Chapter 5

Discussion

5.1 Why not using whole voxels as features in classification

In our system, a ROI process was applied to select some regions where reveal significant differences between two groups as better features in classification. Although whole voxels preserve all data information, it is intuitive that some information may be useless and may interfere with performances of post-processings. Except for this consideration and limitations in practice, we also designed an experiment to show that taking some voxels is better than using whole voxels in classification.

The experimental group was composed of eight patients carrying bipolar disorder and ten normal subjects. Eighteen subjects are picked out from the study groups listed in Chapter 4 and their clinical data are summarized in Table 5.1. Brain volume data of all subjects were segmented into grey matter, white matter and CSF respectively. These same tissue data of different subjects were then normalized to a template and modulated to correct their volume changes. The volume of each voxels was collected as our original materials.

Table 5.1: Clinical data of experimental groups.

characteristics	Healthy controls			BD patients		
	$M+F$	M	F	$M+F$	M	F
amount(n)	10	5	5	8	3	5
age years(mean)	31.9	31.2	32.6	34	33	34.6

Two feature selection methods were adopted to assemble materials from original data for the following classification. One is to maintain all voxels of original data and the other is to select some voxels from the original data which depend on brain structural discrepancies between eight patients and ten controls. Thus, a voxel-based morphometric analysis was applied on this study group and a statistical *t*-map was produced to tell the significance of

each voxel. Furthermore, a significant level was set up to be a criterion on voxel selections and some distinguishable voxels are acquired after the original data are thresholded.

A similarity measure which computes distances between unknown samples and known groups was used to decide where those unknown samples belong instead of determining their nature by estimating their probabilities existing in some specific groups. Suppose that an unknown sample $\mathbf{x}$ is needed to be classified and there are c categories $\{g_1, ..., g_c\}$ with corresponding data points $\{m_1, ..., m_c\}$. The distance between $\mathbf{x}$ and g_i is calculated by averaging distances between $\mathbf{x}$ and each data point in category g_i and can be formularized as

$$\mathbf{D}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} \left\| \mathbf{x} - g_{ij} \right\|. \tag{5.1}$$

Then, $\{\mathbf{D}_1, ..., \mathbf{D}_c\}$ are obtained and compared to find the category with the shortest distance of all. The unknown sample $\mathbf{x}$ is classified into the found category because it is more closer to the found category than others. We also used leave-one-out cross validation to evaluate the classification accuracy.

Table 5.2: **Performance of GM classifier.** This classification is based on the grey matter volume differences between ten normal subjects and eight BD patients. The row of total voxels represents first feature selection method which contains whole voxels. The row of distinguishable voxels represents second feature selection method which applies ROI selection on original data.

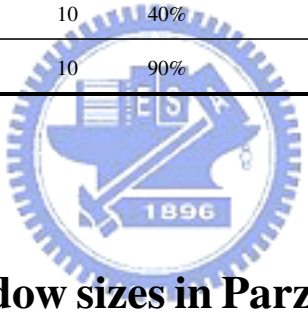
GM classifier	Healthy controls			BD patients		
method	$misclassification$	$total$	$accuracy$	$misclassification$	$total$	$accuracy$
total voxels	3	10	30%	6	8	75%
distinguishable voxels	10	10	100%	8	8	100%

In this experiment, a GM classifier and a WM classifier were built. Their performance with two feature selection methods are shown in Table 5.2 and Table 5.3. It is clear to see that the performance of second feature selection method is better than that of first feature selection method no matter which classifier was used. This result verifies our thought

that some voxels where reach significant differences between two groups are indeed better than others to be features in classification. Moreover, second feature selection method accelerates the system efficiency due to dimensionality reduction. In short, we recommend applying a ROI selection before classification.

Table 5.3: **Performance of WM classifier.** This classification is based on the white matter volume differences between ten normal subjects and eight BD patients. The row of total voxels represents first feature selection method which contains whole voxels. The row of distinguishable voxels represents second feature selection method which applies ROI selection on original data.

