Chapter 6

Classification results

6.1 Overview

The classification process comprises three important points that have to be taken care of: the implementation, the data and the training-testing scheme.

6.1.1 Implementation

The implementation chosen was the LibSVM [5]. It is an integrated software for SVM classification, regression and distribution estimation. It also supports multi-class classification and it provides a MATLAB-based implementation which makes it perfect for our purpose.

Two functions of the LibSVM implementation are used: *svmtrain* and *svmpredict*. The first trains the classifier and builds the model based on the training examples. The second one performs the validation of the model based on the testing examples, i.e., given a new data example not seen in the training step and its ground truth label (depressed or non depressed), it classifies it based on the trained classifier and validates according to the label, returning thus information about the goodness of the classification process.

6.1.2 Data for classification

Choosing the right data is essential to classification success. For our purpose, which is developing a first approach to an automatic depression assessment technique, a simple start point is to choose the most obvious depressed and non depressed examples, which are going to be the easiest to learn and to classify by the SVM. Applying the criterion of depression severity in Section 2.2, which stated that the

depressed sessions are those with a Hamilton score higher or equal to 15 and the non depressed sessions are those with a Hamilton score less or equal to 7, we can choose the most obvious examples of depressive and non-depressive behavior.

Once this criterion is applied to the Spectrum database, 66 depressed sessions and 41 non depressed sessions are kept for classification. These sessions belong to a total of 51 subjects.

Attending to the type of features used, two kinds of experiments are performed:

- **Landmark points as features:** In these experiments the landmarks points of the face will be the start point of the feature extraction process. Not all the 66 landmarks are going to be used, but only the ones of higher interest: eyebrows, eyes and mouth. The remaining points are considered too static to be useful for depression assessment purposes.

- **Shape coefficients as features:** Using the shape coefficients allows us to take into account all the landmark points of the face as a whole (see Section 5.3) and they take part in another set of experiments.

6.1.3 Training-testing algorithm

Two different training-testing algorithms were applied in the experiments. Both algorithms are based in a K-fold cross validation method (see Figure 6.1 and Figure 6.2). In the first one, K is chosen to be equal to the total number of subjects (51), turning out to be a leave-one-subject-out scheme, where the sessions of one subject are left out for testing and the rest are used for training the classifier. The second method relies on taking a certain percentage of the data for training in each fold and using the rest for the validation.

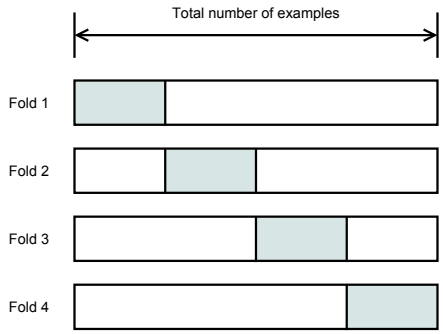

Figure 6.1: 4-fold cross validation method. The testing set is shown in grey.

