

# Report

## Emo models

Feature - lots of params, around 316M

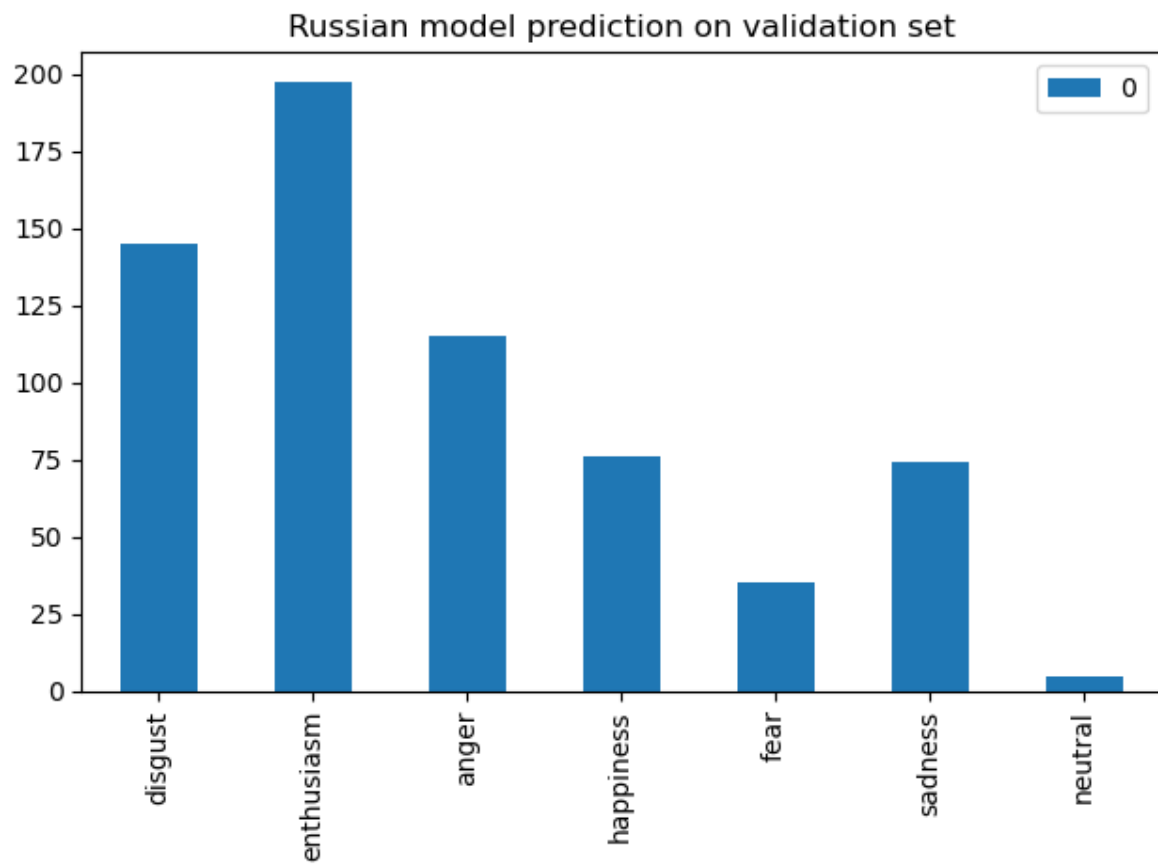
<https://huggingface.co/Aniemore/wav2vec2-xlsr-53-russian-emotion-recognition>  
<https://huggingface.co/Aniemore/wav2vec2-xlsr-53-russian-emotion-recognition>

Russian model predicts 7 classes : happiness, anger, sadness, disgust, fear, enthusiasm, other.

