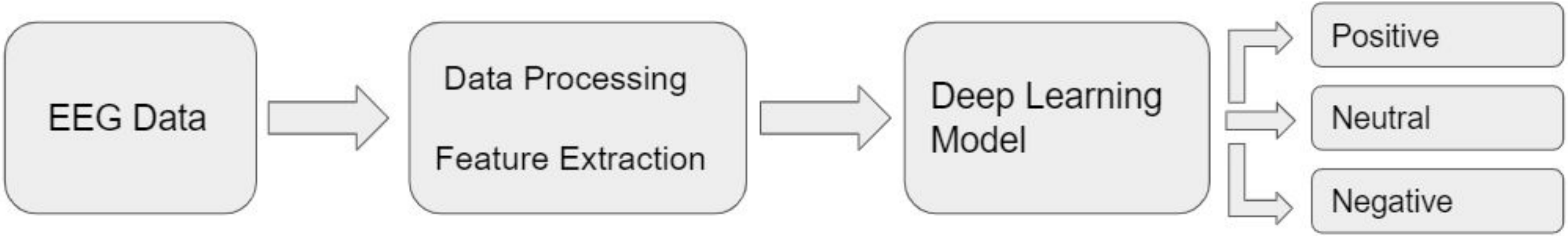
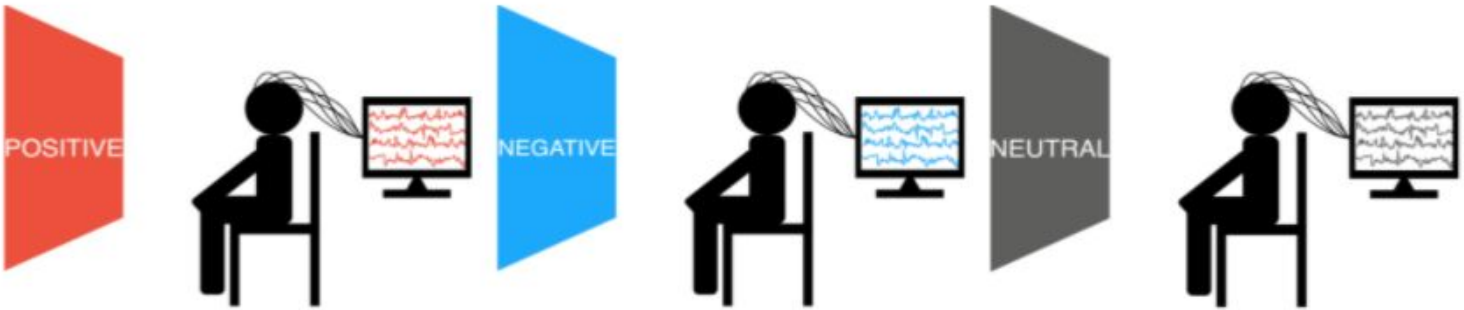
# Deep Learning for Human Emotion Recognition using EEG Recordings

By - Reva, Srimit and Abhishek

# Project description

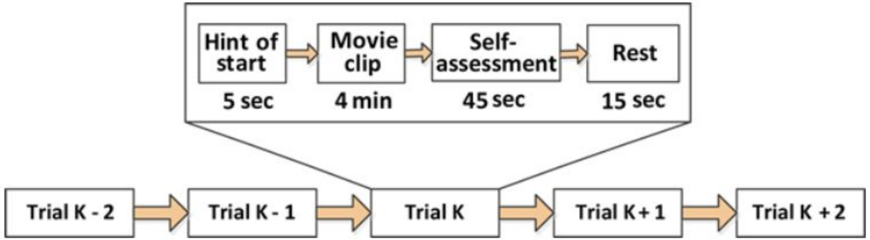
The project is based on application of machine learning and deep learning methods to detect human emotion from electroencephalogram (EEG) signals from SEED dataset and classify as positive, negative or neutral.



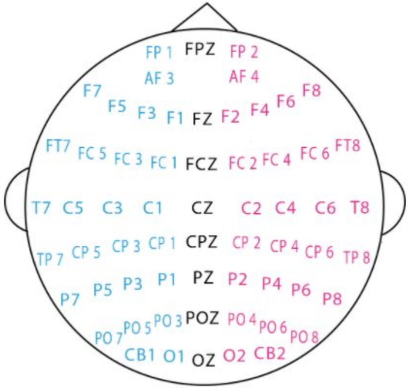
# SEED Dataset

15 Chinese film clips were chosen as stimuli to elicit  positive, neutral or negative emotions. There are totally 15 trials for each experiment.The detailed protocol is shown below:




Fifteen Chinese subjects (7 males and 8 females; MEAN: 23.27, STD: 2.37) participated in the experiments.

The EEG cap according to 10-20 system for 62 channels is as shown.




*images taken from bcmi.sjtu seed dataset description

## Preprocessed data

The dataset consists of 45 .mat files (each file containing 15 trials).
The extracted EEG segment for each trial was downsampled to 200Hz and a bandpass frequency filter of 0-75Hz was applied.
Each trial contains array of shape (channel, data). Number of channels is 62 and datalength vary according to the movie clip.
A array named label contains labels -1 for negative, 0 for neutral and 1 for positive.

## Extracted data

Differential entropy (DE) feature, differential asymmetry(DASM), rational asymmetry (RASM), asymmetry(ASM), differential caudility(DCAU) and power spectral density(PSD).

# PREVIOUS METHODOLOGIES

# Channel Selection Methods:

**Wrapper** - A classification algorithm is used to evaluate the candidate channel subsets, which are generated by a search algorithm. The evaluation of every candidate is obtained by training and testing a specific classification algorithm. Consequently, they are more computationally expensive than filtering techniques and they are prone to overfitting.

