

Method name	Window size, seconds	Pooling	Test RMSE (A)	Test MAE (A)	Test RMSE (V)	Test MAE (V)
Baseline	x	x	0.1084	0.0846	0.1201	0.0929
Video-based	5, 1 sec	AVG Pool	0.1225	0.0955	0.1141	0.0893
Video-based	5, 1 sec	MAX Pool	0.1226	0.0945	0.1288	0.0995
Video-based	5, 1 sec	1D CNN	0.9673	0.9539	0.9682	0.9532
Video-based	10, 2 sec	AVG Pool	0.1159	0.0894	0.1077	0.0850
Video-based	10, 2 sec	MAX Pool	0.1155	0.0901	0.1171	0.0888
Video-based	10, 2 sec	1D CNN	0.9669	0.9539	1.0599	1.0467
Video-based	15, 3 sec	AVG Pool	0.1106	0.0848	0.1029	0.0793
Video-based	15, 3 sec	MAX Pool	0.1135	0.0894	0.1107	0.0839
Video-based	15, 3 sec	1D CNN	1.0572	1.0457	1.0594	1.0466
Video-based	20, 4 sec	AVG Pool	0.1135	0.0887	0.1050	0.0805

Video-based	20, 4 sec	MAX Pool	0.1138	0.0884	0.1273	0.0990
Video-based	20, 4 sec	1D CNN	0.9660	0.9539	0.9670	0.9533
Audio-based	1 sec	x	0.1796	0.1339	0.1669	0.1326
Audio-based	2 sec	x	0.1843	0.1376	0.1698	0.1347
Audio-based	3 sec	x	0.1863	0.1379	0.1665	0.1325
Audio-based	4 sec	x	0.1810	0.1339	0.1672	0.1338
Fusion v1(1)	5, 1 sec	1D CNN, AVG Pool	0.1682	0.1336	0.1835	0.1371
Fusion v1(1)	10, 2 sec		0.1346	0.1064	0.1533	0.1151
Fusion v1(1)	15. 3 sec		0.1391	0.1097	0.1548	0.1165
Fusion v1(1)	20, 4 sec		0.1346	0.1042	0.1589	0.1196
Fusion v1(2)	5, 1 sec	1D CNN, AVG Pool	0.1208	0.0932	0.1106	0.0861
Fusion v1(2)	10, 2 sec		0.1210	0.0937	0.1098	0.0867
Fusion v1(2)	15. 3 sec		0.1261	0.0976	0.1050	0.0827

Fusion v1(2)	20, 4 sec		0.1239	0.0956	0.1035	0.0815
Fusion v1(3)	5, 1 sec	1D CNN, AVG Pool	0.1683	0.1337	0.1852	0.1379
Fusion v1(3)	10, 2 sec		0.1678	0.1341	0.1855	0.1385
Fusion v1(3)	15. 3 sec		0.1713	0.1374	0.1842	0.1379
Fusion v1(3)	20, 4 sec		0.1719	0.1368	0.1836	0.1377
Fusion v2	5, 1 sec	AVG Pool	0.1355	0.1025	0.1206	0.0928
Fusion v2	10, 2 sec		0.1328	0.1034	0.1401	0.1058
Fusion v2	15. 3 sec		0.1470	0.1092	0.1203	0.0925
Fusion v2	20, 4 sec		0.1350	0.1057	0.1494	0.1129
Fusion v3	5, 1 sec	AVG Pool	0.1289	0.1021	0.1232	0.0940
Fusion v3	10, 2 sec		0.1327	0.1033	0.1201	0.0923
Fusion v3	15. 3 sec		0.1292	0.1013	0.1193	0.0914
Fusion v3	20, 4 sec		0.1284	0.0988	0.1174	0.0902

Fusion v4	5, 1 sec	AVG Pool 2x	0.1213	0.0944	0.1113	0.0876
Fusion v4	10, 2 sec		0.1183	0.0930	0.1145	0.0893
Fusion v4	15. 3 sec		0.1246	0.0975	0.1174	0.0923
Fusion v4	20, 4 sec		0.1297	0.0999	0.1197	0.0941