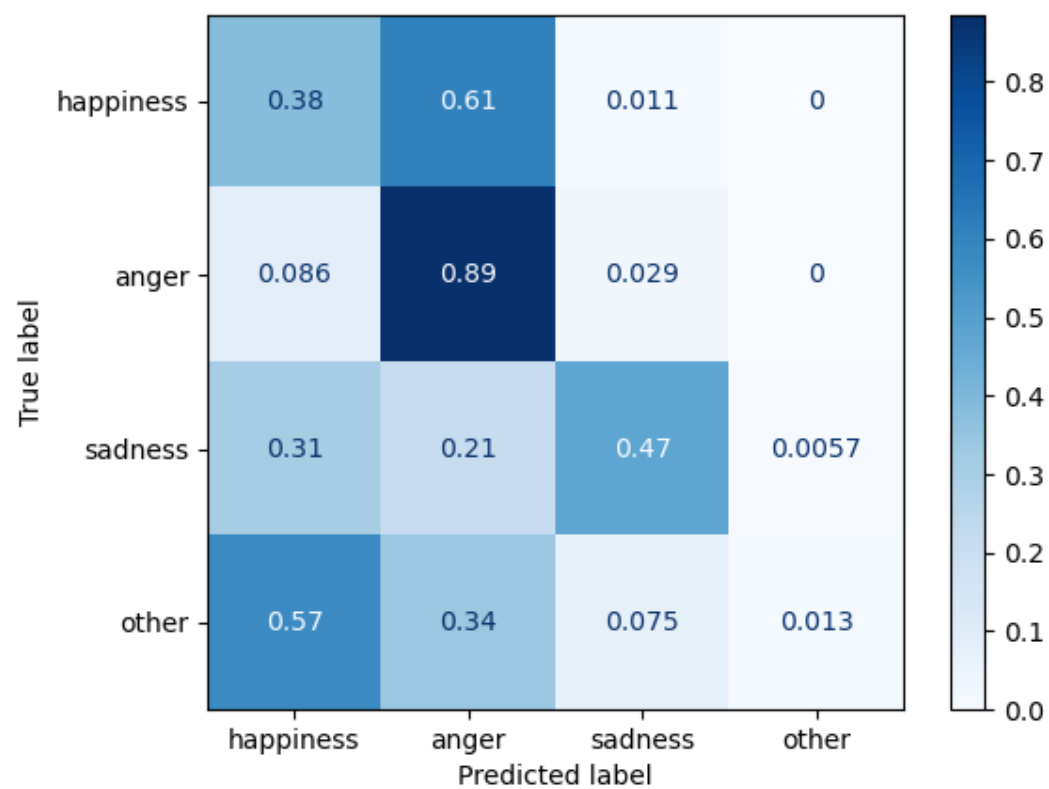
Their dataset - around 3.5h.

## Russian model's results on our eval data:

Histogram of predictions:

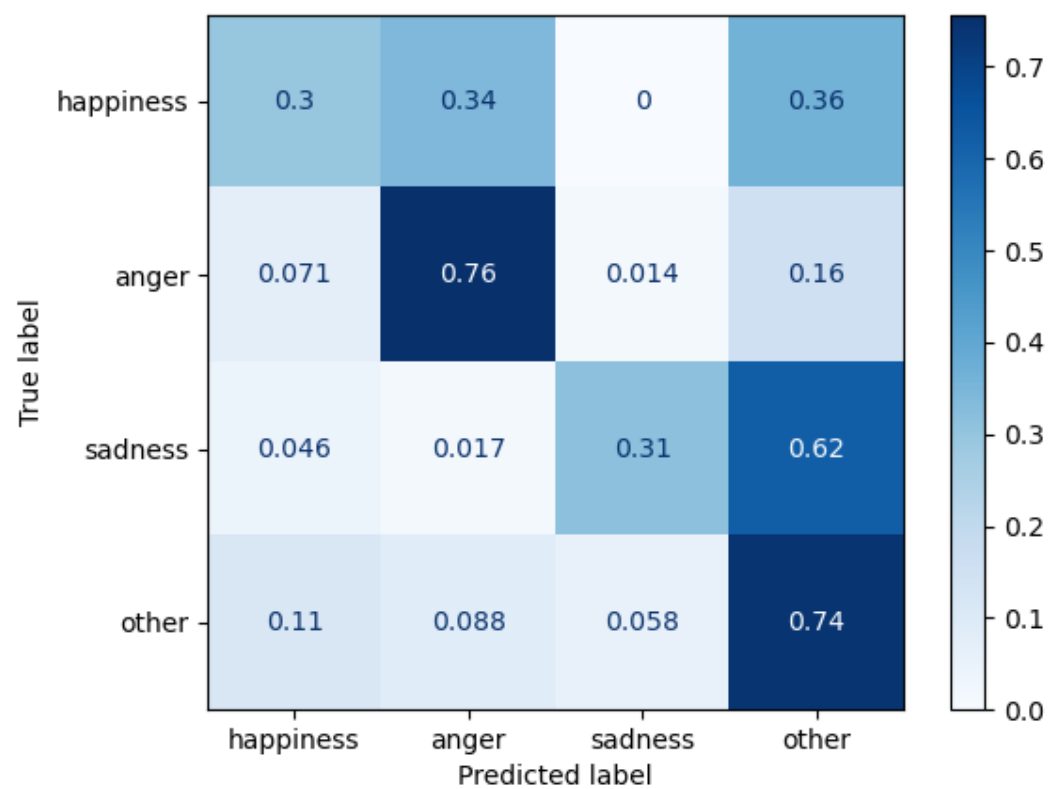


1.) Disgust - anger; fear - sadness, enthusiasm - happiness:  
accuracy = 0.29; f1 = 0.29



2.) disgust, fear, enthusiasm - other:

f1 = acc = 0.56

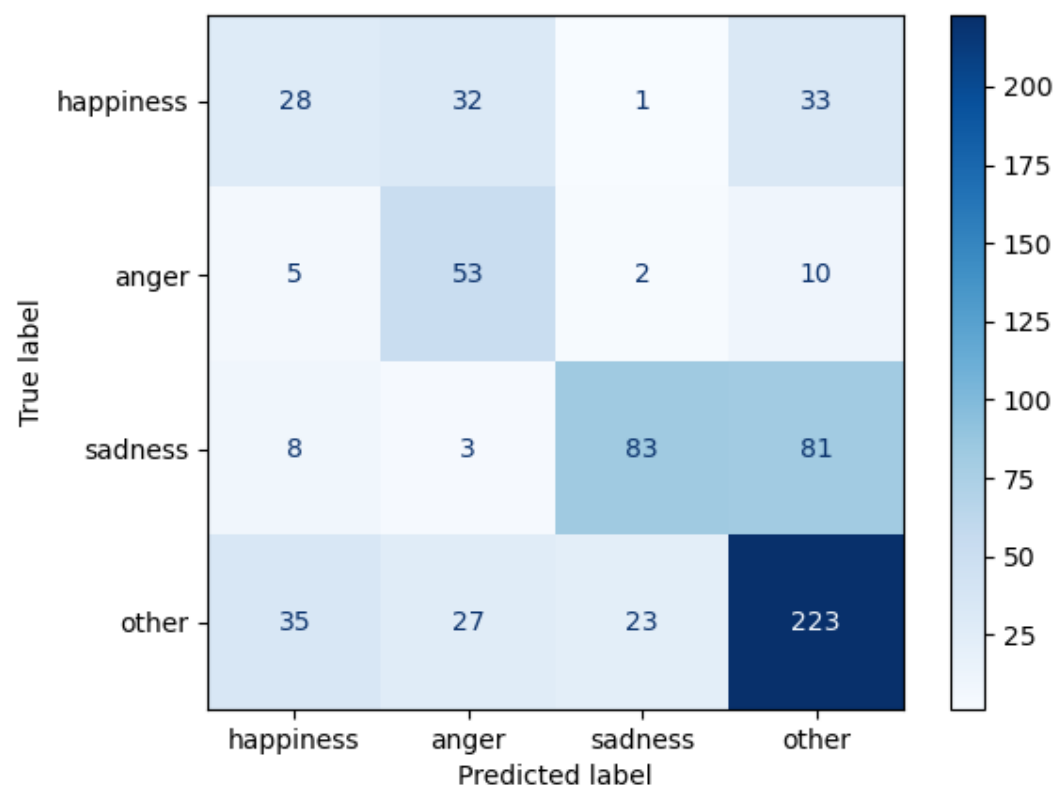


3.) fear - sadness, enthusiasm, disgust - other

f1 = acc = 0.598



The same without normalization:



4.) fear - sadness, disgust - anger, enthusiasm - other

acc = 0.493



The second tested one:

<https://huggingface.co/harshit345/xlsr-wav2vec-speech-emotion-recognition>

Low scores, might just try using as pre-trained.