**Filtering** - Filtering techniques use an independent evaluation criterion to evaluate the candidate channel subsets, which are generated using a search algorithm. Advantages of the high speed, independence from the classifier, and scalability, but they suffer from low accuracy, because  they do not consider the combinations of different channels

# Channel Selection Methods:

**Embedded** - The selection process is included in the construction of the classifier, and the criteria is used to select the channel in the classifier learning process. Embedded techniques acquire an interaction between the channel selection and the classification.

**Hybrid** - fusion of a filtering and a wrapper attempting to take advantage of the two models to avoid the pre-specification of a stopping criterion.

# Channel Selection Methods (Previous Papers):

**Rizon** - **asymmetric ratio (AR)** based channel selection method for human emotion recognition from EEG signals.

**Zheng -** proposed a method based on **deep neural networks** to learn the average absolute weight distribution to select the **optimal EEG channels**, and obtain preferable experimental results.

**Gupta -** proposed a **flexible analytic wavelet transform (FAWT)** based on six known channels for emotion recognition.

**Bajaj -** proposed **multiwavelets decomposition** based features for EEG emotion classification, and the result is performed well. A new method for emotion recognition using multiwavelet transform with multiclass least squares support vector machine (MC-LS-SVM), and it provided classification accuracy of 84.79% for emotions.

# Feature extraction:

**Hjorth Features:**

- Activity: $A_{\boldsymbol{\xi}} = \frac{\sum_{t=1}^{T} (\boldsymbol{\xi}(t) - \mu)^2}{T}$
- Mobility: $M_{\boldsymbol{\xi}} = \sqrt{\frac{\mathrm{var}(\dot{\boldsymbol{\xi}}(t))}{\mathrm{var}(\boldsymbol{\xi}(t))}}$
- Complexity: $C_{\boldsymbol{\xi}} = \frac{M(\dot{\boldsymbol{\xi}}(t))}{M(\boldsymbol{\xi}(t))}$.

**Non-Stationary Index (NSI):**

The NSI is defined as the standard deviation of all means, where higher index values indicate more inconsistent local averages.

# Feature Extraction:

**Fractal Dimension (FD):**

Several methods like Sevcik’s method , Fractal Brownian Motion , Box-counting , or Higuchi algorithm are used to calculate FD. The Higuchi algorithm is known to produce results closer to the theoretical FD values than Box-counting.

**Higher Order Crossings (HOC):**

Captures the oscillatory pattern of EEG. A sequence of high-pass filters is applied to the zero-mean time series Z(t):

$$\Im_k\{Z(t)\} = \nabla^{k-1} Z(t)$$

# Feature Extraction:

**Band Power:**
Discrete Fourier Transform is used to extract band features.Fast fourier transform (FFT), Short-time fourier transform (STFT) or estimation of power spectra density (PSD) using Welch’s method are the ways to compute DFT.

**Higher Order Spectra (HOS):**
The bispectrum Bis represents the Fourier Transform of the third order moment of the signal. Bicoherence Bic is simply the normalized bispectrum.

**Hilbert-Huang Spectrum (HHS):**
HHS, non-linear method, showed to be more resistant against noise than SPG. HHS for each signal,  is done via empirical mode decomposition (EMD) to arrive at intrinsic mode functions (IMFs) to represent the original signal.

# Feature Extraction:

**Discrete Wavelet Transform:**
Discrete wavelet transform, which decomposes the signal in different approximation and detail levels corresponding to different frequency ranges, while conserving the time information of the signal. Correspondence of frequency bands and wavelet decomposition levels depends on the sampling frequency.

# Models Used:

According to different papers and comprehensive reviews of Deep Learning techniques for Emotion Recognition, the following models are very popular :

1. CNNs (Convolutional Neural Networks)
2. DBNs (Deep Belief Networks)
3. RNNs (Recurrent Neural Networks)
4. A hybrid model of the above

Most popular and efficient function is the ReLU function, used in more than 75% of all papers.

# Models Used:

| | | | | |
|---|---|---|---|---|
| [63] | CNN<br>5 conv layers<br>1 dense layer | Signal values | All channels | 81% |
| [64] | MLPNN<br>4 hidden layers | PSD features | All channels | 82% |
| [65] | RNN<br>2 LSTM layers<br>1 dense layer | Signal values | All channels | 87% |
| [66] | CNN<br>2 conv layers)<br>2 dense layers | Image - Fourier feature maps | Channels combined in image | 87% |
| [67] | CNN<br>2 conv layers<br>1 dense layer | Image - 3D Grid | Channels combined in image | 88% |
| [50] | DBN<br>(3 RBM's)<br>1 dense layer | PSD features | All channels | 89% |

# PRESENT METHODOLOGIES AND ADVANCES

# Present Methodologies and Advances

Implemented in Keras, Plan to implement in PyTorch

- Started with Raw Data on CNNs (Ran out of memory on Colab)
- Proceeded to use PSD features (Problem all features is 10^7)
- Finally settled on using DE (Normalize + extend to same length)

Models -

- Started with CNNs, (maximum accuracy achieved - 87%)
- Progressed to RNNs, (LSTM model achieved - 89%)
- Hybrid model (RNN → CNN → Dense - 93%)
- All models combined, Ensemble Model (97%)

# Challenges Faced

Some of the key challenges faced while training the models were -

- Dataset small in number of examples (only 675 examples)
- Frequently Overfit to training data
- Loss tended to increase after a few epochs (Exploding gradients)
- Running out of memory in Colab
- Accuracy reached saturation at 92% - 93%

Solutions devised for the same -

- Regularization in the form of Batch normalization and dropout implemented
- Used DE features and reduced depth of models to reduce complexity
- Ensemble modelling to increase accuracy past saturation point