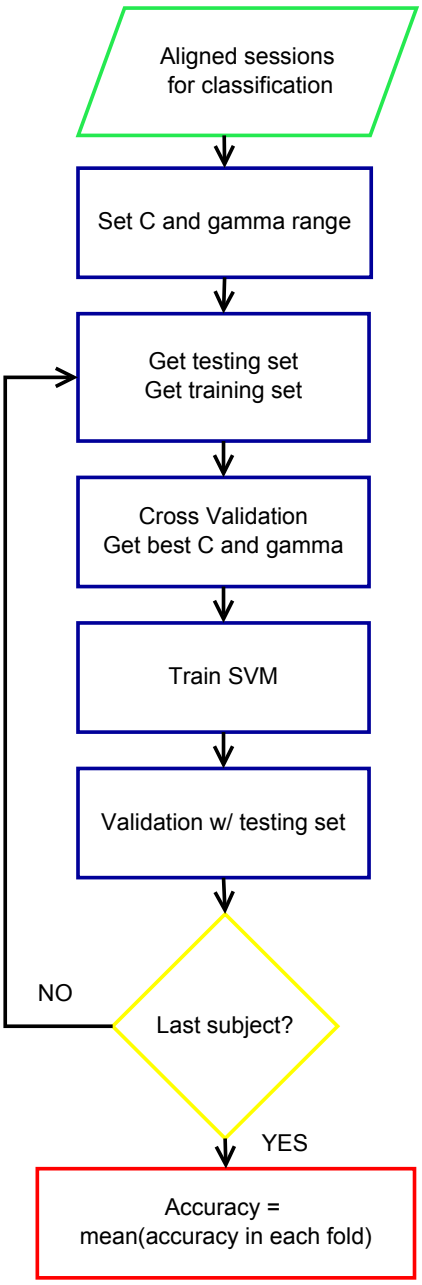

Figure 6.2: Flow chart of the general training-testing scheme used for the experiments of Sections 6.2 and 6.3.

6.2 Leave-one-subject-out results

6.2.1 Overview

As explained in the introduction of this chapter, the leave-one-subject-out algorithm relies on a 51-fold cross validation method, where in each fold the sessions of one subject are used for the validation step, while the remaining sessions train the SVM. Employing a classical leave-one-subject out technique would result in including in both the training and testing step sessions from the same subject (with no doubt highly correlated) and in a better performance of the classifier. However, the goal of these experiments is being able to classify depressed and non-depressed sessions from completely "unseen" subjects and that is the reason why the leave-one-out method was discarded.

The results are presented in a table, in which the performance of the classifier is summarized by the following parameters (see Section 4.3 for their definitions):

- **Acc.** = Accuracy.
- **TPR** = True Positive Rate.
- **TNR** = True Negative Rate.
- **F** = F parameter.
- **Area ROC** = Area under the ROC curve (ideally 1).

The experiments focus on specific parts of the face (eyebrows, eyes, mouth) and on the shape coefficients. Furthermore, the classification results using all possible combination of these features, except the shape coefficients, are presented. The shape coefficients are excluded in these combinations in order to avoid redundant information during classification, since they are a representation of the behavior of the whole face.

All the experiments using single features are repeated using different lengths of intervals in the group level (150, 300 and 600 frames). The experiments using combination of features are only performed with 300-frames intervals, due to the better performance of this interval length in the experiments using single features. This interval length represents the number of frames of the video sequence that are grouped together before obtaining the final vector of features.

6.2.2 Eyes

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.498	0.606	0.390	0.547	0.530
300 frames	Mean Median Std	Mean Median Min Max	0.671	0.758	0.585	0.698	0.701
600 frames	Mean Median Std	Mean Median Min Max	0.652	0.742	0.561	0.681	0.707

Table 6.1: Classification results for the eyes.

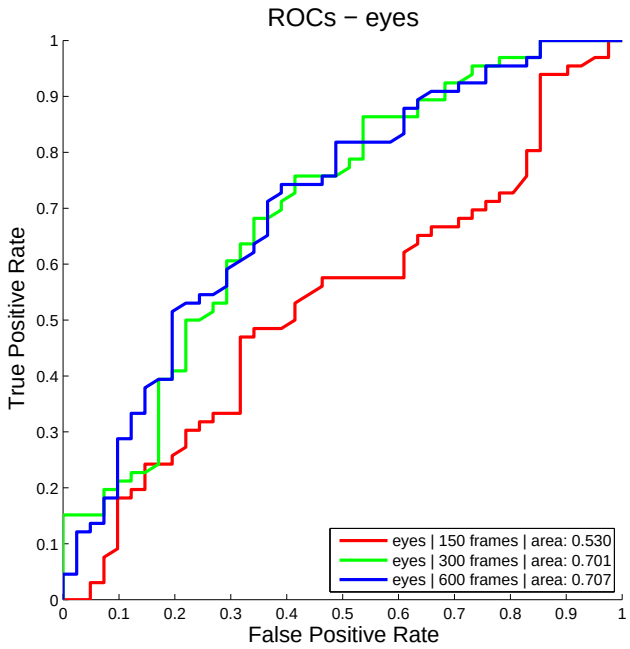

Figure 6.3: ROC curves for the classification using the eyes and 150, 300 and 600 frames intervals in the global level.