WM classifier	Healthy controls			BD patients		
method	$misclassification$	$total$	$accuracy$	$misclassification$	$total$	$accuracy$
total voxels	4	10	40%	7	8	87.5%
distinguishable voxels	9	10	90%	8	8	100%



5.2 Influences of window sizes in Parzen-window approach

In Parzen-window density estimation approach, the parameter, window size, plays an important role in the accuracy in our classification model. As mentioned in section 3.3, the estimated density will be oversmoothed if the window size is large and seem noisy if it is small. So, it is easy to have inaccurate density estimation if a bad window size is used. Also, an inaccurate estimated density will lead to a poor outcome of the posterior probability and make a wrong prediction on a classification system. Thus, it must be careful to choose the window size.

We designed a simple experiment to tell influences of varying the window size in Parzen-window approach on the classification model. The experimental group was composed of fifteen patients suffering bipolar disorder and 76 normal subjects and it was the same as the study group used in section 4.3. Brain volume data of all subjects were seg-

mented into grey matter, white matter and CSF parts respectively. Here, only GM partitions were used to construct a GM classifier to verify the effects of varying the window size on this GM classifier. GM images of all subjects were then normalized to a customized GM template, modulated to correct their volume changes and masked with a GM mask. Both of the customized GM template and the GM mask were obtained in section 4.3. The remainder features were collected as our original materials.

A principal component analysis was then applied on the original materials to find proper representations of them with fewer variables. Here, only the variance-based principal component selection method was used to choose some more useful characteristics as features for classification processes. We accounted for 60% of the variance in the reference set for a total of 36 eigenvectors. The parameter, prior, was set 0.165 and multivariate Gaussian distribution was used as the window function in density estimation. We employed a leave-one-out cross validation to verify our classification results. Now, we changed the parameter, window size, to test the effects by varying it.

Variations of predicted probabilities of each BD patient are illustrated in Figure 5.1 where reveals that the abnormal possibility of each BD patient is getting lower as the window size is getting larger. It is reasonable that the amplitudes of density function of normal and BD groups vary as the window size changes. Moreover, because the amount of both normal and abnormal groups are limited and the normal group is much larger and tighter than the abnormal group, the amplitude of density function of normal group is getting larger than that of BD group when the window size gets large gradually. Figure 5.2 illustrates the corresponding classification accuracy of both normal and abnormal groups. Due to the variations of both density functions in increasing the window size, the probability of belonging to normal group is getting higher. So, the classification accuracy of the BD group is getting lower and lower and that of normal group becomes high, when the window size is over 7.

In our thesis, we proposed an efficient visualization method to decide the window size

