Improving mental disorder detection through emotion recognition

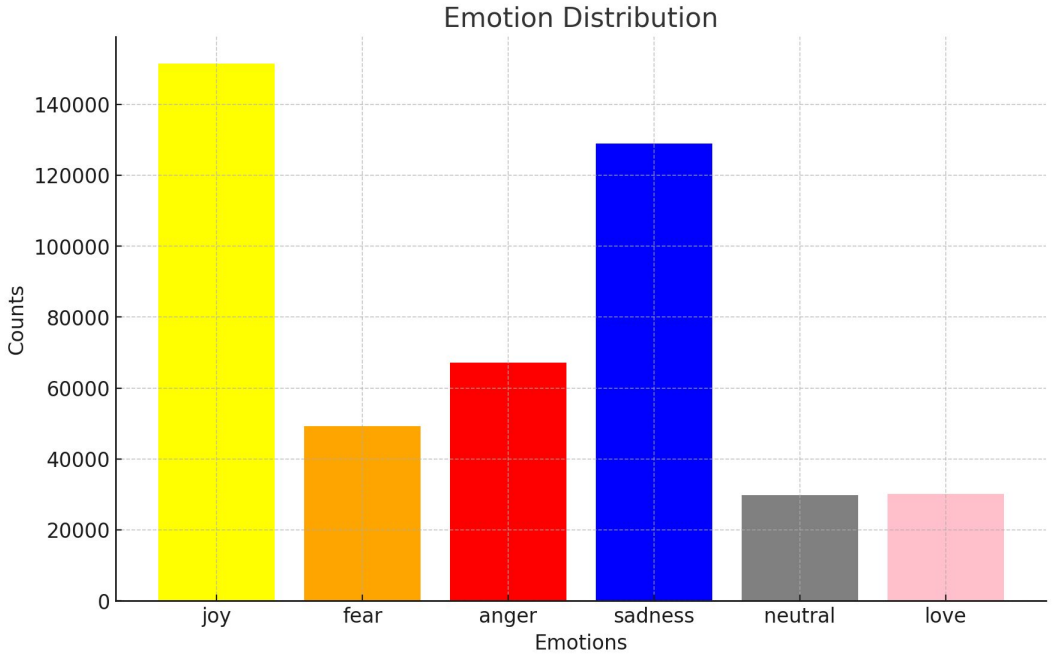
Mihail Bogdan Chirobocea, Răzvan George Costea

University of Bucharest

Data

- Emotion dataset
  - Ensemble of 7 public datasets
  - Merged emotions
  - No duplicates
  - No significant unbalance
  - Removed too long texts
  - Total of 450K samples

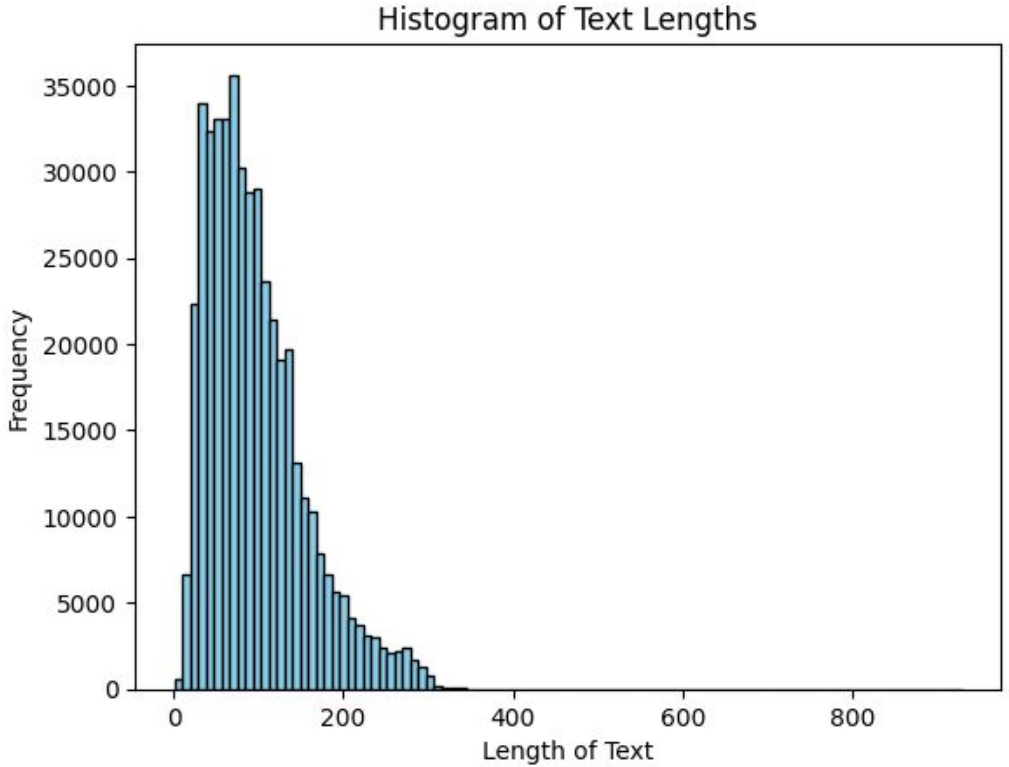
- Depression dataset
  - 8000 negative samples
  - 2000 positive samples



