

Latest Emotion Classification results

Results by model

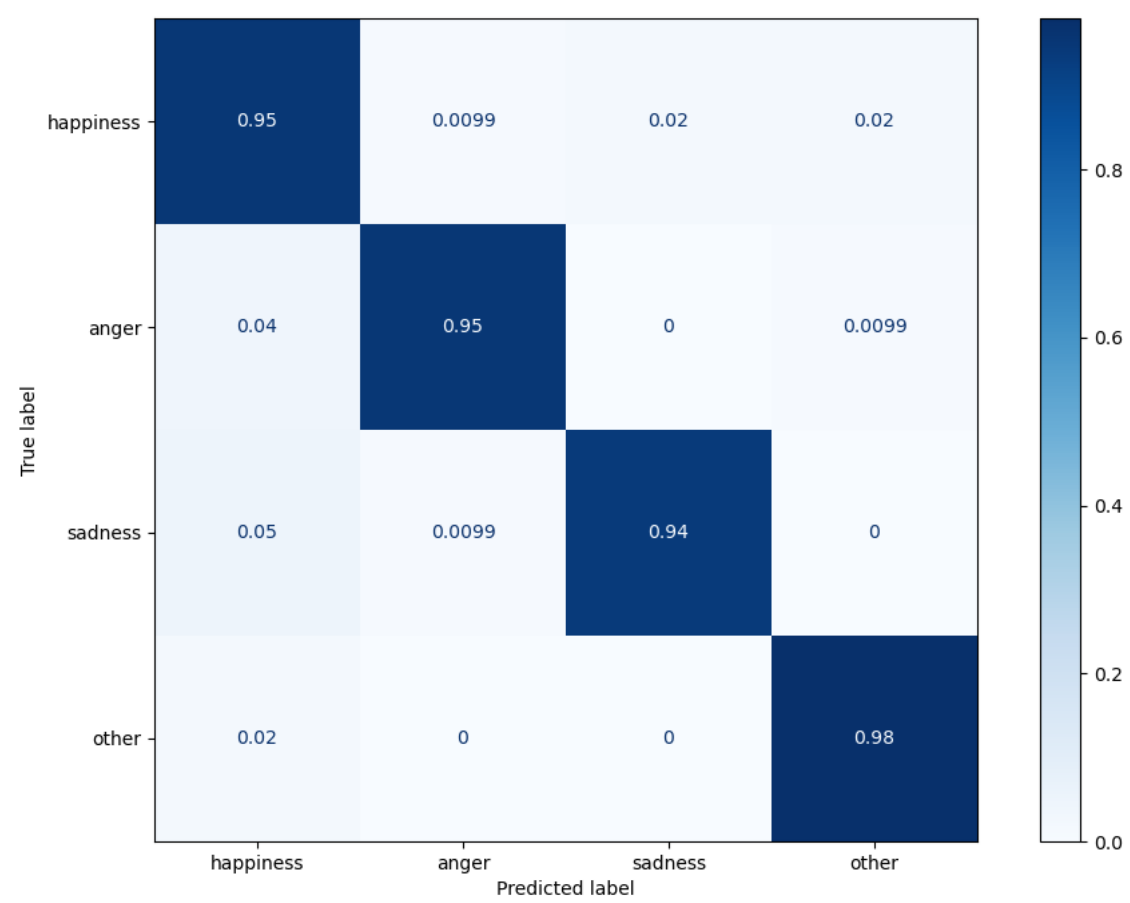
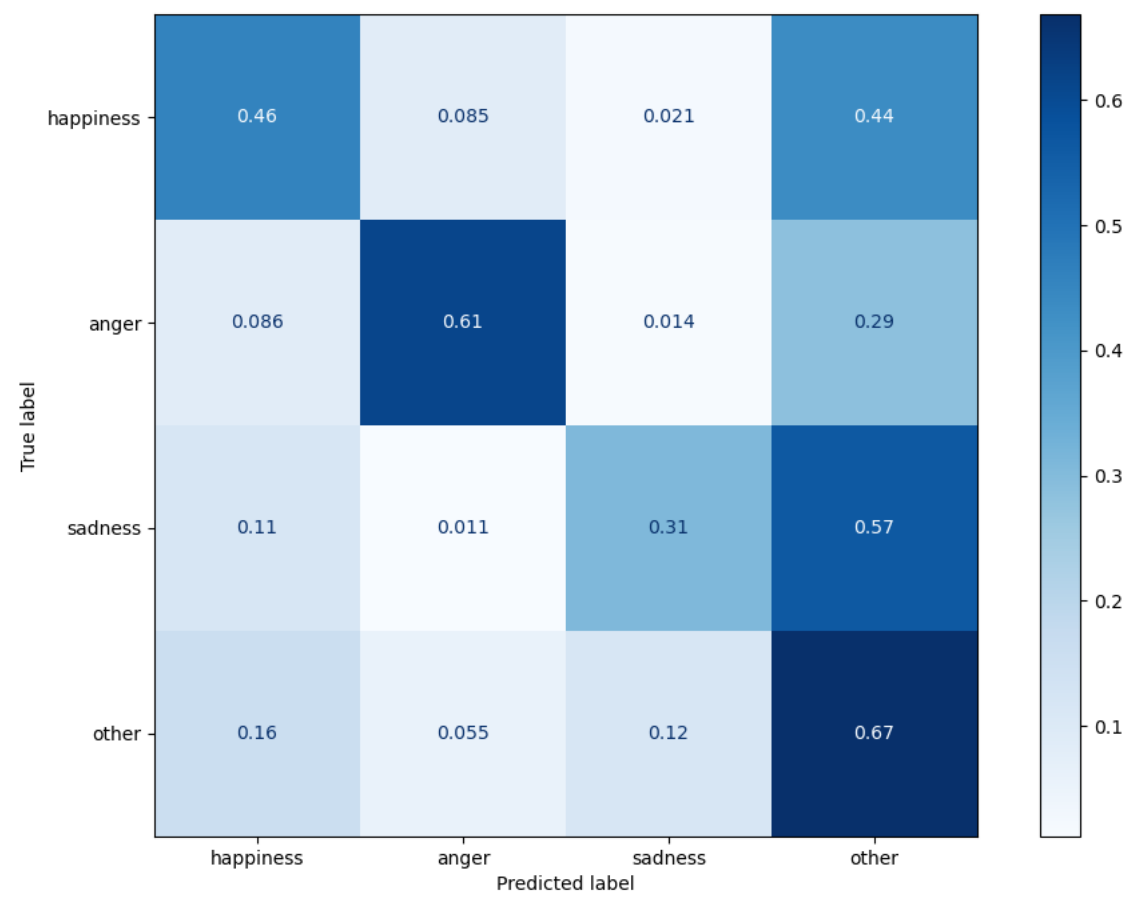
model type	best test acc	train accuracy	eval acc 1 (improv)	eval acc 2 (pp)	eval acc improv augn	eval acc pp augm	API acc	dataset
russian baseline (not fine-tuned on our data)	-	-	0.598	0.374			-	-
Russian (19)	0.730	0.986	0.675	0.822			0.68	emo_audio_pp_relabelled_threshold_2_test_train_split
russian (21)	0.9125	0.9999	0.535	0.955	0.529	0.908	0.824	emo_audio_PP_threshold_2_and_other_from_relabelled_threshold_3
russian (23)	0.9625	0.992	0.544	0.960	0.561	0.941	0.87625	emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
russian (24)	0.930	0.996	0.550	0.913				emo_audio_PP_threshold_2_and_other_from_relabelled_threshold_3
xlsr (pre-trained, not fine-tuned) (22)	0.9625	0.986	0.516	0.958				emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
xlsr (no pre-training, just the architecture) (25)	0.720	0.944	0.406	0.731				emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
wav2vec small (no pre-training, just the architecture) (26)	0.7025	0.931	0.456	0.785				emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
lstm big (37 layers)	0.25	0.25	0.11	0.25				emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
lstm small (11 layers)	0.625	0.933	0.465	0.644				emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
conv1d big	0.61	0.898	0.388	0.611				emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
demucs lstm (33)	0.551	0.951	0.439	0.634				emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
xlsr (pre-trained, not fine-tuned) (34)	0.779	0.993	0.564	0.968	0.532	0.911	0.91	russian_emo_dataset_un-relabelled + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
russian relabelled + pp tresh3 (35)	0.922	0.987			0.593	0.953	0.91	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
russian relabelled + pp tresh3 (50% 8k - 16k) audeering	0.926	0.968			0.593	0.97	0.86	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3
russian relabelled + pp tresh3 (100% 8k - 16k) audeering	0.91	0.98			0.589	0.943		
38 (attention, no resampling)	0.892	0.946			0.668	0.928	85.625	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3 + upwork_emo_pp_2022-08-15_4sec_part_1_fixed_threshold_2
39 (attention, 50% resampled)	0.908	0.977			0.636	0.97	86.875	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3 + upwork_emo_pp_2022-08-15_4sec_part_1_fixed_threshold_2
40 (attention, 100% resampled)	0.846	0.945			0.429	0.807	65.875	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3 + upwork_emo_pp_2022-08-15_4sec_part_1_fixed_threshold_2
41 (mean, no resampling)	0.882	0.942			0.607	0.948	86.875	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3 + upwork_emo_pp_2022-08-15_4sec_part_1_fixed_threshold_2
42 (mean, 50% resampled)	0.921	0.967			0.621	0.96	90.625	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3 + upwork_emo_pp_2022-08-15_4sec_part_1_fixed_threshold_2
43 (mean, 100%)	0.923	0.963			0.629	0.965	84.375	russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3 + upwork_emo_pp_2022-08-15_4sec_part_1_fixed_threshold_2
44 (mean, 100% lowpass)	0.915	0.978			0.621	0.96		russian_emo_dataset_tresh2 + emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3 + upwork_emo_pp_2022-08-15_4sec_part_1_fixed_threshold_2
45 (mean, 50%, tresh2)	0.793	0.903			0.668	0.767		russian_emo_dataset_tresh2 + emo_audio_PP_threshold_2 + upwork_08-15_4sec_part_1_fixed_threshold_2

