification by nine-layer convolutional...
neural network with parametric rectified linear unit and rank-based stochastic pooling. Journal of Computational Science, 2018. 27: p. 57-68