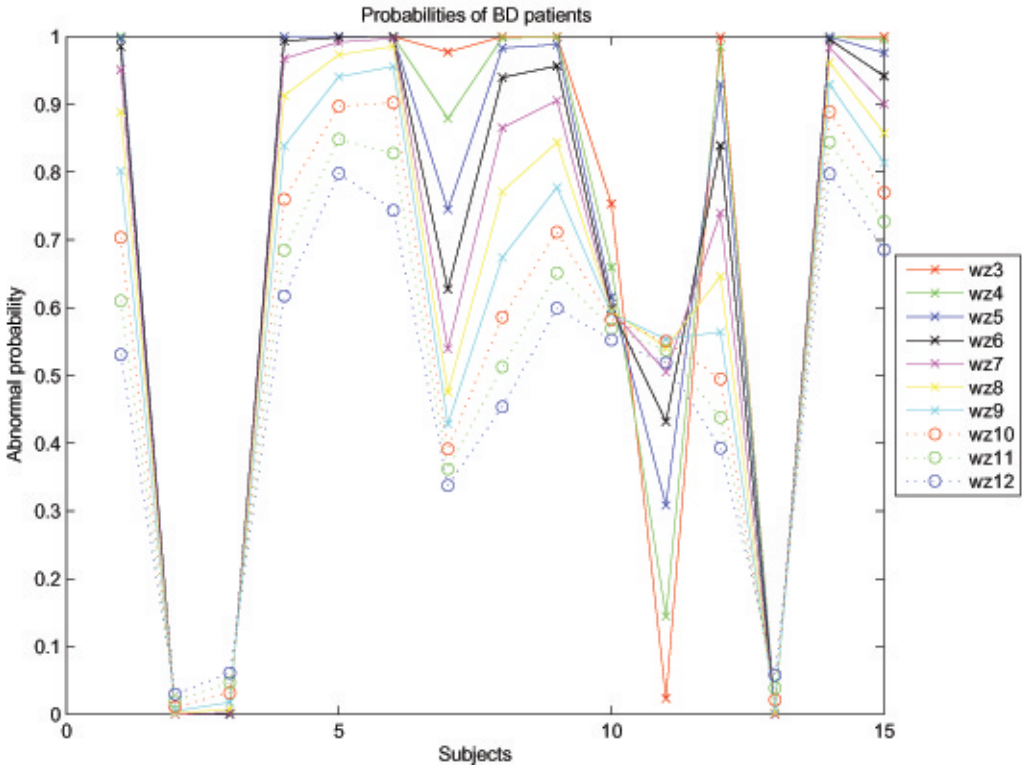


Figure 5.1: **Predictions of BD patients by varying window size.** Here lists the predictions of fifteen test subjects by the BD classifier constructed in our work. Test subjects were clinical diagnosed as BD patients by professional doctors. The value of window size is changed to test influences on classification. This figure shows that the predictions become less accurate when the window size gets large.

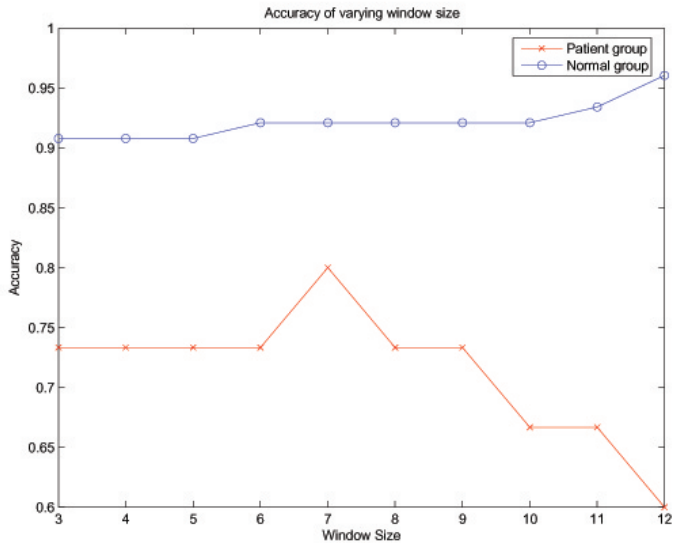


Figure 5.2: **Classification accuracy on the BD classifier.** This graph illustrates the variations of classification accuracy of the normal group and the BD group. When the window size gets large, the accuracy of the BD group becomes less accurate and that of the normal group becomes more accurate.

used in density estimation instead of trying lots of values to find a better classification results. After finding a proper representation for the training data, we picked out the first two principal components as axes to form a coordinate system and projected the data into this new space. Then, we sketched the distribution and the density function of two groups in different window size as shown in Figure 5.3 where the window size varied from 1 to 8. An improper window size makes the density function oversmoothed or erratic. According to the variations of density function in each kind of window size, we could easily determine which one may be good to use in estimation by visualization.