6.2.3 Eyebrows

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.626	0.667	0.585	0.641	0.645
300 frames	Mean Median Std	Mean Median Min Max	0.666	0.697	0.634	0.676	0.752
600 frames	Mean Median Std	Mean Median Min Max	0.632	0.727	0.537	0.664	0.695

Table 6.2: Classification results for the eyebrows.

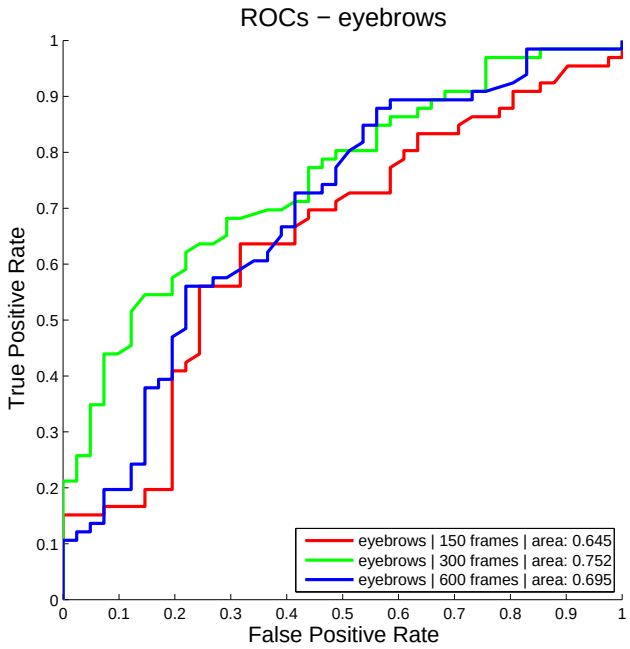

Figure 6.4: ROC curves for the classification using the eyebrows and 150, 300 and 600 frames intervals in the group level.

6.2.4 Mouth

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.611	0.636	0.585	0.621	0.582
300 frames	Mean Median Std	Mean Median Min Max	0.749	0.742	0.756	0.748	0.779
600 frames	Mean Median Std	Mean Median Min Max	0.592	0.697	0.488	0.631	0.632

Table 6.3: Classification results for the mouth.

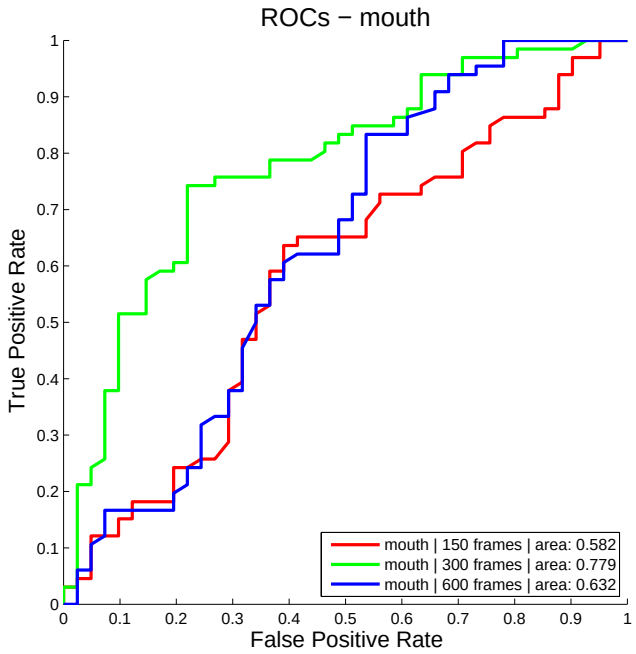

Figure 6.5: ROC curves for the classification using the mouth and 150, 300 and 600 frames intervals in the group level.

