

## Sentiment Analysis Results on the CMU-MOSI Training and Validation Sets

We report the unimodal and multimodal results on the training and validation sets of the CMU-MOSI database in Tables 1 and 2.

In these tables, “S” is the single-task learning model; “S+P” is the bi-task learning model with polarity classification as the auxillary task; “S+I” is the bi-task learning model with intensity classification as the auxillary task; “S+P+I” is the tri-task learning model.

To evaluate the performance of sentiment score prediction, we report both Pearson’s correlation coefficient (CC, higher is better) and mean absolute error (MAE, lower is better) between predictions and annotations of sentiment scores.

CC	S	S+P	S+I	S+P+I
Vocal	0.485	0.389	0.242	0.319
Visual	0.310	0.291	0.219	0.272
Verbal	0.760	0.767	0.767	0.761
EF	0.831	0.822	0.855	0.839
TFN	0.925	0.913	0.879	0.925
LF	0.861	0.849	0.851	0.865
HF	0.870	0.853	0.858	0.854
MAE	S	S+P	S+I	S+P+I
Vocal	1.102	1.172	1.264	1.213
Visual	1.231	1.260	1.302	1.299
Verbal	0.806	0.804	0.774	0.798
EF	0.727	0.743	0.684	0.727
TFN	0.552	0.596	0.646	0.579
LF	0.668	0.712	0.679	0.678
HF	0.654	0.712	0.664	0.699

Table 1: Results on CMU-MOSI training set.

CC	S	S+P	S+I	S+P+I
Vocal	0.140	0.154	0.132	0.093
Visual	0.093	0.032	0.142	0.162
Verbal	0.559	0.501	0.485	0.522
EF	0.527	0.520	0.506	0.513
TFN	0.533	0.529	0.477	0.558
LF	0.503	0.468	0.472	0.453
HF	0.523	0.451	0.513	0.480
MAE	S	S+P	S+I	S+P+I
Vocal	1.388	1.368	1.389	1.394
Visual	1.408	1.416	1.418	1.413
Verbal	1.154	1.197	1.190	1.170
EF	1.189	1.180	1.194	1.209
TFN	1.196	1.216	1.231	1.166
LF	1.208	1.240	1.255	1.230
HF	1.189	1.237	1.190	1.215

Table 2: Results on CMU-MOSI validation set.