

Facial Action Unit Detection Using Kernel Partial Least Squares - Supplemental Material

Tobias Gehrig and Hazım Kemal Ekenel
Facial Image Processing and Analysis Group, Institute for Anthropomatics
Karlsruhe Institute of Technology
D-76131 Karlsruhe, P.O. Box 6980 Germany
{tobias.gehrig, ekenel}@kit.edu

1. Introduction

In this document we present additional results corresponding to the experiments shown in [1].

A. ROC Curves

The ROC curves for the AU estimates are shown in this section.

A.1. Evaluation on a Single Dataset

A.1.1 Experiment on the CK+ Dataset with Eye Labels

See Figure 1.

A.1.2 Experiment on the CK+ Dataset with Automatic Eye Detection

See Figure 2.

A.1.3 Experiment on the GEMEP-FERA Dataset

See Figure 3.

A.2. Evaluation across Datasets

A.2.1 Generalization from Constrained to less Constrained Condition

See Figure 4.

A.2.2 Generalization from less Constrained to Constrained Condition

See Figure 5.

B. F1-Score

The F1-Scores for the AU estimates are shown in this section. If no threshold optimization is performed then the thresholds are set to 0.5 for the PLS-based approaches and

Table 1. F1 scores in % on CK+ using eye labels. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF PLS	linear SVM	RBF SVM
1	176	78.1	77.5	69.6	71.5
2	117	80.4	76.2	78.9	76.7
4	193	74.2	75.9	72.8	68.0
5	102	77.5	76.2	74.3	73.8
6	123	72.8	68.2	67.0	65.7
7	120	64.0	51.0	51.9	42.3
9	75	84.3	84.2	84.5	83.0
11	34	15.0	5.7	14.6	0.0
12	131	84.7	81.9	78.3	80.0
15	94	60.3	51.5	52.6	49.6
17	201	77.4	78.3	73.6	76.8
20	79	64.8	57.1	49.6	28.0
23	60	35.2	28.6	28.9	14.3
24	58	38.2	26.7	14.1	9.0
25	324	85.4	86.5	86.5	86.1
26	50	15.6	7.4	5.9	0.0
27	81	85.9	83.0	84.6	77.7
AVG		72.3	69.5	67.4	64.4

0.0 for the SVM-based approaches. Otherwise thresholds are optimized using *equal error rate* (EER) or *F1 score* as metrics [2] on either the training folds of the LOSO scheme or the whole training data in case of the cross-dataset tests.

B.1. Evaluation on a Single Dataset

B.1.1 Experiment on the CK+ Dataset with Eye Labels

See Table 1 for F1 scores without threshold optimization, Table 2 for F1 scores using threshold optimization based on EER and Table 3 for F1 scores using threshold optimization based on F1 score.

B.1.2 Experiment on the CK+ Dataset with Automatic Eye Detection

See Table 4 for F1 scores without threshold optimization, Table 5 for F1 scores using threshold optimization based on EER and Table 6 for F1 scores using threshold optimization based on F1 score.