Data

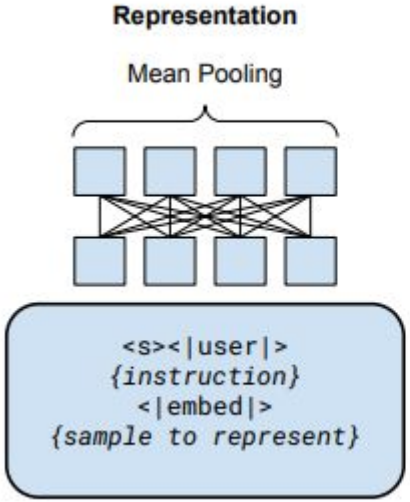
- Emotion dataset
  - Ensemble of 7 public datasets
  - Merged emotions
  - No duplicates
  - No significant unbalance
  - Removed too long texts
  - Total of 450K samples

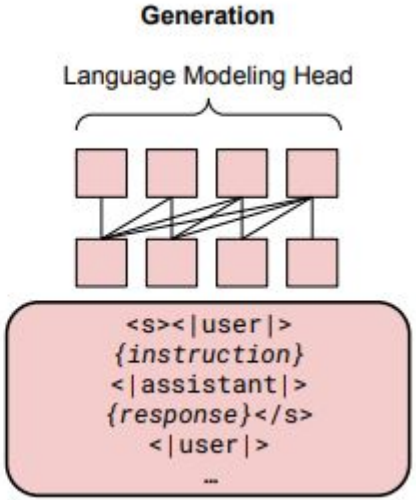
- Depression dataset
  - 8000 negative samples
  - 2000 positive samples



Embeddings - GritLM

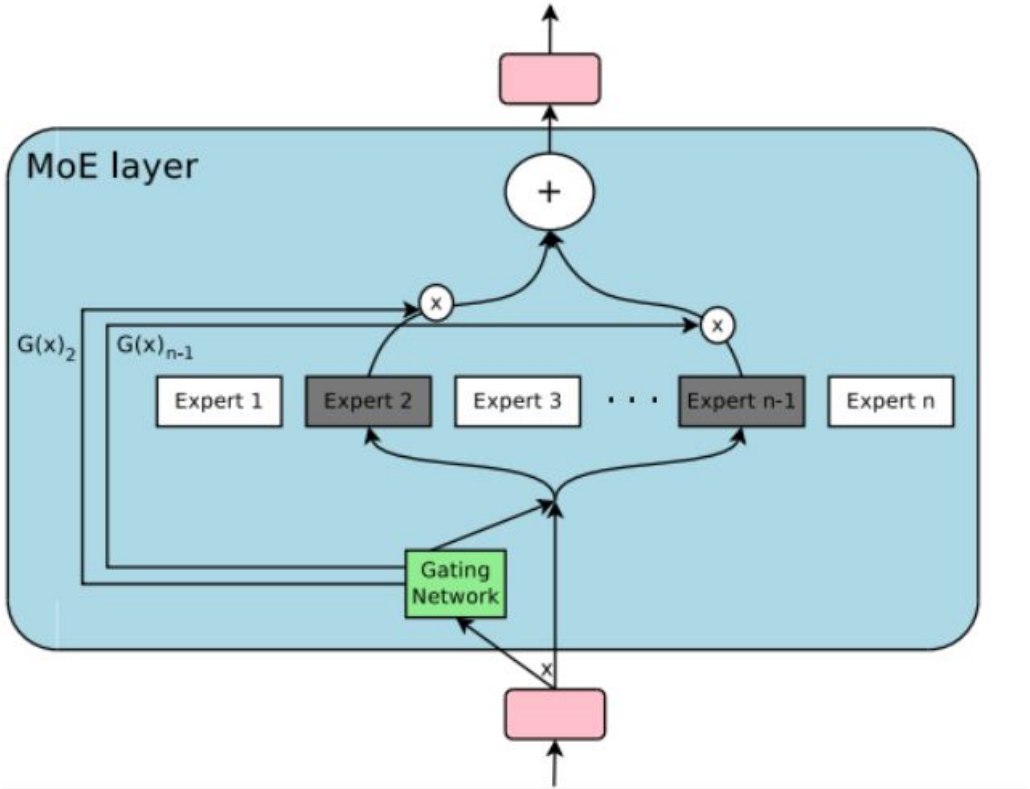
- Based on Mistral
  - No typoes issues
  - Huge model
- Bidirectional attention
  - Better encoding





Model - MLP with MoE layers

- Each layer is a MoE layer
- Hidden size:
  - 8192
  - 256
  - 32
- Boost performance (~2%)

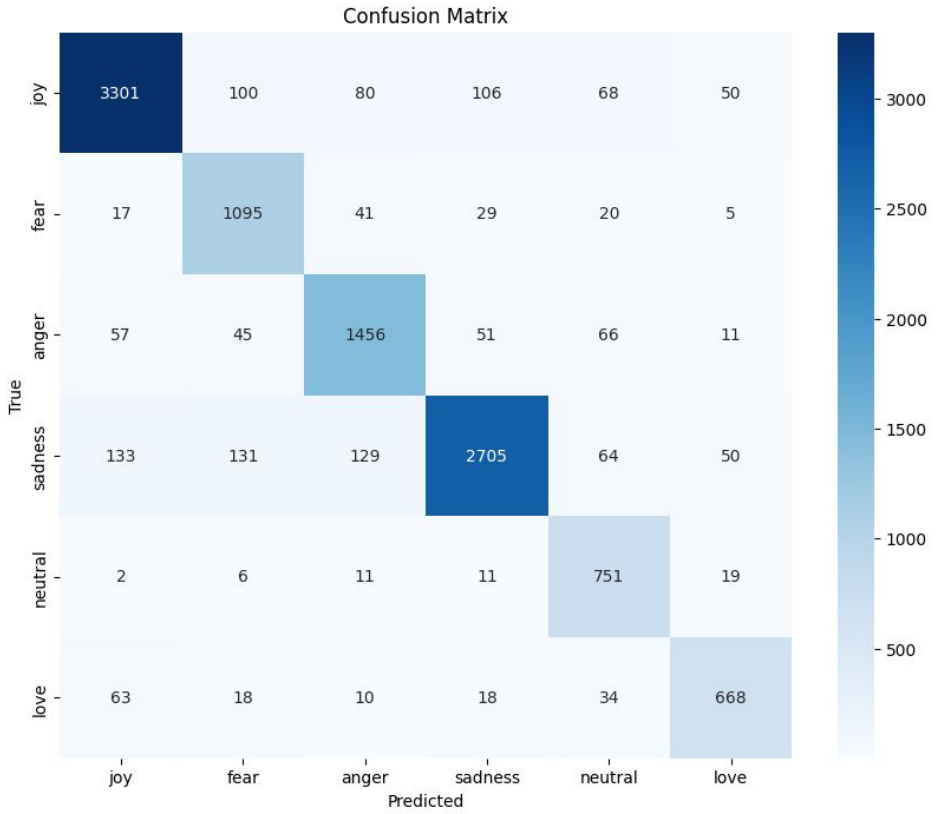


Model - Optimisation

- Weighted cross entropy loss to overcome the unbalanced classes
- Label smoothing to improve generalization
- Dropout on input features to act like masking, in order to improve generalization and remove biases
- Dropout between MoE layers to improve generalizations
- Weight decay to overcome overfitting
- Noise and contrast augmentations to improve generalizations
- Cosine Scheduler to overcome both overfitting and local minimum issues

Emotion detection

- Macro f1: 0.86
- Weighted f1: 0.87



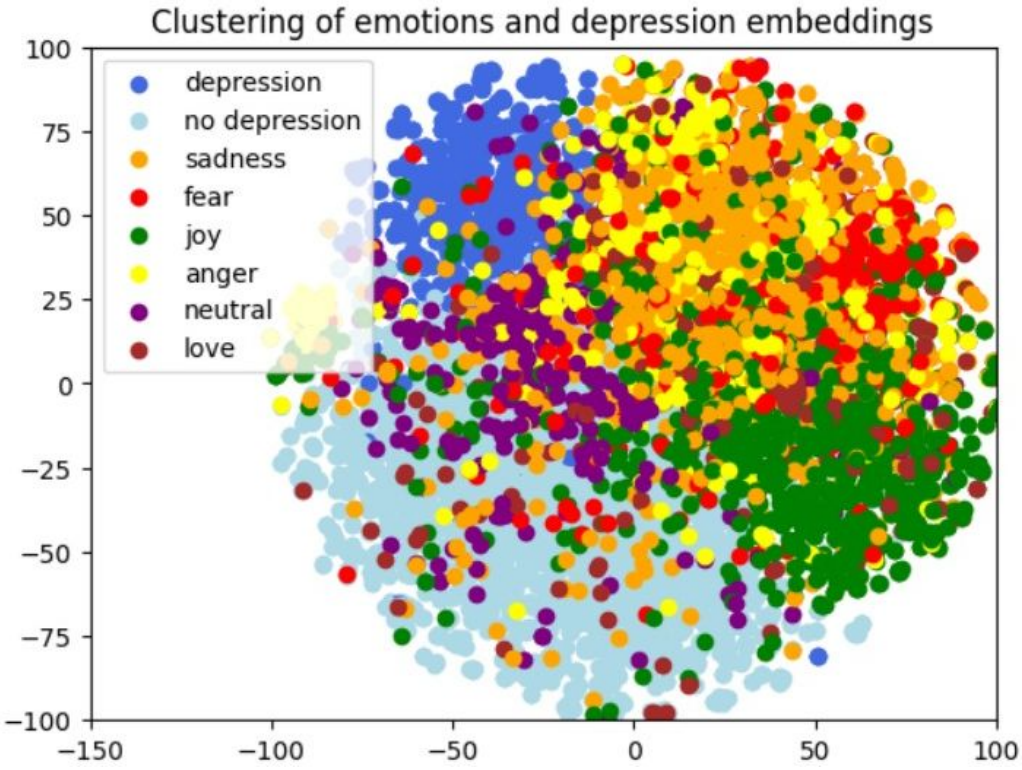
Feature visualization

- We want to see if emotions and depression features are in the same neighbourhood

- Sampled 10000 feature vectors from each dataset (emotion and depression)

- We employ t-SNE method for reducing the dimension from 4096 to 2

Feature Visualization

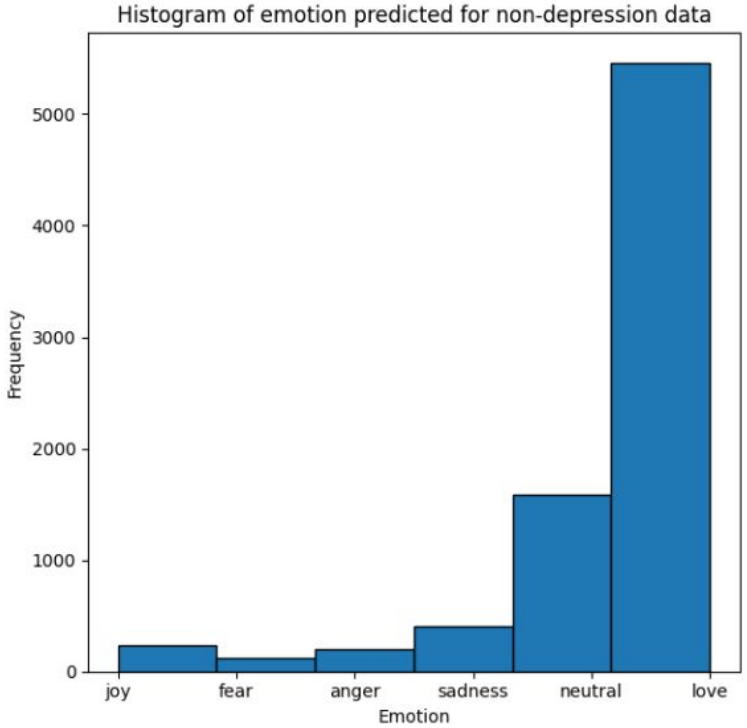


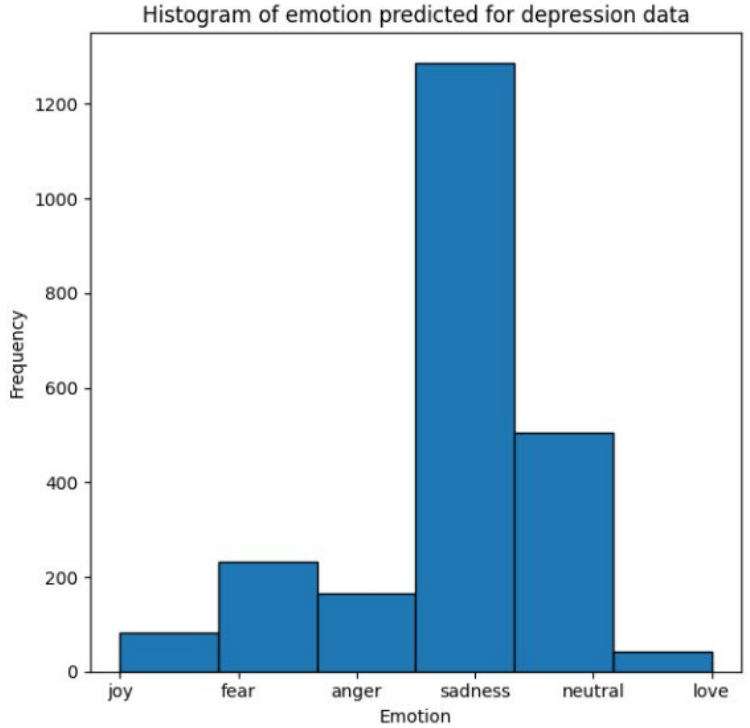
Depression detection

- We want to see what emotions are predicted by he pretrained model for depression data

- Feed the depression data to the model and separate the predicted emotions for both depressive and non-depressive condition

Depression detection





Depression detection

- Next, we want to predict the mental disorder directly. The number data has 4125 train samples, 3094 validation samples, 3095 test samples

- The trained MLP on emotions is used for fine tuning on depression data (except the output layer, which is the classification head)

- The model is fine tunes in the same settings and with the same hyperparameters (except the number of epochs) as the emotion model

- The metrics on the test data:
  - Macro f1: 1.0
  - Weighted f1: 1.0

Depression detection