Other potential candidate for use as pre-trained:

<https://huggingface.co/audeerling/wav2vec2-large-robust-12-ft-emotion-msp-dim>

*"The model expects a raw audio signal as input and outputs predictions for arousal, dominance and valence in a range of approximately 0...1."*

Optimizers - not documented, AdamW most likely.

## Results from fine-tuning the Russian model

Three tested variations:

1. Frozen pre-trained model, fine-tuning the classification layer.

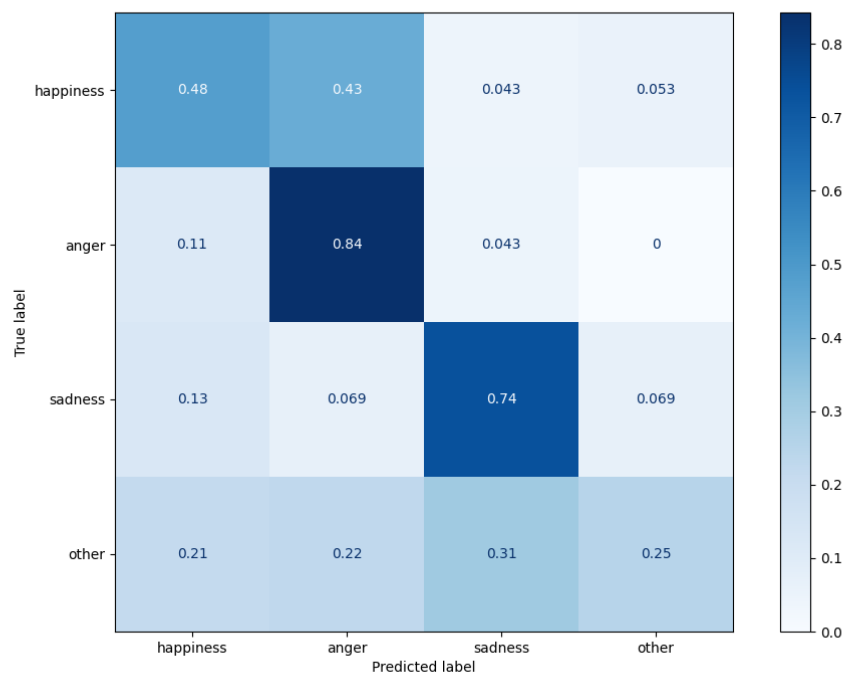
2. Training whole model for classification.
3. Training whole model to output embeddings with contrastive loss.

Trained model's size is 1.2GB

1.) Frozen pre-trained model, fine-tuning the classification layer.

test acc = 64.5%, train acc = 69.5%, eval acc = 47.9% (but trained just 10 epochs, hasn't converged)

eval matrix:

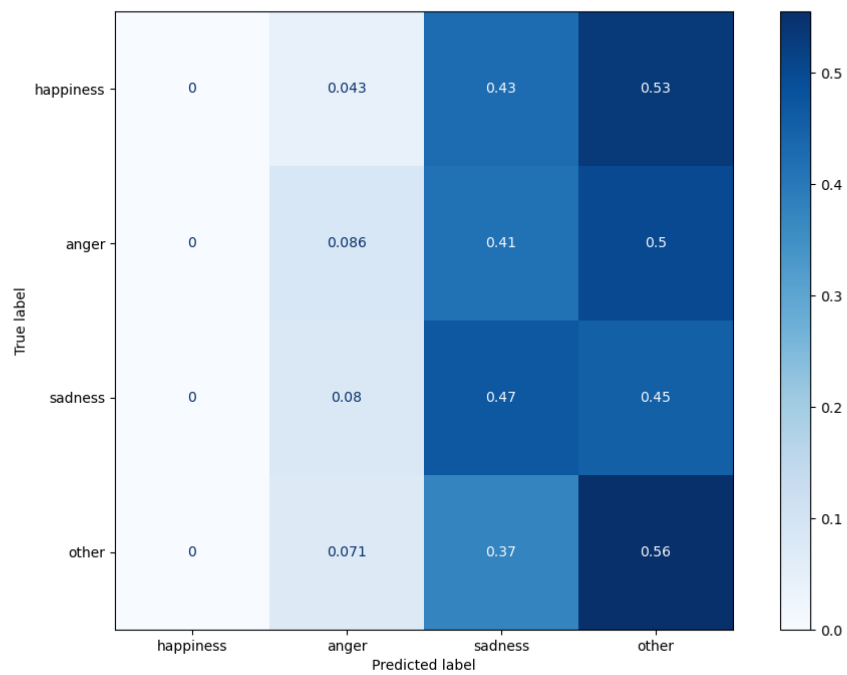


3.) Training whole model to output embeddings with contrastive loss.

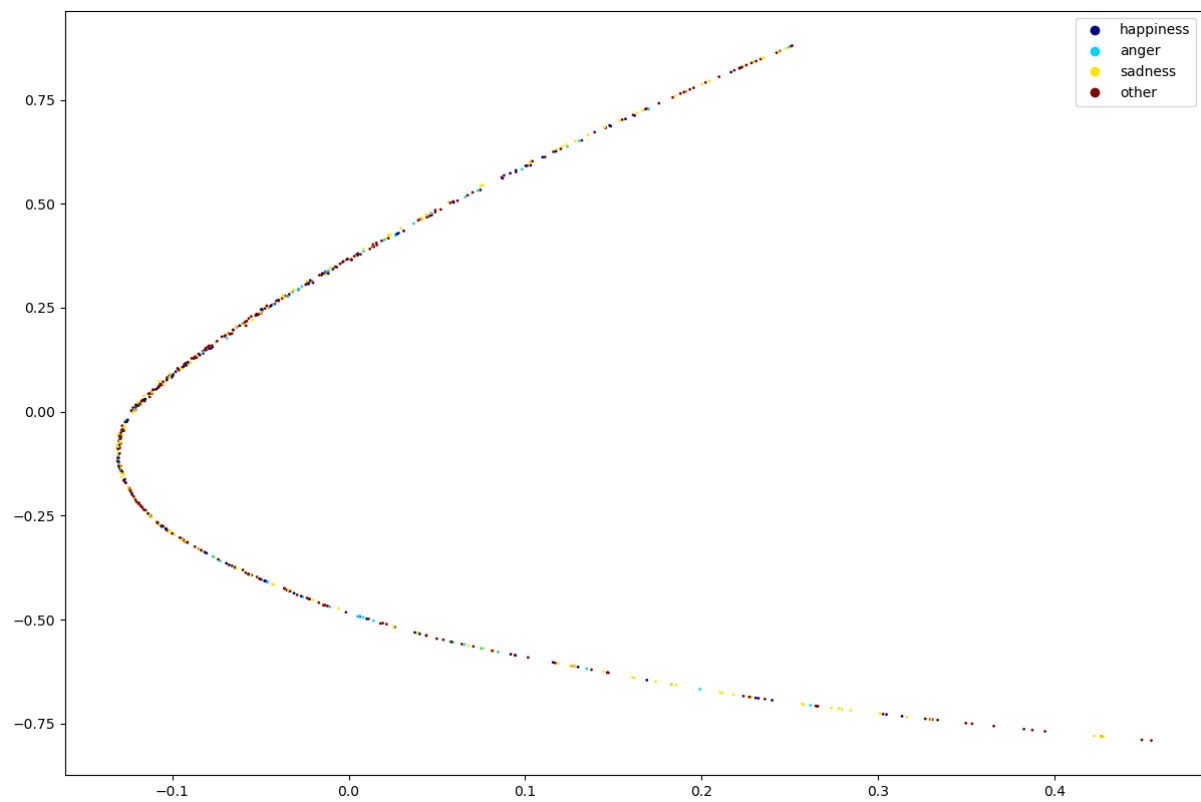
test acc = 39.3%, train acc = 37.8%, eval acc = 40.0%

eval matrix:





eval embeddings:



train embeddings:

