Compact LSTM-based model for classifying high dimensional multi-modality data

Barath Mohan U, Vidhi Sinha

Problem Definition

- High dimensional data requires feature selection/reduction – usually hand-picked or trial-and-error.
- EEG datasets recorded from human subjects are often limited in size.
- To be able to predict from multi-modal data (EEG, speech, thinking) – need to handle high dimensionality in a data limited scenario.

Technical Details

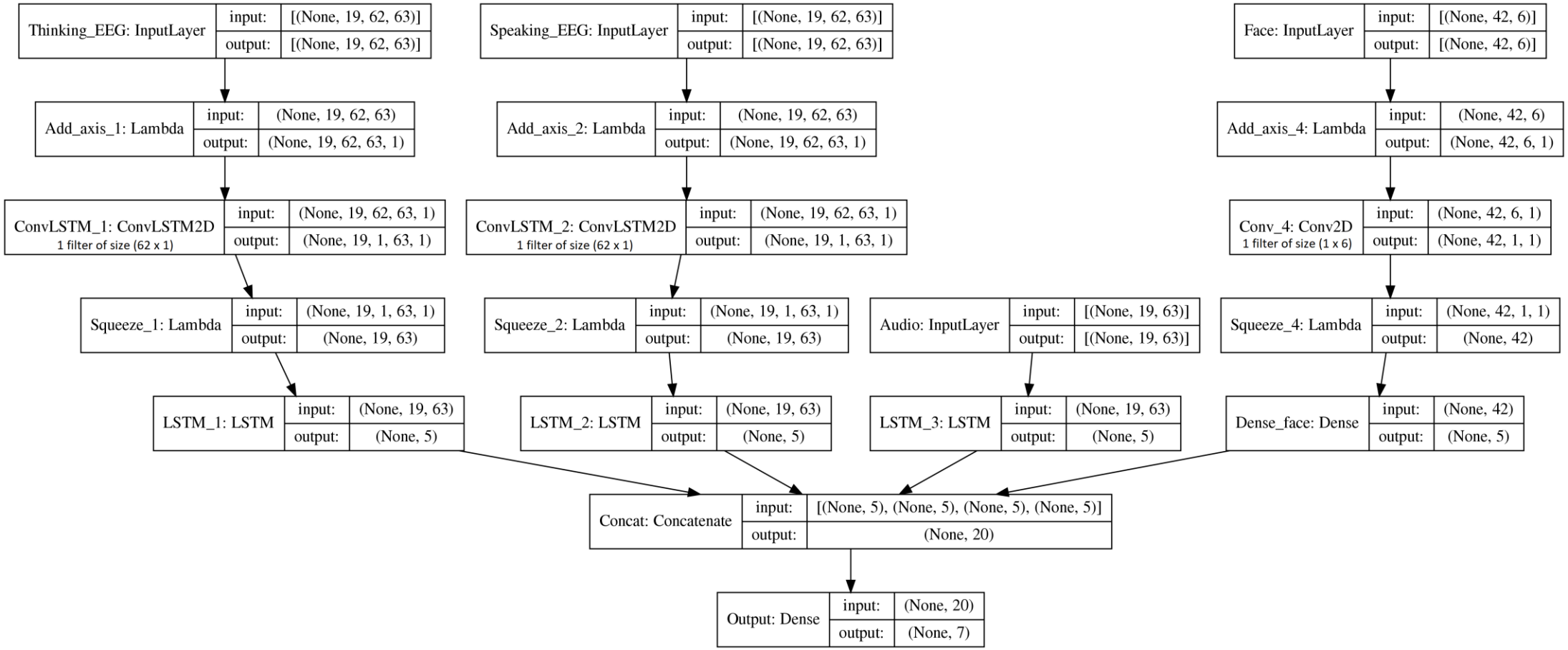
- EEG recording was done by the authors on 14 participants where each participant looked at screen for 30-40 minutes.
- They were shown mono-syllabic prompts(/iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/) and 4 words derived from Kent's list of phonetically-similar pairs (i.e., pat, pot, knew, and gnaw
- 4 stages - rest (relax and clear thoughts), stimulus (auditory utterance of prompt), imagined (imagine the prompt) and speaking (participant spoke it aloud).