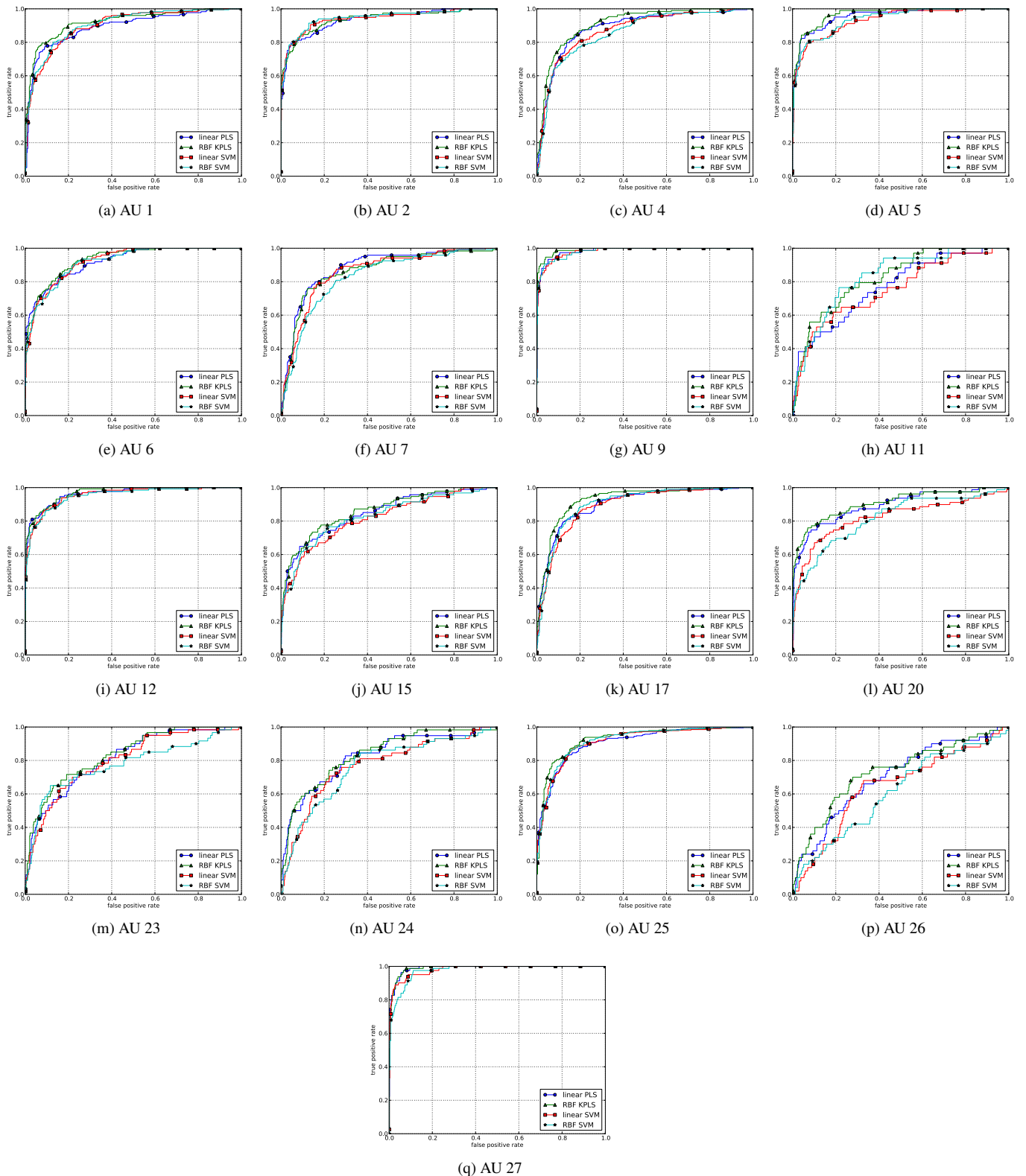


Figure 1. ROC curves on CK+ using eye labels.

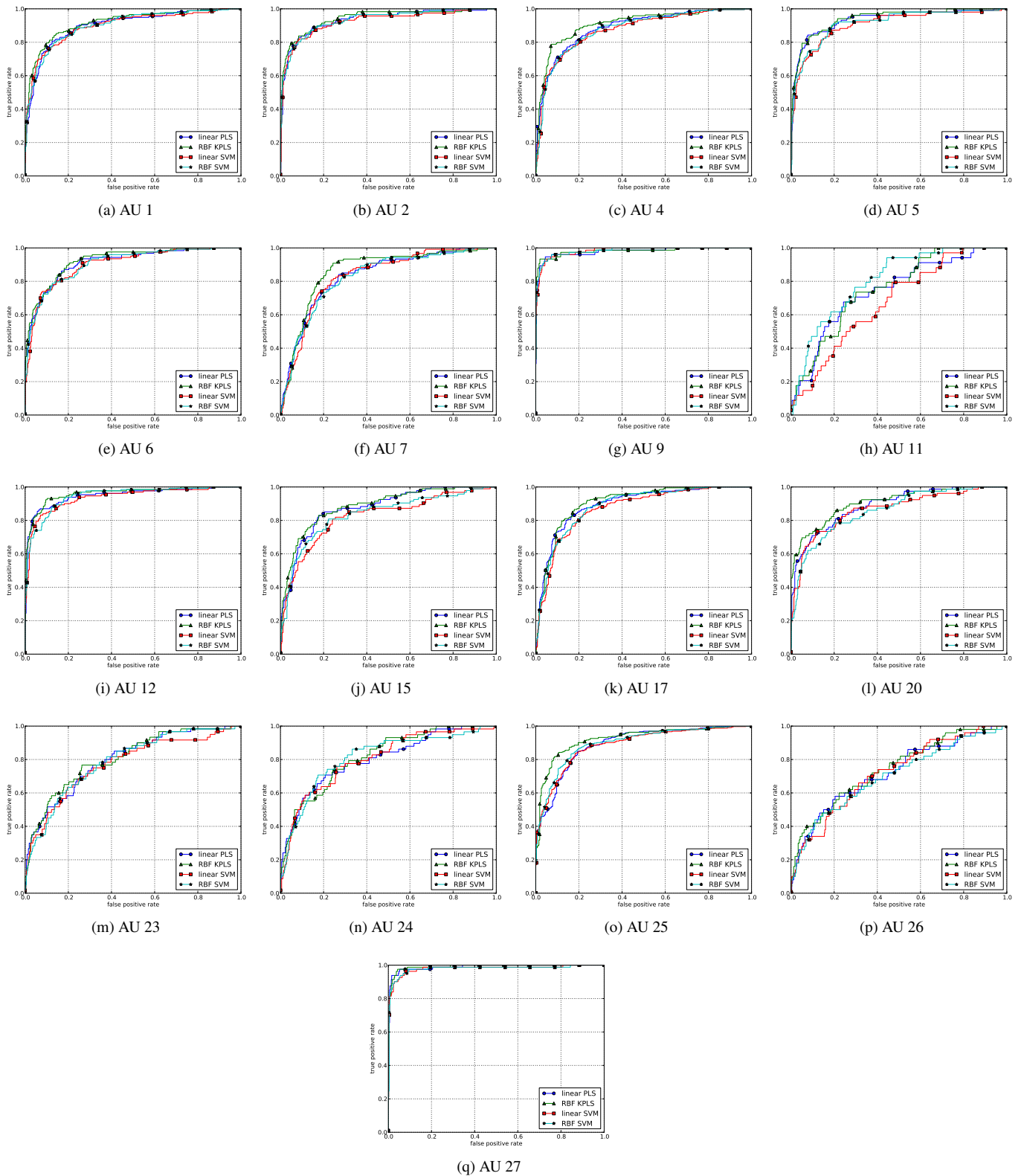


Figure 2. ROC curves on CK+ using automatic eye detection.

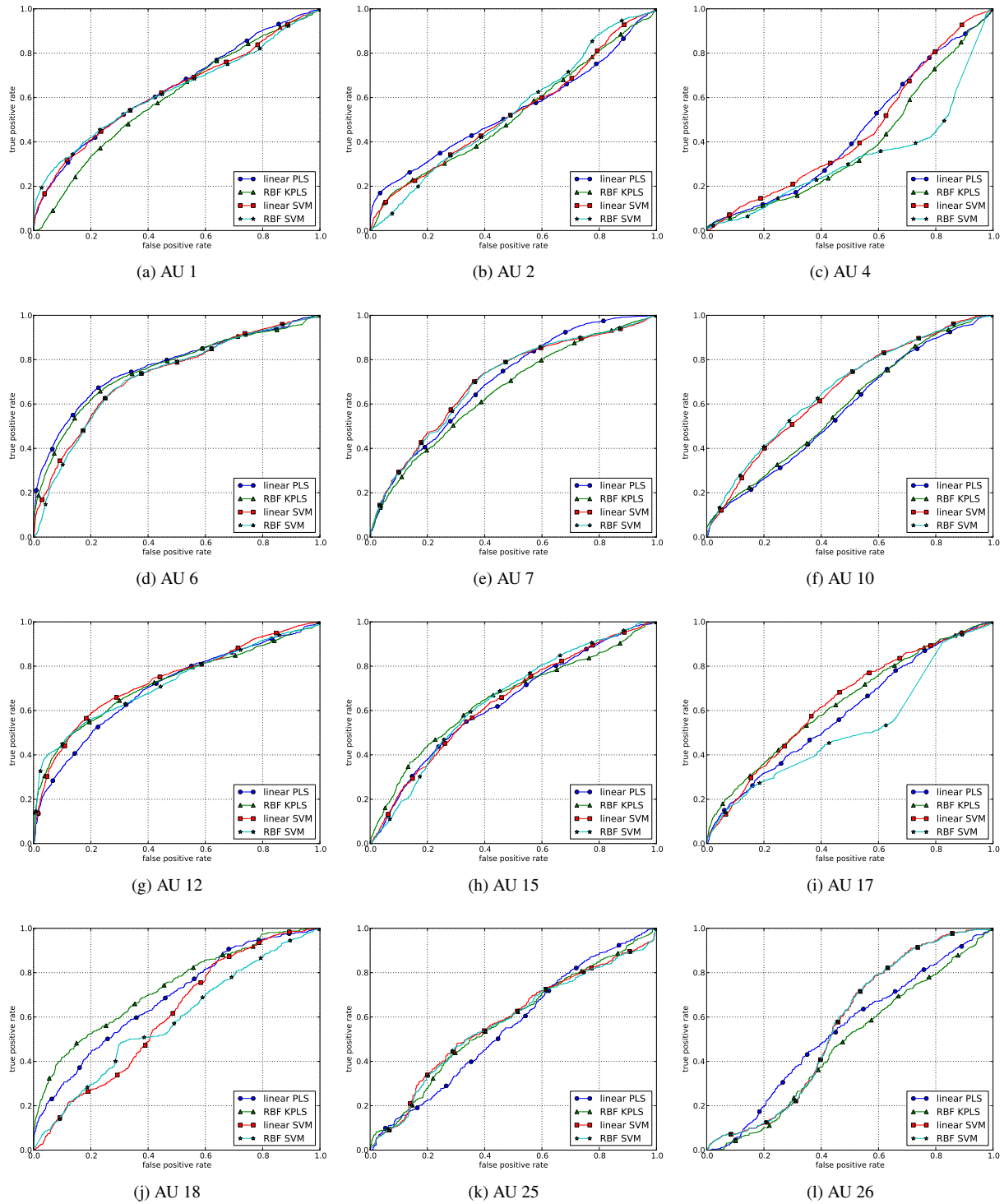


Figure 3. ROC curves on GEMEP-FERA.

B.1.3 Experiment on the GEMEP-FERA Dataset

See Table 7 for F1 scores without threshold optimization, Table 8 for F1 scores using threshold optimization based on

EER and Table 9 for F1 scores using threshold optimization based on F1 score.

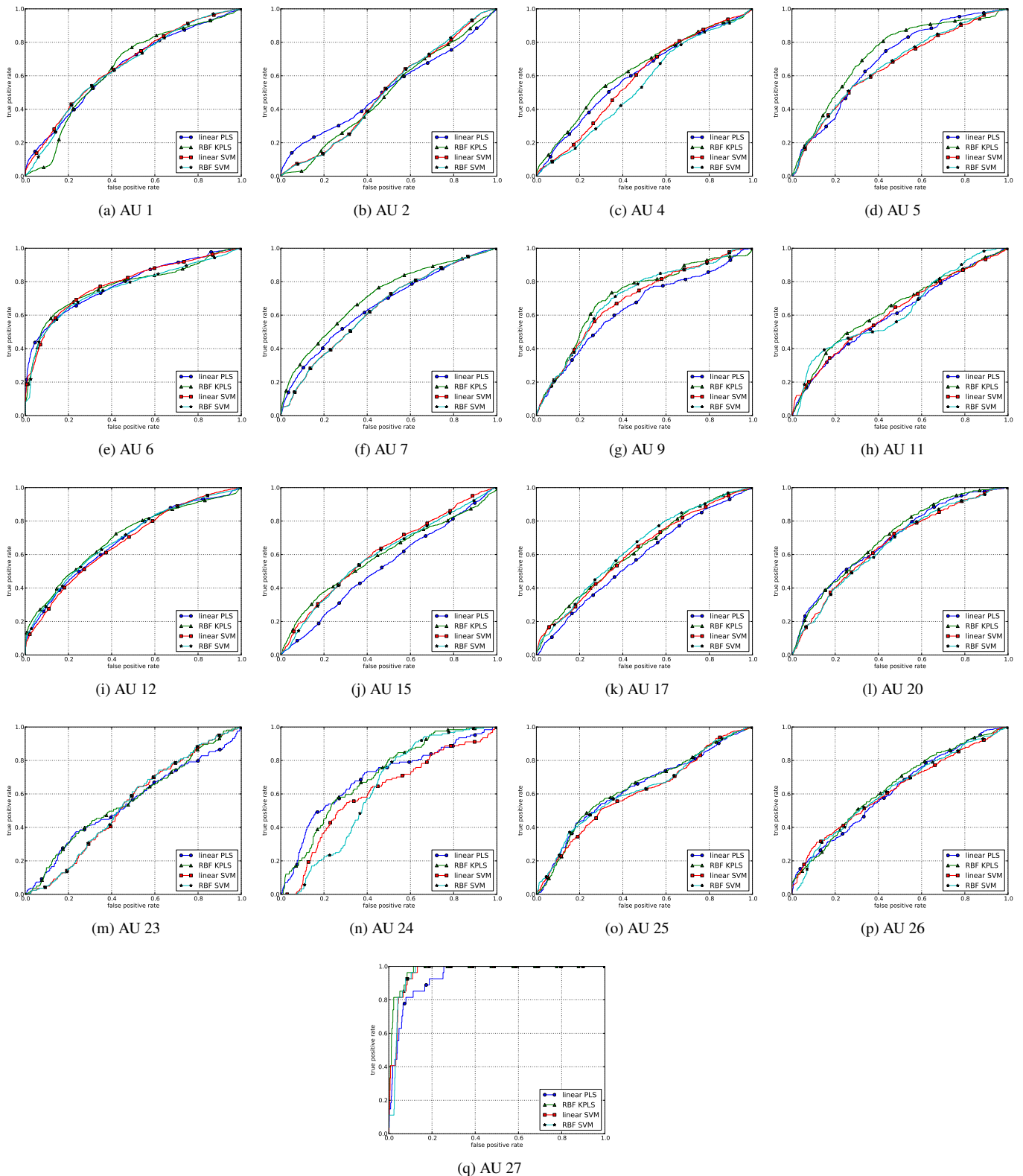


Figure 4. ROC curves in on GEMEP-FERA using classifiers trained on CK+.

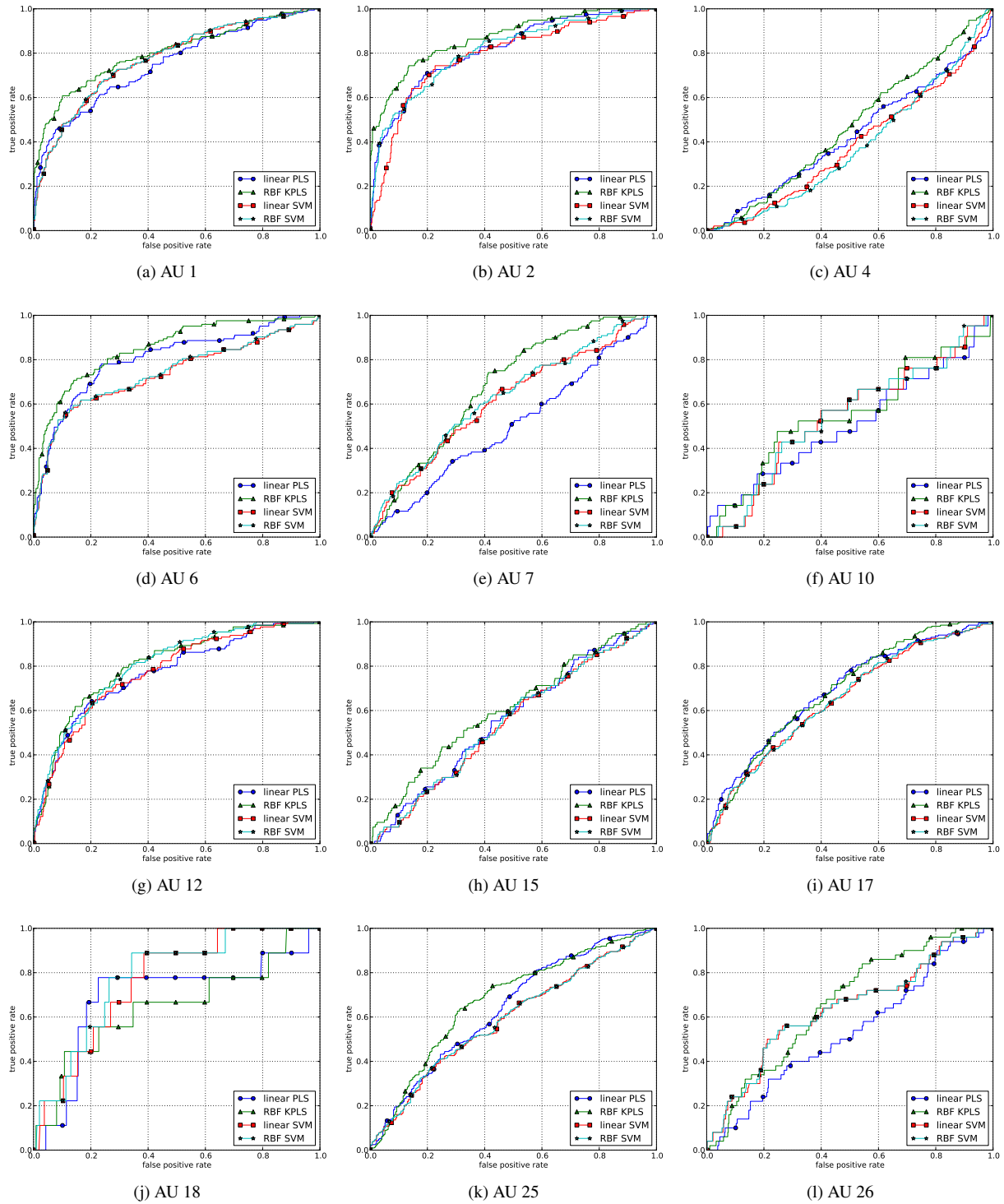


Figure 5. ROC curves on CK+ using classifiers trained on GEMEP-FERA.

B.2. Evaluation across Datasets

B.2.1 Generalization from Constrained to less Constrained Condition

See Table 10 for F1 scores without threshold optimization, Table 11 for F1 scores using threshold optimization based

Table 4. F1 scores in % on CK+ using automatic eye detection. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF PLS	linear SVM	RBF SVM
1	176	76.0	77.2	74.5	72.3
2	117	77.3	78.5	73.6	71.0
4	193	74.3	77.4	71.4	70.8
5	102	75.5	70.4	65.6	67.4
6	123	68.7	68.2	68.4	63.7
7	120	49.3	45.6	43.8	41.5
9	75	84.1	82.2	81.4	85.1
11	34	13.6	5.7	12.0	0.0
12	131	81.5	80.7	78.4	70.1
15	94	53.6	47.1	50.3	31.3
17	201	74.4	77.4	70.7	70.8
20	79	63.7	57.7	49.2	33.0
23	60	34.1	20.0	21.3	16.9
24	58	33.3	20.9	22.8	15.4
25	324	83.6	86.5	83.7	85.1
26	50	18.5	14.0	24.7	13.3
27	81	90.4	85.9	85.7	87.2
AVG		69.7	68.5	66.1	62.9

Table 5. F1 scores in % on CK+ using automatic eye detection and threshold optimization based on EER. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	176.0	63.3	65.7	63.5	63.7
2	117.0	57.4	58.2	55.5	55.7
4	193.0	63.5	65.1	64.1	65.5
5	102.0	50.1	51.8	47.9	49.4
6	123.0	56.3	55.7	53.6	53.9
7	120.0	47.2	52.0	50.6	51.1
9	75.0	49.7	49.0	48.5	48.5
11	34.0	15.0	17.1	15.0	20.1
12	131.0	59.4	60.8	56.9	59.5
15	94.0	45.9	42.7	40.6	40.2
17	201.0	65.4	64.5	63.8	65.2
20	79.0	39.2	42.3	37.5	38.8
23	60.0	28.0	27.8	25.0	26.3
24	58.0	24.9	29.0	28.7	31.5
25	324.0	76.1	78.1	76.8	77.5
26	50.0	21.2	19.8	21.6	19.8
27	81.0	53.1	54.0	51.3	51.3
AVG		56.2	57.4	55.5	56.4

Table 6. F1 scores in % on CK+ using automatic eye detection and threshold optimization based on F1 score. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	176.0	66.8	67.7	67.9	65.5
2	117.0	63.8	61.4	60.9	61.1
4	193.0	64.8	68.7	68.1	67.6
5	102.0	57.6	58.5	53.3	53.3
6	123.0	60.6	59.9	58.5	58.2
7	120.0	48.6	53.1	53.3	51.8
9	75.0	53.7	53.9	52.0	53.0
11	34.0	19.7	18.9	15.7	24.1
12	131.0	66.5	65.4	64.0	63.9
15	94.0	49.2	47.3	46.5	47.4
17	201.0	65.0	66.7	63.1	63.7
20	79.0	48.1	50.0	46.5	47.7
23	60.0	32.5	31.9	32.8	31.2
24	58.0	31.5	34.7	36.6	34.7
25	324.0	75.2	77.3	75.2	76.7
26	50.0	25.3	27.4	28.1	26.2
27	81.0	56.4	56.6	58.8	58.2
AVG		59.3	60.5	59.2	59.2

Table 7. F1 scores in % on GEMEP-FERA. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	1584	45.3	34.1	39.4	39.5
2	1618	35.0	29.8	31.8	25.0
4	1342	11.5	6.6	21.6	16.3
6	1780	63.5	61.0	58.6	58.9
7	2100	51.3	43.0	51.9	51.2
10	2008	47.5	44.1	54.7	54.8
12	2692	67.3	68.1	67.9	66.3
15	1014	21.3	10.4	31.0	21.8
17	820	18.9	13.2	27.4	19.4
18	417	19.4	13.4	8.8	2.8
25	874	19.1	6.0	67.3	67.7
26	544	1.3	0.0	30.5	29.7
AVG		41.9	36.6	46.5	44.0

Table 8. F1 scores in % on GEMEP-FERA using threshold optimization based on EER. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	1584.0	46.4	44.8	45.7	45.5
2	1618.0	38.3	36.4	38.4	37.8
4	1342.0	29.3	24.1	26.0	34.3
6	1780.0	62.1	60.8	58.4	58.8
7	2100.0	58.2	55.2	60.6	60.2
10	2008.0	45.8	46.5	54.1	55.8
12	2692.0	64.4	66.1	67.2	64.4
15	1014.0	35.2	38.0	37.5	39.1
17	820.0	27.4	30.7	33.3	34.4
18	417.0	20.1	22.5	18.0	16.7
25	771.0	55.5	60.5	59.6	58.9
26	478.0	45.9	40.9	44.0	44.3
AVG		48.4	47.9	49.9	50.4

Table 9. F1 scores in % on GEMEP-FERA using threshold optimization based on F1 score. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	1584.0	44.0	38.1	37.4	37.0
2	1618.0	28.7	37.2	44.7	24.8
4	1342.0	20.3	31.5	44.6	35.4
6	1780.0	61.7	60.6	58.0	59.3
7	2100.0	61.6	55.3	61.2	61.1
10	2008.0	55.4	55.9	58.6	57.9
12	2692.0	68.9	64.1	64.5	61.1
15	1014.0	34.1	36.0	36.7	37.5
17	820.0	25.4	28.5	28.0	35.5
18	417.0	17.2	29.8	15.8	16.8
25	771.0	61.0	36.3	32.1	32.0
26	478.0	52.5	11.8	52.9	52.5
AVG		49.0	46.8	50.0	47.2

Table 2. F1 scores in % on CK+ using eye labels and threshold optimization based on EER. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	176.0	62.5	66.7	65.0	65.8
2	117.0	55.8	56.5	55.9	56.3
4	193.0	64.6	65.9	61.4	60.5
5	102.0	51.5	54.6	49.3	50.1
6	123.0	54.9	54.8	54.5	55.2
7	120.0	51.8	51.9	51.7	51.5
9	75.0	49.7	50.9	48.3	47.7
11	34.0	15.6	18.1	16.8	20.2
12	131.0	61.9	62.1	59.4	60.1
15	94.0	41.1	40.7	37.7	39.1
17	201.0	64.8	68.2	65.3	67.4
20	79.0	37.9	39.9	36.0	37.2
23	60.0	28.7	29.4	25.4	26.1
24	58.0	27.7	28.4	29.3	28.4
25	324.0	78.5	79.3	79.9	79.6
26	50.0	20.7	21.0	21.4	21.7
27	81.0	52.2	52.0	51.6	50.6

Table 13. F1 scores in % on CK+ using classifiers trained on GEMEP-FERA. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	176	52.2	51.0	48.9	48.4
2	117	54.5	60.3	38.9	11.0
4	193	11.4	1.0	25.3	16.2
6	123	54.6	58.8	53.2	53.0
7	120	13.8	9.6	26.3	29.6
10	21	7.5	9.3	8.1	7.3
12	131	47.6	50.9	47.2	45.6
15	94	20.9	14.9	13.2	13.4
17	201	35.6	13.2	45.1	45.0
18	9	0.0	0.0	0.0	0.0
25	324	22.9	1.8	68.7	68.7
26	50	0.0	0.0	19.3	18.8
AVG		31.5	23.2	43.4	40.2

Table 14. F1 scores in % on CK+ using classifiers trained on GEMEP-FERA and threshold optimization based on EER. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	176.0	56.0	62.2	60.2	59.7
2	117.0	49.7	49.9	53.7	52.2
4	193.0	23.2	12.7	12.8	8.3
6	123.0	57.8	54.7	56.0	56.8
7	120.0	15.7	23.4	26.8	30.4
10	21.0	6.3	6.8	10.3	9.6
12	131.0	46.3	48.0	47.4	46.1
15	94.0	27.8	29.9	15.6	16.6
17	201.0	54.5	45.5	50.3	49.9
18	9.0	5.9	6.7	7.6	7.5
25	324.0	54.5	10.7	64.9	64.6
26	50.0	15.2	0.0	21.8	21.8
AVG		43.2	32.5	44.7	44.1

Table 15. F1 scores in % on CK+ using classifiers trained on GEMEP-FERA and threshold optimization based on F1 score. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	176.0	54.7	53.9	59.9	59.2
2	117.0	55.4	64.3	50.5	55.2
4	193.0	13.9	2.0	6.2	1.0
6	123.0	57.6	62.6	51.2	50.7
7	120.0	13.7	14.0	28.1	29.6
10	21.0	6.9	10.1	10.1	9.9
12	131.0	46.5	48.6	45.5	45.0
15	94.0	24.4	18.4	9.7	9.8
17	201.0	50.1	24.1	40.0	40.0
18	9.0	4.9	0.0	0.0	0.0
25	324.0	58.6	30.2	69.0	68.8
26	50.0	15.5	0.0	24.8	25.1
AVG		42.3	31.9	42.4	42.0

Table 10. F1 scores in % on GEMEP-FERA using classifiers trained on CK+. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	1584	50.0	45.5	48.9	48.1
2	1618	31.0	12.1	19.3	18.6
4	1342	42.0	34.2	37.5	28.7
5	735	25.0	17.6	24.2	24.3
6	1780	56.9	14.4	36.1	14.3
7	2100	9.1	0.0	9.1	8.8
9	392	8.1	0.0	10.8	6.7
11	512	1.1	0.0	15.6	0.0
12	2692	47.0	11.1	29.8	22.2
15	1014	14.8	1.7	22.9	3.9
17	820	21.8	12.1	21.2	15.3
20	480	26.4	5.2	19.2	16.8
23	163	0.0	0.0	0.0	0.0
24	124	4.5	0.0	0.0	0.0
25	874	28.6	29.1	67.4	70.8
26	544	13.6	0.0	8.8	0.0
27	27	12.9	26.8	11.2	3.9
AVG		31.8	14.7	27.6	20.9

Table 11. F1 scores in % on GEMEP-FERA using classifiers trained on CK+ and threshold optimization based on EER. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	1584.0	49.9	40.1	50.5	50.0
2	1618.0	37.6	8.2	42.4	42.3
4	1342.0	41.8	20.1	42.6	40.7
5	735.0	28.2	7.6	32.8	32.0
6	1780.0	62.6	8.1	59.1	59.7
7	2100.0	20.4	0.1	41.9	42.2
9	392.0	17.9	0.5	16.8	15.3
11	512.0	17.9	0.0	23.1	19.4
12	2692.0	61.8	1.8	56.2	58.1
15	1014.0	24.3	0.0	35.6	35.3
17	820.0	22.2	8.1	24.2	22.9
20	480.0	22.5	0.0	22.1	20.9
23	163.0	4.0	0.0	2.0	2.0
24	124.0	8.0	0.0	6.5	0.0
25	771.0	67.4	67.9	64.9	66.4
26	478.0	43.8	0.4	47.2	46.9
27	27.0	10.6	17.0	4.4	4.7
AVG		41.0	11.3	44.1	44.0

Table 12. F1 scores in % on GEMEP-FERA using classifiers trained on CK+ and threshold optimization based on F1 score. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	1584.0	49.9	40.1	50.0	49.7
2	1618.0	37.3	8.2	25.6	25.2
4	1342.0	41.8	20.1	42.6	35.8
5	735.0	25.3	7.6	30.6	29.8
6	1780.0	60.4	8.1	48.3	41.5
7	2100.0	21.2	0.1	16.7	16.9
9	392.0	12.7	0.5	11.1	14.2
11	512.0	14.2	0.0	21.4	0.6
12	2692.0	59.7	1.8	38.6	37.5
15	1014.0	16.0	0.0	28.9	25.0
17	820.0	17.3	8.1	24.7	24.5
20	480.0	25.3	0.0	23.6	22.2
23	163.0	1.1	0.0	1.3	0.7
24	124.0	6.6	0.0	0.0	0.0
25	771.0	68.6	67.9	67.8	69.9
26	478.0	36.3	0.4	28.8	4.7
27	27.0	11.2	17.0	8.2	9.3
AVG		39.3	11.3	34.1	31.2

Table 3. F1 scores in % on CK+ using eye labels and threshold optimization based on F1 score. AVG is the weighted average over the individual results, depending on the number of positive samples given by in the column N.

AU	N	linear PLS	RBF KPLS	linear SVM	RBF SVM
1	176.0	68.0	69.3	65.8	67.0
2	117.0	63.2	62.6	61.5	61.3
4	193.0	66.9	69.4	61.0	63.5
5	102.0	57.8	59.3	55.5	56.4
6	123.0	62.1	60.4	59.3	61.5
7	120.0	54.4	54.6	54.0	53.5
9	75.0	54.3	55.8	53.8	54.3
11	34.0	21.7	25.6	23.6	25.9
12	131.0	68.7	68.1	63.6	65.7
15	94.0	51.3	47.6	45.0	45.9
17	201.0	66.4	68.2	64.8	65.4
20	79.0	44.0	50.8	45.9	47.4
23	60.0	38.3	41.0	34.5	36.6
24	58.0	37.8	37.8	33.9	37.1
25	324.0	76.5	76.0	79.0	76.1
26	50.0	22.9	23.0	22.0	22.4
27	81.0	58.0	56.8	56.7	56.2
AVG		60.9	61.5	59.2	59.7

on EER and Table 12 for F1 scores using threshold opti-

mization based on F1 score.

B.2.2 Generalization from less Constrained to Constrained Condition

See Table 13 for F1 scores without threshold optimization, Table 14 for F1 scores using threshold optimization based on EER and Table 15 for F1 scores using threshold optimization based on F1 score.

References

- [1] T. Gehrig and H. K. Ekenel. Facial Action Unit Detection Using Kernel Partial Least Squares. In *1st IEEE Int'l Workshop on Benchmarking Facial Image Analysis Technologies (BeFIT 2011)*, Barcelona, Spain, Nov. 2011. 1
- [2] J. M. Girard and J. F. Cohn. Criteria and metrics for thresholded AU detection. In *1st IEEE Int'l Workshop on Benchmarking Facial Image Analysis Technologies (BeFIT 2011)*, Barcelona, Spain, Nov. 2011. 1