1.) classification, Russian model (frozen conv layers, lr = 1e-5, batch_size = 8, dropout=0.5)

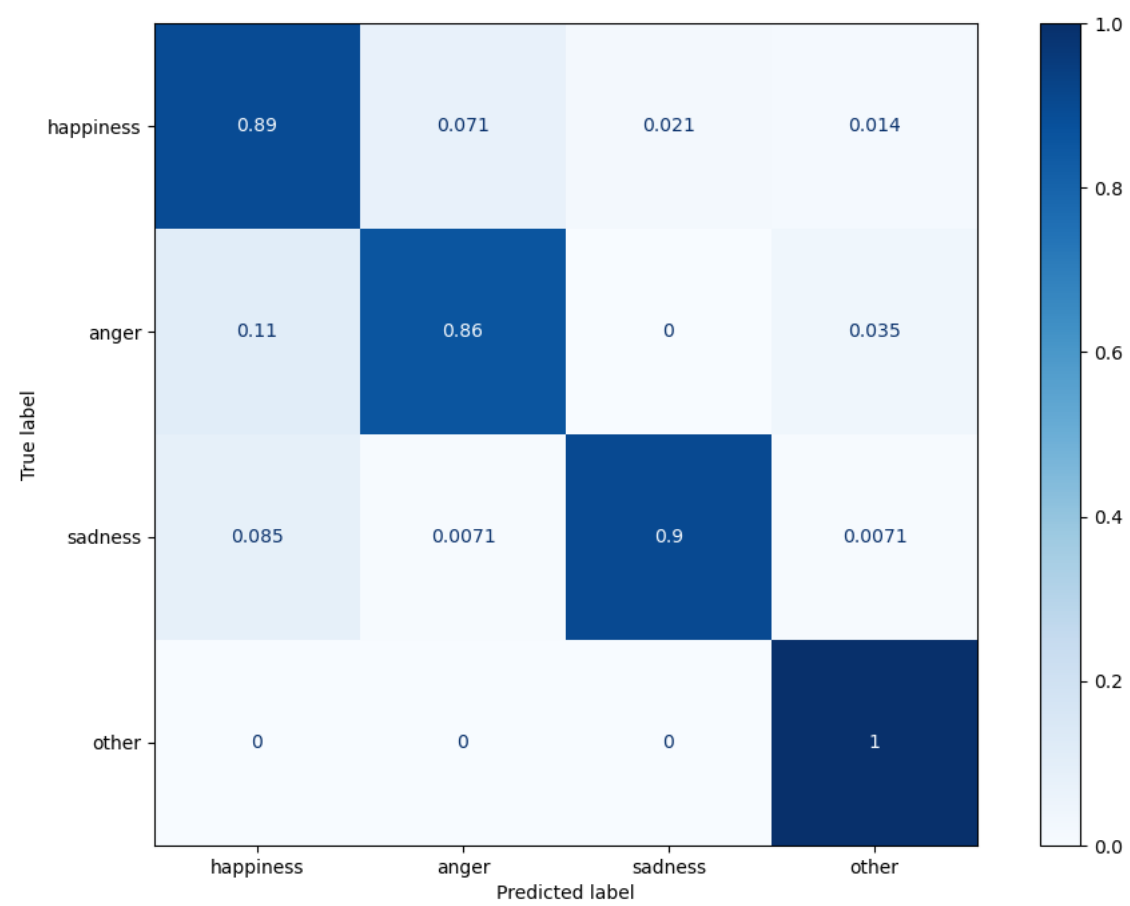
train accuracy = 0.999, test acc = 0.9125, eval acc_1 = 0.535, eval acc_2 = 0.955

dataset - emo_audio_PP_threshold_2_and_other_from_relabelled_threshold_3 (previously trained with emo_audio_pp_relabelled_threshold_2_test_train_split eval_acc_1 = 67.5%, eval_acc_2 = 82%)

eval matrix 1(left) and 2:



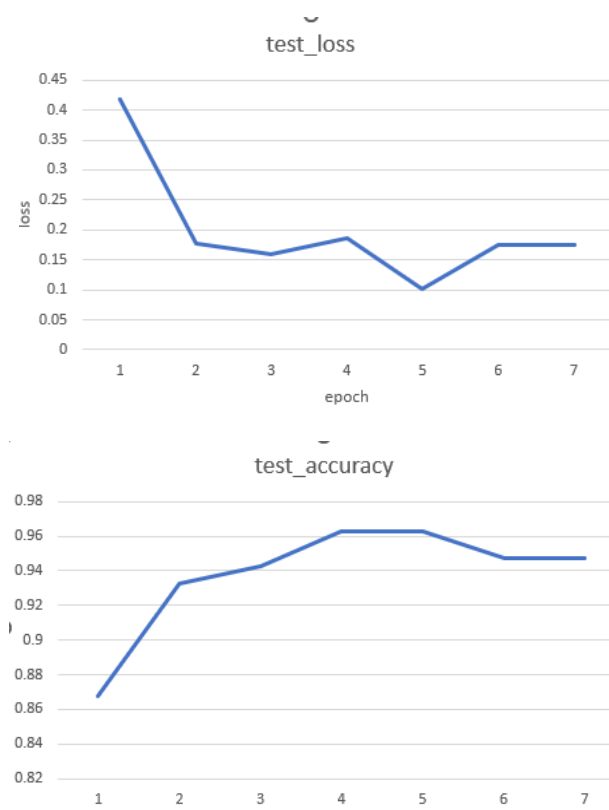
test matrix:



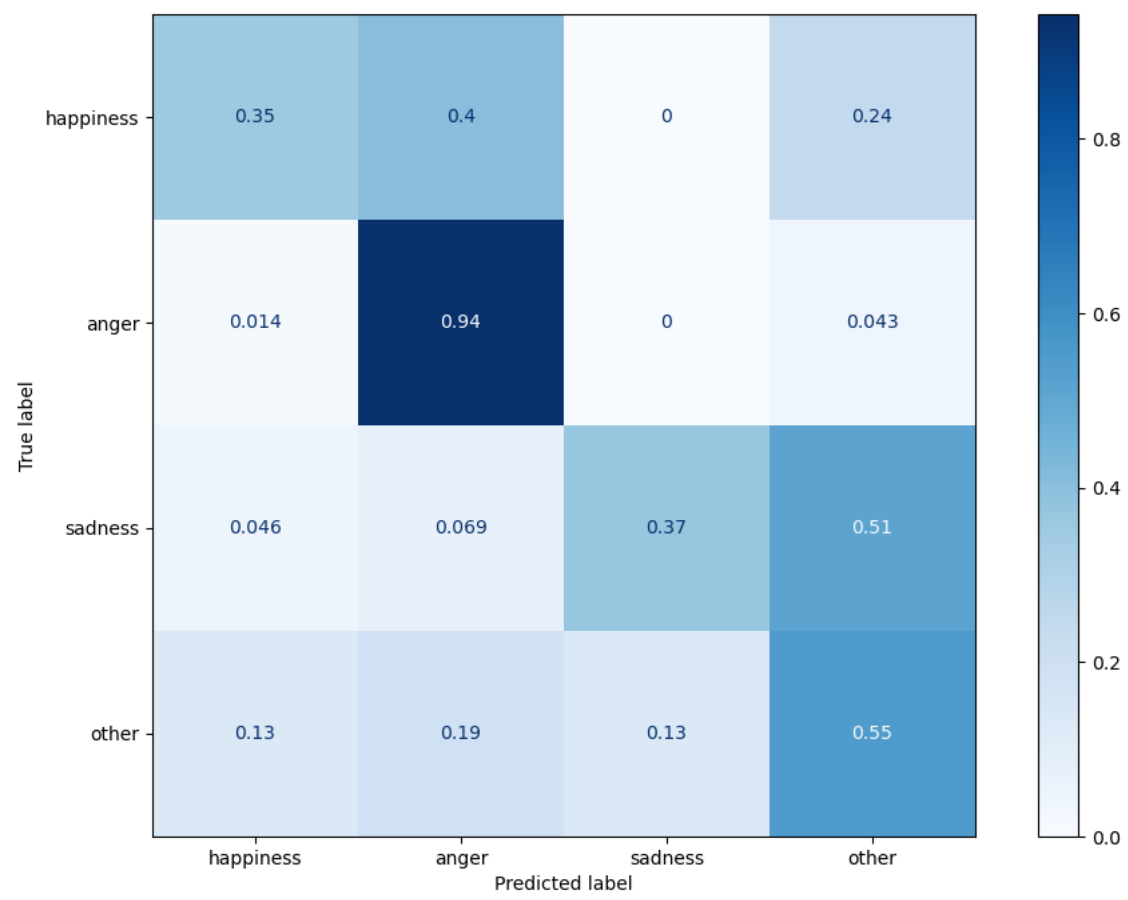
2.) xlsr (pre-trained)