6.2.5 Shape coefficients

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.583	0.727	0.439	0.636	0.664
300 frames	Mean Median Std	Mean Median Min Max	0.761	0.864	0.659	0.783	0.794
600 frames	Mean Median Std	Mean Median Min Max	0.712	0.864	0.561	0.750	0.746

Table 6.4: Classification results for the shape coefficients.

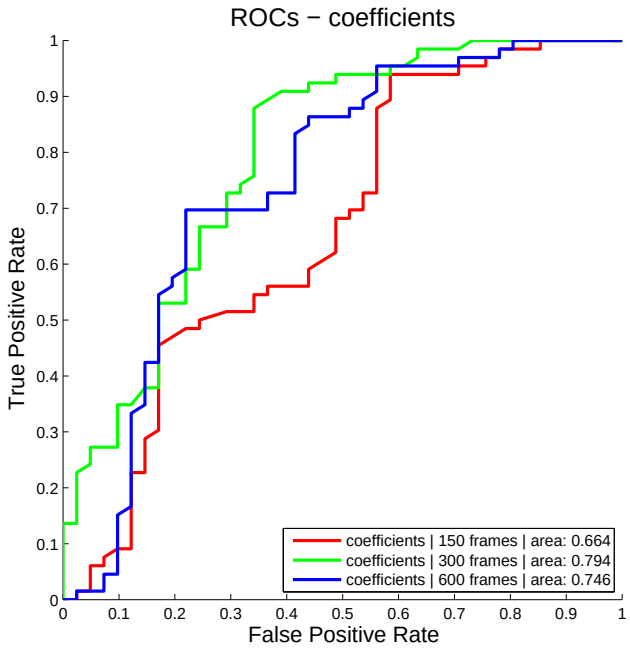

Figure 6.6: ROC curves for the classification using the shape coefficients and 150, 300 and 600 frames intervals in the group level.

6.2.6 Combination of features

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
Eyebrows + mouth							
300 frames	Mean Median Std	Mean Median Min Max	0.650	0.667	0.634	0.656	0.746
Eyebrows + eyes							
300 frames	Mean Median Std	Mean Median Min Max	0.560	0.682	0.439	0.608	0.662
Eyebrows + eyes + mouth							
300 frames	Mean Median Std	Mean Median Min Max	0.691	0.773	0.610	0.715	0.758
Eyes + mouth							
300 frames	Mean Median Std	Mean Median Min Max	0.722	0.833	0.610	0.750	0.747

Table 6.5: Classification results for all possible combinations of eyebrows, eyes and mouth features.

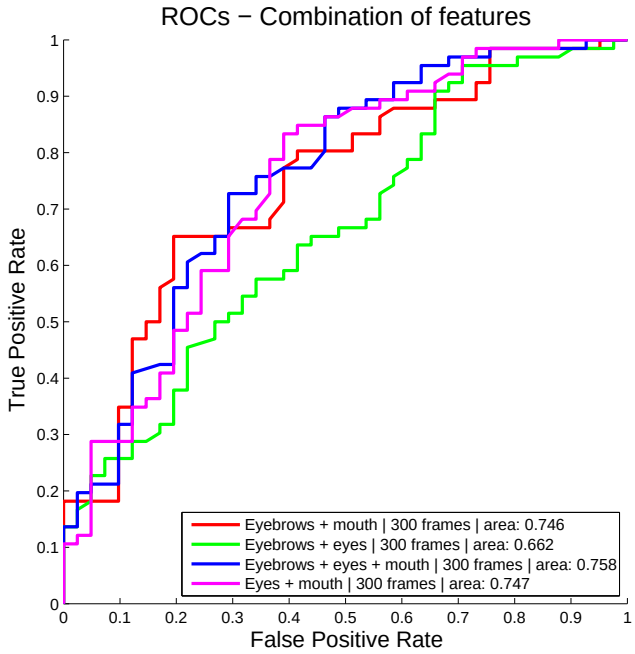

Figure 6.7: ROC curves for the classification using all possible combinations of features and 300-frames intervals in the group level.

6.2.7 Summary and discussion

As a summary, Table 6.6 and Table 6.7 show the accuracy of each of the experiment presented in the previous section. The discussion on the results will be based on this measure.

Feature \ Interval	150 frames	300 frames	600 frames
Eyes	0.498	0.671	0.652
Eyebrows	0.626	0.666	0.632
Mouth	0.611	0.749	0.592
Shape coefficients	0.583	0.761	0.712

Table 6.6: Accuracy of the classification using single features.

Among all the interval lengths chosen for the experiments (150, 300 and 600 frames, corresponding to 5, 10 and 15 seconds, respectively), the 300-frames interval shows the best results. This is more noticeable in the case of the shape coefficients

Features / Interval	Eyebrows Eyes	Mouth Eyes	Mouth Eyebrows	Eyes Eyebrows Mouth
300 frames	0.560	0.722	0.650	0.691

Table 6.7: Accuracy of the classification using combinations of features.

features, which greatly improve the classification performance with 300 frames in comparison with the 150 and 600 frames intervals.

The eyes and the eyebrows do not seem to be determinant to discriminate between depressed and non-depressed subjects, since the classification results for these features do not exceed the 68% of accuracy. However, the same cannot be said about the mouth and the shape coefficients. These features are able to achieve the 75% of accuracy in the classification. A successful result was expectable in the case of the shape coefficients if we take into account that they are meant to be a representation of the whole face and thus carry a wealth of information of it, but not in the case of the mouth. The mouth itself is able to discriminate with a reasonable accuracy between the two classes. If we carefully think about it, the mouth can express many different feelings: happiness – by a smile –, sadness – by depressing the lip corners –, surprise – by a jaw drop –, etc. Furthermore, it also carries the information of the speech, both the velocity and the amount of time the person speaks.

The experiments using combinations of features were only performed using 300-frames intervals in the group level, due to the better results this interval length has shown in those using single features in comparison with 150 and 600 frames. Analyzing the accuracy of these experiments we can conclude that none of them give satisfactory results, due to different factors. The performance of *Eyebrows+Eyes* is close to randomness, which is expectable if we take into account that we are combining two features that used as single features for classification give poor results. Regarding the rest of the examples (*Eyebrows+Mouth*, *Mouth+Eyes* and *Mouth+Eyes+Eyebrows*), all of them include the mouth, but in combination with less discriminative features, resulting in a lower performance than using just the mouth itself as a feature.

In conclusion, the best result obtained by the leave-one-subject-out scheme is achieved using the shape coefficients as features and 300-frames interval length in the group level of the feature extraction process.

6.3 75% training - 25% testing

6.3.1 Overview

In the following pages, the results of the classification experiments using a 75% of the data for the training and 25% for the validation (it can be seen as a 4-fold) are presented. Strictly speaking, the percentage of data used for the training-testing is just approximate. Again, as it was done discarding the classical leave-one-out method, for the training and testing set to be completely uncorrelated, sessions of the same subject cannot be contained in both sets. This fact was taking into account when the two sets were built and therefore not the same number of sessions are present in the training and testing sets in each iteration of the algorithm (note that there are not the same number of sessions for each of the subjects).

The results are presented in a table, in which the performance of the classifier is summarized by the following parameters (see Section 4.3 for their definitions):

- **Acc.** = Accuracy.
- **TPR** = True Positive Rate.
- **TNR** = True Negative Rate.
- **F** = F parameter.
- **Area ROC** = Area under the ROC curve (ideally 1).

The experiments focus on specific parts of the face (eyebrows, eyes, mouth) and on the shape coefficients. Furthermore, the classification results using all possible combination of these features, except the shape coefficients, are presented. The shape coefficients are excluded in these combinations in order to avoid redundant information during classification, since they are a representation of the behavior of the whole face.

All the experiments using single features are repeated using different lengths of intervals in the group level (150, 300 and 600 frames). The experiments using combination of features are only performed with 300-frames intervals, due to the better performance of this interval length in the experiments using single features. This interval length represents the number of frames of the video sequence that are grouped together before obtaining the final vector of features.