Technical Details

- Features like mean, std deviation, median, etc. were calculated along with their 1st and 2nd derivatives. For audio and facial data, similar features were calculated.
- Pearson correlation coefficient was computed between the class labels and data to rank the features and select N€[5,100] of them for final analysis.
- Feed forward neural network (FFNN) with 2 hidden layers of 40 nodes each, having 'elu' activation was trained. The output was softmax activation for 7 classes.
- SVM-rbf kernel model was trained on training set using 'one vs all multi-class method'

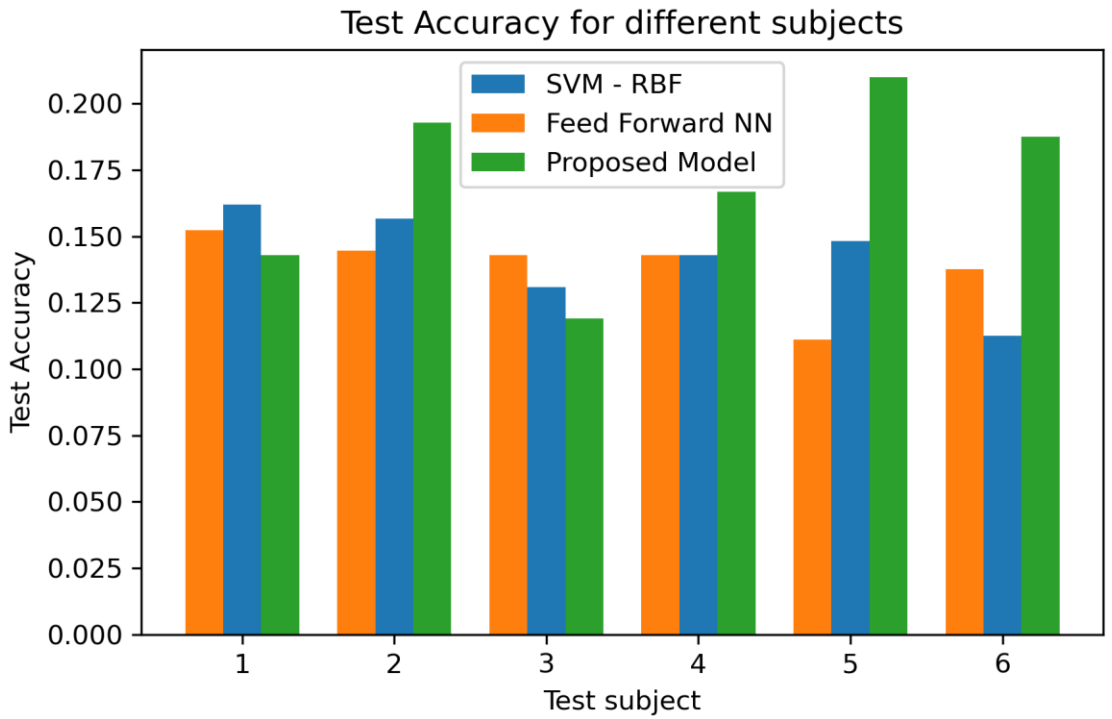
Technical Details

Proposed Model Architecture



Total Parameter = 5509

Results



Model	Average test accuracy
SVM – RBF kernel	0.1422
Feed-forward neural network	0.1385
Proposed Model	0.1698

Novel Contributions

- The proposed model does not require the features to be hand-picked before feeding the data to it.
- It handles the high dimensional data by exploiting the relationship between the features and sharing weights between them, thus reducing the parameters.