train accuracy = 0.986, test acc = 0.9625, eval acc_1 = 0.516, eval acc_2 = 0.958

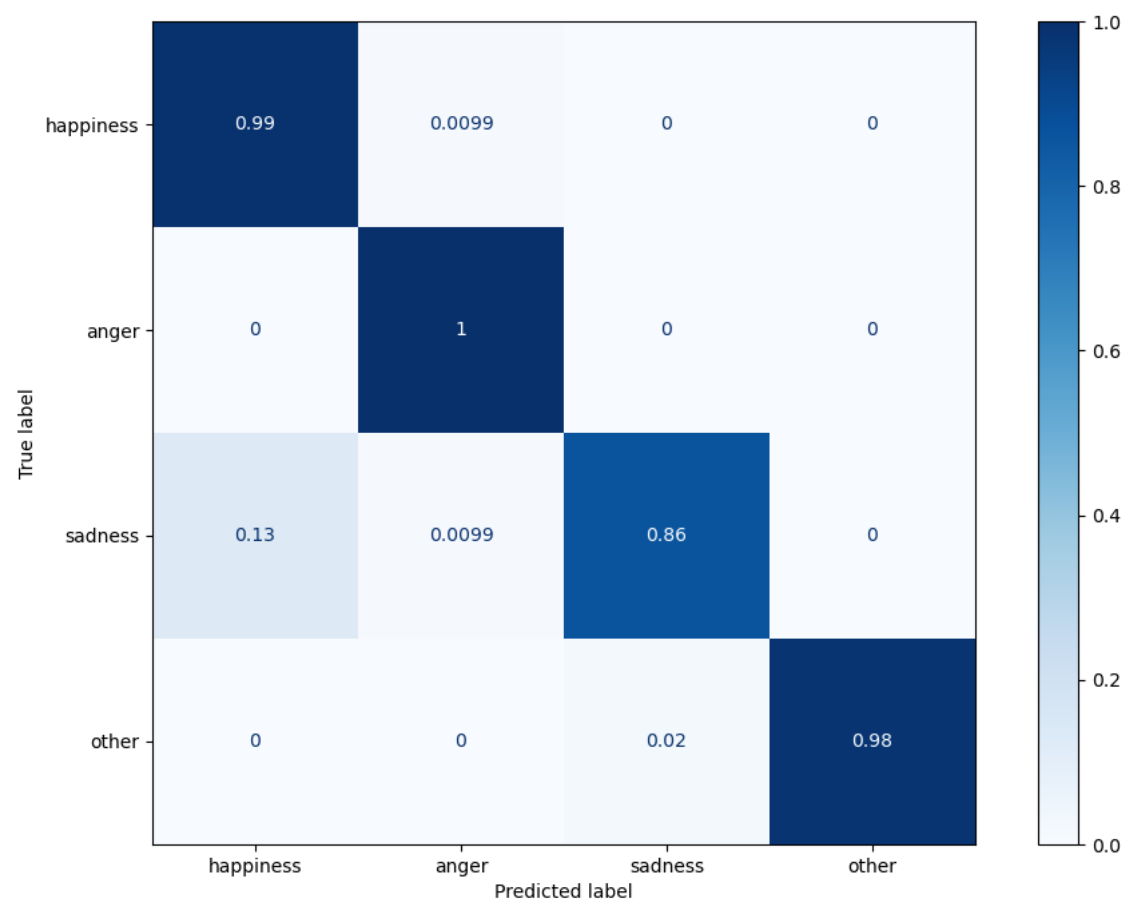
emo_audio_PP_threshold_3_and_other_from_relabelled_threshold_3



eval1



eval2



test