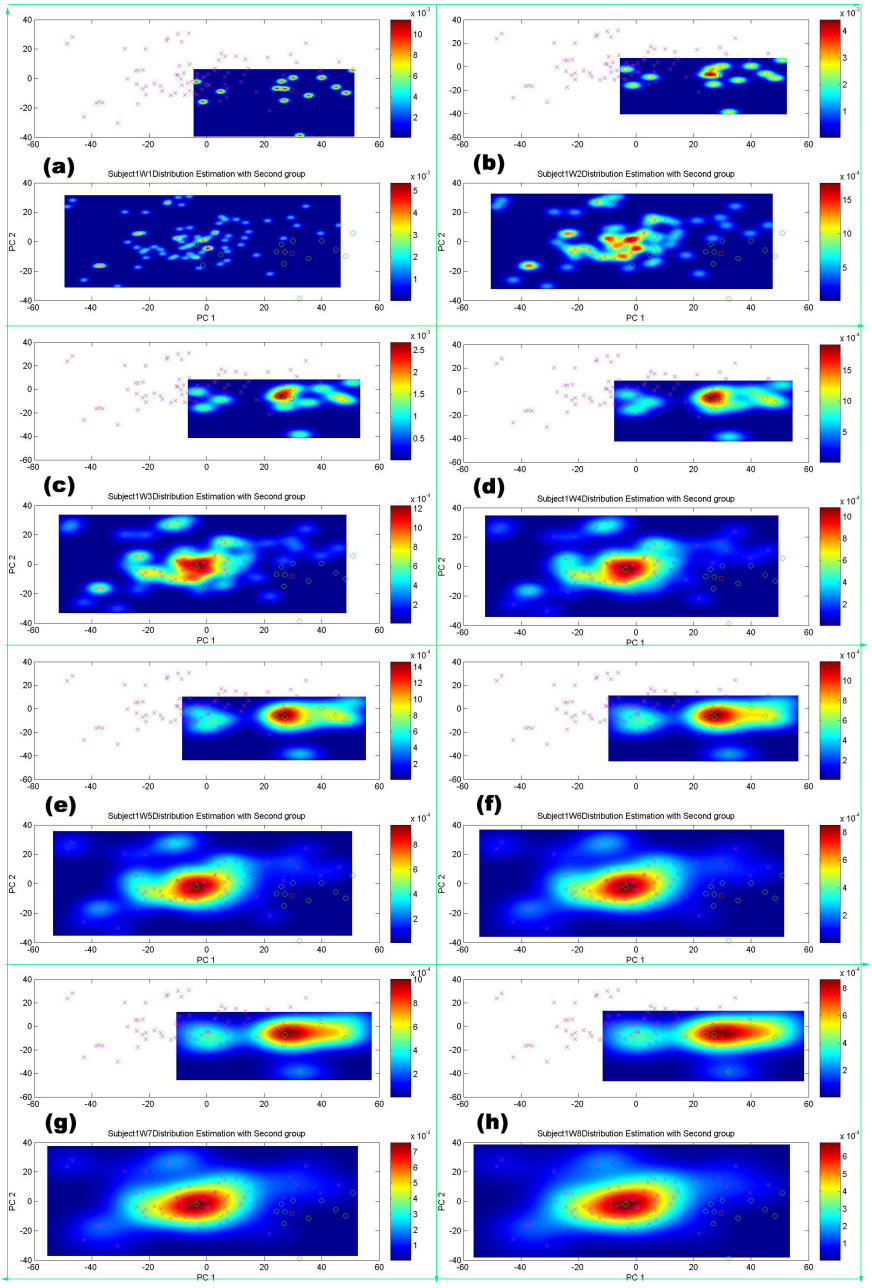


Figure 5.3: **Visualization of density functions in varying the window size.** From Figure (a) to Figure (h) are density function estimated by varying window size from 1 to 8. For each figure, the above graph is the density function of BD group and the below one is that of normal group. When the window size equals to 4 or 5, we could see that both distributions are smooth enough. Therefore, we will use these two values to estimate the density functions.

5.3 Comparisons between variance-based PC selection and significant-based PC selection

We use the receiver operating characteristic (ROC) curve to evaluate the performance of classifiers constructed in our work. Because the classification accuracy of the SCA3 classifier reached 100%, we only sketched the ROC curve of BD classifiers to compare the efficiency with two different PC selection methods. In our proposed classification model, the final classifier was composed of GM, WM and CSF classifiers with corresponding optimal parameters. Parameters indicate how many eigenvectors were used in the classification model and a variance ratio method was used to decide it. We varied the variance ratio from 10% to 100% in our experiments.

Combining the GM, WM and CSF classifiers, there are 1000 combinations to determine characteristics of the final classifier. Figure 5.4 illustrates the ROC curves of the BD classifier with variance-based PC selection and that with significant-based PC selection. Due to the limitation of small sample size, the ROC curves of both classifiers are rough. Therefore, we applied a curve fitting technique to fit a smooth curve exponentially. The ROC curve of the BD classifier with significant-based PC selection method is closer to the top left corner than that of the BD classifier with variance-based PC selection. Once having a ROC curve of a classifier, the best performance could be found if we decide the risk of misdetection. In our work, the risk of misdetection was set 1.5 and the risk of false alarm was set 1. Then, the best performance with significant-based PC selection reached a FP rate, 16.5%, and a TP rate, 89.9%, and that with variance-based PC selection reached a FP rate, 24.6%, and a TP rate, 89.9%. It showed that the performance of significant-based classifiers was better than that of variance-based classifiers.

The corresponding PAUC indices are listed in Table 5.4. The particular region is that the FP rate ranges from 0.026 to 0.3 because the value of actual ROC curve of the BD

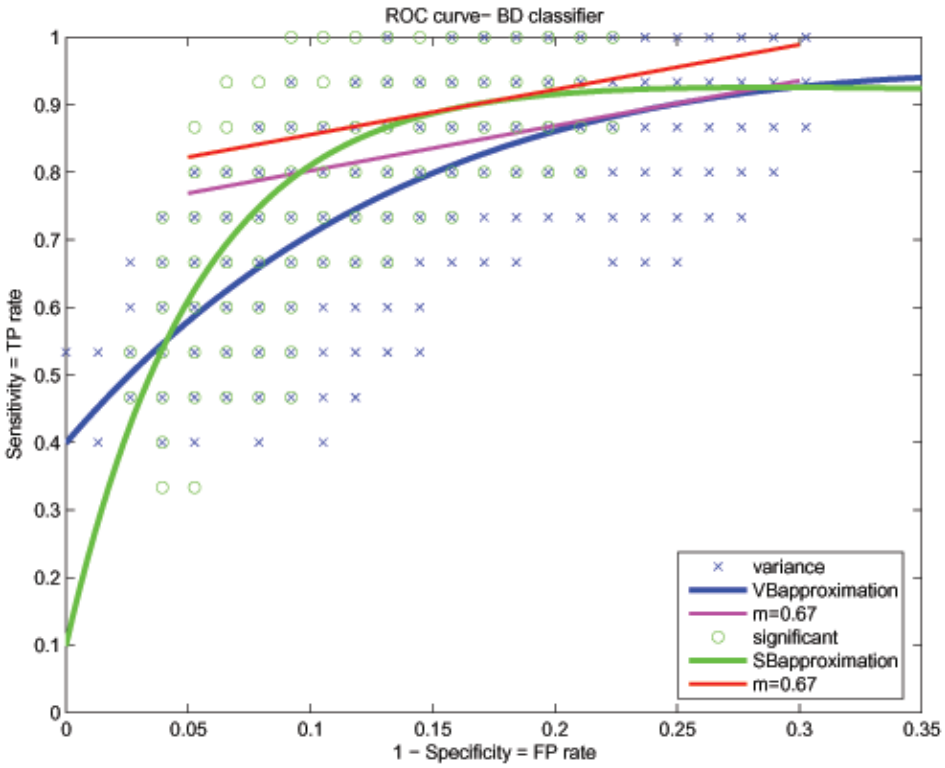


Figure 5.4: **ROC curves of the BD classifier with two PC selection method.** The blue line is the actual curve of the classifier with variance-based PC selection method and a curve fitting, applied on it, leads to the blue curve. The green line is the actual curve of the classifier with significant-based PC selection method and the green curve is the fitting curve. The green curve is closer to the top left corner than the blue curve. Therefore, it seems that using significant-based PC selection is better than using variance-based PC selection. The position where the green curve intersects with the red line represents the best performance of the BD classifier with significant-based PC selection method. The position where the blue curve intersects with the magenta line indicates the best performance of the other classifier.

Table 5.4: **PAUC indices for ROC curves of two BD classifiers.** The partial area was calculated in a specific region where the FP rate ranges from 0.026 to 0.3. As this specific region denotes 1, we have $0 \leq \text{PAUC} \leq 1$. It is clear that PAUC index with significant-based PC selection is greater than that with variance-based PC selection.

Methods	PAUC indices
Variance-based PC selection	0.78366
Significant-based PC selection	0.83617

classifier with significant-based PC selection begins when FP rate equals 0.026. The PAUC index of the BD classifier with significant-based PC selection method is larger than that with variance-based PC selection method and is referred to achieve a better classification performance than the other.

5.4 Cross-group testing

In our thesis, a diagnosis system is proposed with several parallel classification models. For an unknown subject, it is easy to diagnose whether he or she is attacked by a specific disease or not. However, the diagnosis report may result in a complication. That is, the test subject is informed that he or she may sicken with many diseases. Although it might certainly happen that a subject suffers from various illnesses at the same time, a subject usually has only a particular disorder at a time. Thus, if a test subject is truly a patient with disease A and he is told that he has both disorders, disease A and disease B, with probabilities P% and Q% respectively, it makes reasonable that the probability P% should be higher than the probability Q%.