6.3.2 Eyes

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.540	0.627	0.452	0.577	0.570
300 frames	Mean Median Std	Mean Median Min Max	0.619	0.761	0.476	0.666	0.686
600 frames	Mean Median Std	Mean Median Min Max	0.584	0.716	0.452	0.633	0.618

Table 6.8: Classification results for the eyes.

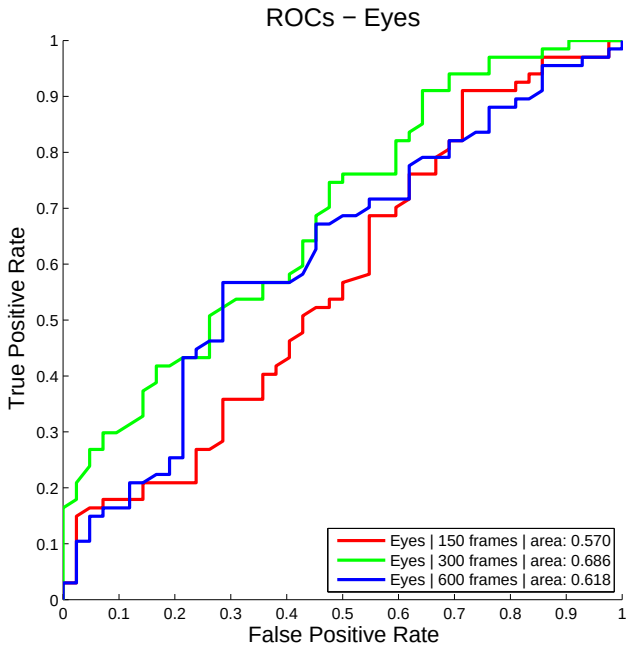

Figure 6.8: ROC curves for the classification using the eyes and 150, 300 and 600 frames intervals in the group level.

6.3.3 Eyebrows

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.541	0.582	0.500	0.559	0.566
300 frames	Mean Median Std	Mean Median Min Max	0.608	0.716	0.500	0.646	0.688
600 frames	Mean Median Std	Mean Median Min Max	0.589	0.701	0.476	0.630	0.623

Table 6.9: Classification results for the eyebrows.

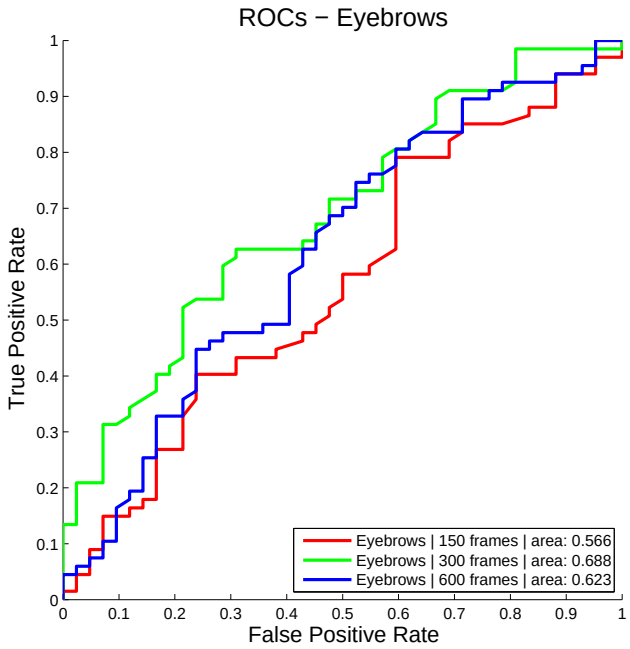

Figure 6.9: ROC curves for the classification using the eyebrows and 150, 300 and 600 frames intervals in the group level.

6.3.4 Mouth

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.583	0.761	0.405	0.646	0.630
300 frames	Mean Median Std	Mean Median Min Max	0.696	0.701	0.690	0.698	0.739
600 frames	Mean Median Std	Mean Median Min Max	0.601	0.701	0.500	0.637	0.658

Table 6.10: Classification results for the mouth.

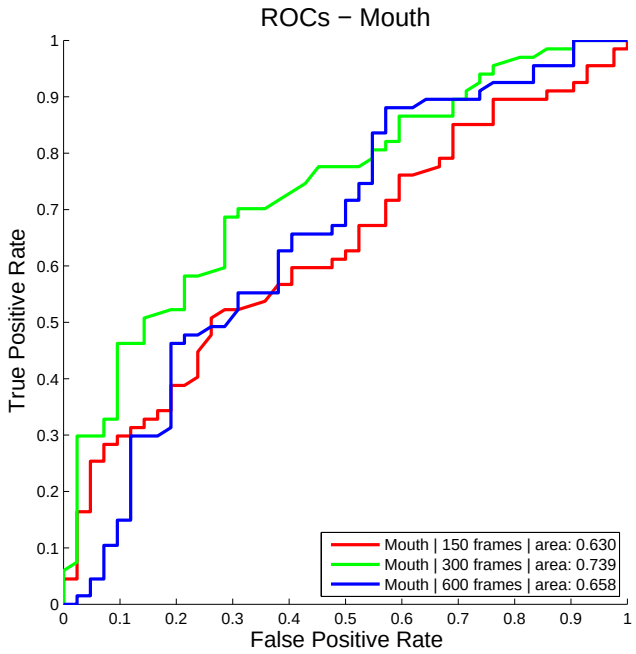

Figure 6.10: ROC curves for the classification using the mouth and 150, 300 and 600 frames intervals in the group level.

6.3.5 Shape coefficients

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
150 frames	Mean Median Std	Mean Median Min Max	0.659	0.866	0.452	0.717	0.679
300 frames	Mean Median Std	Mean Median Min Max	0.713	0.806	0.619	0.737	0.732
600 frames	Mean Median Std	Mean Median Min Max	0.663	0.851	0.476	0.717	0.669

Table 6.11: Classification results for the shape coefficients.

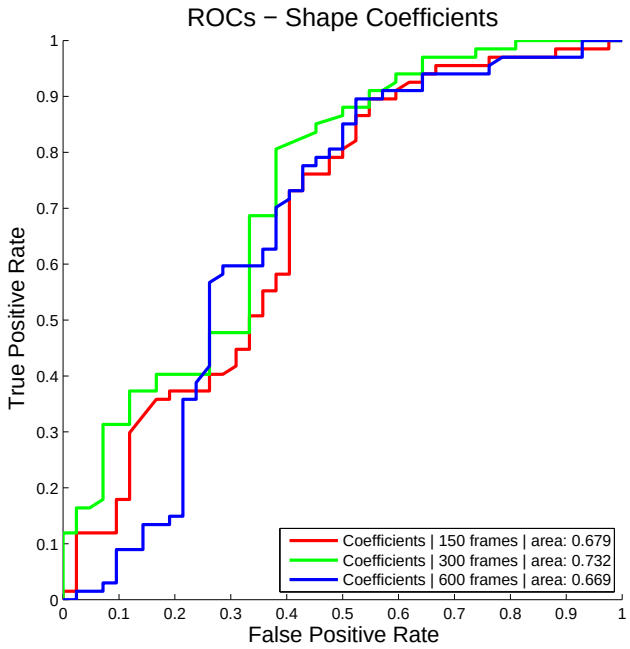

Figure 6.11: ROC curves for the classification using the shape coefficients and 150, 300 and 600 frames intervals in the group level.

6.3.6 Combination of features

Interval Length	Group Level	Global Level	Acc.	TPR	TNR	F	Area ROC
Eyebrows + mouth							
300 frames	Mean Median Std	Mean Median Min Max	0.692	0.716	0.667	0.699	0.770
Eyebrows + eyes							
300 frames	Mean Median Std	Mean Median Min Max	0.653	0.806	0.500	0.699	0.780
Eyebrows + eyes + mouth							
300 frames	Mean Median Std	Mean Median Min Max	0.663	0.731	0.595	0.685	0.757
Eyes + mouth							
300 frames	Mean Median Std	Mean Median Min Max	0.696	0.821	0.571	0.730	0.740

Table 6.12: Classification results for all possible combinations of eyebrows, eyes and mouth features.

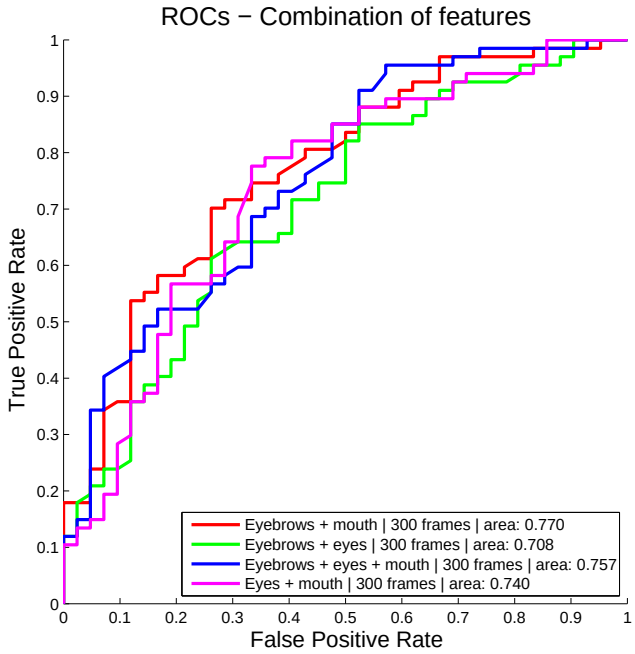

Figure 6.12: ROC curves for the classification using all possible combinations of features and 300-frames intervals in the group level.

6.3.7 Summary and discussion

As a summary, Table 6.13 and Table 6.14 show the accuracy of each of the experiment presented in the previous section. The discussion on the results will be based on this measure.

Feature \ Interval	150 frames	300 frames	600 frames
Eyes	0.540	0.619	0.584
Eyebrows	0.541	0.608	0.589
Mouth	0.583	0.696	0.601
Shape coefficients	0.659	0.713	0.663

Table 6.13: Accuracy of the classification using single features.

As we can easily see comparing the tables of this section with those of Section 6.2 (Tables 6.6 and 6.7), the overall results are worse when using a 75% - 25% training-testing scheme than with the leave-one-subject-out method. This fact is

Features / Interval	Eyebrows Eyes	Mouth Eyes	Mouth Eyebrows	Eyes Eyebrows Mouth
300 frames	0.653	0.696	0.692	0.663

Table 6.14: Accuracy of the classification using combinations of features.

easy to understand if we think that it is easier for the SVM to correctly classify one, two or at most, three sessions (depending on the subject we are leaving out) than classifying one quarter of the sessions (around 25). To this we have to sum up the fact that during the training step the classifier has less examples to learn from, making its task even harder. However, in general, the performance does not greatly decrease, which gives robustness to our classifier.

The relative performance (taking just into account the 25% - 75% results) of the different features for classification is very similar to what was found with the leave-one-subject-out algorithm. The best results are given by the shape coefficients and the mouth and 300-frames segmentation, this time with an accuracy of around 70% (5% of accuracy loss). The poor performance of the eyes and the eyebrows reinforce the theory that these two features are not determinant to distinguish between the depressed and the non-depressed class.

Regarding the results using combination of features, in this case the combination of the mouth with non-discriminant features, such as the eyes and the eyebrows, does not seem to affect as much as in the case of the leave-one-subject-out scheme, since the results do not differ much from the ones obtained using the mouth as a single feature.

In conclusion, the best result obtained by the 75% - 25% testing algorithm is achieved using the shape coefficients as features and 300-frames interval length in the group level of the feature extraction process.