In this work, we have three data sets: a healthy control set, a BD set and a SCA3 set. All bipolar disorder patients were diagnosed to suffer from BD by doctors, but we have no

Table 5.5: **Classification of SCA3 patients on the BD classifier with variance-based PC selection method.** This table shows predictions of patients carrying SCA3 by the BD classifier and by the SCA3 classifier with variance-based PC selection method. Columns from two to four indicate individual predictions of the GM, WM and CSF classifiers which compose of the BD classifier. The column, Result (BD), presents the final prediction, the maximum of columns from two to four, by the BD classifier. The most right column reveals predictions of them by using the SCA3 classifier.

SCA3 patients	GM prediction (%)	WM prediction (%)	CSF prediction (%)	Result (BD) (%)	Result (SCA3) (%)
1	0.999968	0.381076	0.999808	0.999968	0.999995
2	0.999983	0.921121	0.996982	0.999983	1
3	0.93203	0.988987	0.273474	0.988987	1
4	0.999579	0.666803	0.234573	0.999579	1
5	0.796025	0.197827	0.040213	0.796025	0.782884
6	0.990275	0.175605	0.396005	0.990275	0.99247

idea whether they suffer from SCA3 or not. So do SCA3 patients. Therefore, we did an experiment for cross-group testing. We expected that, for all SCA3 patients, the possibility of belonging to the SCA3 group is larger than that to the BD group and the possibility of being in the BD group is higher than that in the SCA3 group for all BD patients.

Table 5.5 and Table 5.6 show the prediction results of SCA3 patients by the BD classifier with variance-based PC selection method and with significant-based PC selection method respectively. With variance-based PC selection method, all patients carrying SCA3 have almost the same possibility of sickening with bipolar disorder and with SCA3. It seems that the distributions of the BD group and the SCA3 group are hard to differentiate from each other in our classification model. On the other hand, using significant-based PC selection method, the probability of taking bipolar disorder is lower than that of suffering SCA3 for each SCA3 patient. This phenomenon conforms to our expectation and seems that both distributions of the BD group and the SCA3 group could be separate in this classification model.

Table 5.7 shows the prediction results of patients with bipolar disorder by the SCA3

Table 5.6: **Classification of SCA3 patients on the BD classifier with significant-based PC selection method.**This table shows predictions of patients carrying SCA3 by the BD classifier and by the SCA3 classifier with significant-based PC selection method. Columns from two to four indicate individual predictions of the GM, WM and CSF classifiers which compose of the BD classifier. The column, Result (BD), presents the final prediction, the maximum of columns from two to four, by the BD classifier. The most right column reveals predictions of them by using the SCA3 classifier.

SCA3 patients	GM prediction (%)	WM prediction (%)	CSF prediction (%)	Result (BD) (%)	Result (SCA3) (%)
1	0.642423	0.000012	0.700042	0.700042	1
2	0.677553	0.000158	0.966791	0.966791	1
3	0.850826	0.170933	0.516088	0.850826	1
4	0.757999	0.006987	0.591706	0.757999	1
5	0.012332	0.001971	0.322434	0.322434	0.999932
6	0.21259	0.036317	0.797595	0.797595	1

classifier with variance-based PC selection method. Using variance-based PC selection method, most of BD patients are predicted to have low possibility to sickening with SCA3 but there is a BD patient, numbered five, who may have SCA3 in a high possibility, 98.9%. This result seems that the BD group and the SCA3 group should be easy to separate from each other in this classification model. However, it contradicts with the results of testing patients carrying SCA3 by the BD classifier. Thus, the prediction results from a classification model with variance-based PC selection method look inconsistent.

Table 5.8 displays the prediction results of patients with bipolar disorder by the SCA3 classifier with significant-based PC selection method. With significant-based PC selection method, all patients suffering bipolar disorder are predicted that they have low probability of being attacked by SCA3. Also, it looks that the distributions of the BD group and the SCA3 group could be differentiated from each other in this classification model and seems more consistent with the prediction results of testing SCA3 patients by the BD classifier. As a result, we might use a cartoon-like representation of three sets, shown in Figure 5.5, to exhibit the relationship of all three groups in a hypothetic classification space although

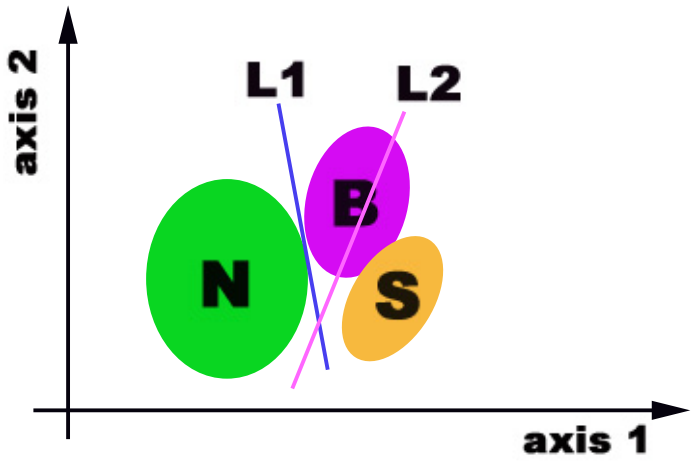


Figure 5.5: **Cartoon-like representation of classification of three groups.** The normal group, BD group and SCA3 group are represented by the green, the purple and the orange ellipses written with N, B and S individually. Assume that the line, L1, represents a predicted boundary between the normal group and the BD group found by the BD classifier, and the line, L2, represents a predicted boundary between the normal group and the SCA3 group found by the SCA3 classifier. It is reasonable that SCA3 patients would be predicted to sicken with bipolar disorder by the BD classifier constructed in our work.

we proposed a separate classification space for a particular disease in our work. Since the BD group may be in the middle between the normal group and the SCA3 group, it makes sense that SCA3 patients may be predicted to taking bipolar disorder by the BD classifier although they may not suffer from it actually. Fortunately, a probabilistic prediction result could tell how close a test subject is to a specific group and make the system worthful.

Finally, comparing the results of cross-group testing with two different PC selection method, a classifier constructed with a significant-based PC selection method makes a more accurate and consistent diagnosis than that with a variance-based PC selection method. Thus, using the significant-based PC selection method to construct a classifier will indeed improve the performance of a classification model.

Table 5.7: **Classification of BD patients on the SCA3 classifier with variance-based PC selection method.**This table shows predictions of patients suffering BD by the SCA3 classifier and by the BD classifier with variance-based PC selection method. Columns from two to three indicate individual predictions of the GM and WM classifiers which compose of the SCA3 classifier. The column, Result (SCA3), presents the final prediction, the maximum of columns from two to four, by the SCA3 classifier. The most right column reveals predictions of them by using the BD classifier.

BD patients	GM prediction (%)	WM prediction (%)	Result (SCA3) (%)	Result (BD) (%)
1	0	0	0	0.997847
2	0	0	0	0.999999
3	0	0	0	0.996431
4	0	0	0	0.999994
5	0	0.989137	0.989137	1.000000
6	0	0	0	0.999999
7	0	0.000005	0.000005	0.744795
8	0	0	0	0.983158
9	0	0	0	0.999999
10	0	0	0	0.808894
11	0	0	0	0.712419
12	0	0	0	0.987404
13	0	0	0	0.999999
14	0	0	0	1.000000
15	0	0	0	0.999996

Table 5.8: **Classification of BD patients on the SCA3 classifier with significant-based PC selection method.**This table shows predictions of patients suffering BD by the SCA3 classifier and by the BD classifier with significant-based PC selection method. Columns from two to three indicate individual predictions of the GM and WM classifiers which compose of the SCA3 classifier. The column, Result (SCA3), presents the final prediction, the maximum of columns from two to four, by the SCA3 classifier. The most right column reveals predictions of them by using the BD classifier.

BD patients	GM prediction (%)	WM prediction (%)	Result (SCA3) (%)	Result (BD) (%)
1	0	0.032993	0.032993	0.951285
2	0	0.000008	0.000008	0.999999
3	0	0.000108	0.000108	0.997940
4	0	0	0	0.999852
5	0	0	0	1.000000
6	0	0.000002	0.000002	0.996268
7	0	0.000002	0.000002	0.826220
8	0	0	0	0.638330
9	0	0	0	1.000000
10	0	0	0	0.978042
11	0	0.000036	0.000036	0.951610
12	0	0.000009	0.000009	0.985007
13	0	0	0	0.993619
14	0	0	0	1.000000
15	0	0	0	0.999919