Novelties in Intelligent Digital Systems
C. Frasson et al. (Eds.)

© 2021 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA210099

Prediction of Amusement Intensity Based on Brain Activity

Gabrielle TOUPIN [a], Mohamed S. BENLAMINE[a] and Claude FRASSON [a]
[a] *University of Montréal*

Abstract. Amusement can help modulate psychological disorders and cognitive functions. Unfortunately, algorithms classifying emotions still combine multiple positive emotions into a unique emotion, namely joy, making it hard to use amusement in a real-life setting. Here we train a Long-Short-Term-Memory (LSTM) on electroencephalography (EEG) to predict amusement on a categorical scale. Participants (n=10) watched and rated 120 videos with various funniness levels while their brain activity was recorded with an Emotiv Headset. Participant's ratings were divided into four bins of amusement (low, medium, high & very high) based on the participant's ranking's percentile. Nested cross-validation was used to validate the models. We first left out one video from each participant for the final model's validation and a leave-one-group-out technique was used to test the model on an unseen participant during the training phase. The nested cross-validation was tested on sixteen different videos. We created an LSTM model with five hidden layers, vatch size of 256 and an input layer of 14 x 128 (number of electrodes x 1 sec of recording) and four nodes representing the different levels of amusement. The best model obtained during the training phase was tested on the unseen video. While the level of accuracy between the validation videos varies slightly (mean=57.3%, std=13.7%), our best model obtained an accuracy of 82,4%. This high accuracy supports the use of brain activity to predict amusement. Moreover, the validation process we design conveys that models using this technique are transferable across participants and videos.

Keywords. amusement, LSTM, EEG, emotiv, emotions, humour

1. Introduction

1.1. Context & Motivation

Humour is a social behaviour that allows people to break the ice, relax the atmosphere, or gently pass a criticism [1]. It is a complex cognitive process that can result in an emotional state of amusement and can trigger laughter [2]. Research in positive psychology induces amusement to modulate psychological disorders, such as schizophrenia and depression [3-5]. This positive emotion can also benefit cognitive functions such as memory [5,6]. In addition to having different research uses, amusement differs from joy in terms of facial expressions, physiological signals, and feelings [7,8]. Nevertheless, predicting emotions still widely combines these positive emotions together [9]. Only a handful of studies can predict different positive emotions [8,10]. Thus, a better understanding and prediction of amusement would benefit research using amusement as a regulator, as well as new technologies.

The development of new models to predict emotions in artificial intelligence is on the rise. Those algorithms are trained to predict emotions based on facial expressions, electroencephalography (EEG) or physiological signals, such as electrocardiography (ECG). Algorithms using facials expressions to predict emotions can be complicated when used in real-world applications and experiments. First, using a filming process requires specific settings where the participant always faces a camera, making it notably difficult for moving subjects and situations where faces are hidden (e.g., virtual reality headset, wearing a mask, etc.). So far, algorithms based on artificial intelligence, like Emotient, are better than humans at classifying basic emotions when they are typical, exaggerated and static. However, the accuracy drastically drops when used on spontaneous, dynamic and mixed emotions [11-13]. While there is more work to be done in this area, using brain activity and physiology might be a better choice to train algorithms to predict emotion since it does not require the participant to be static in front of the camera and physiological signals cannot be intentionally controlled. With the use of new technologies like Emotiv (https://www.emotiv.com/), where the headset is affordable, requires minimum setup and connects via Bluetooth, new setup experiments and real-life applications are conceivable.

Researchers use different estimators and features to train algorithms to detect emotion based on EEG signals. If we look at the machine learning side, studies use estimators such as support vector machines (SVM), Naive Bayes (NB) and K-nearest neighbours (KNN) to classify emotions [8,10,12]. When looking at deep learning, there is no consensus on which algorithm is best for emotion classification [13]. In their study, Alhagry [14] reaches an accuracy score over 85% with a Long-Short Term Memory algorithm (LSTM) to predict the intensity of arousal and valence of the emotion base on EEG. LSTM is a recurrent neural network (RNN) architecture used in the field of deep learning. This algorithm is promising since it can learn from complex data and predict both on a continuous and categorical scale. Feeding raw EEG data allows us to create algorithms that do not require transformed data, which takes time to compute. Furthermore, LSTM can take more information into account than classical machine learning techniques, meaning that even some artifacts or movements detected by the EEG headset could help define the amusement intensity. Therefore, we propose to train an LSTM algorithm to predict amusement intensity with EEG data acquired with the Emotiv headset. This study brings new insights into the prediction of emotion intensity and amusement.

1.2. Objective

This paper's objective is to develop a deep-learning model that can predict the amusement's intensity of the participant based on its brain activity. We trained an LSTM to predict the categorical score of amusement (low, medium, high, very high) based on one second of brain activity from 14 electrodes. When developing our model, we took special care to ensure that our model was transferable to new participants and new visual content.

2. Methods

2.1. Participants

Ten participants (7 women, 3 men) were recruited for this experiment. They were approached on social media and were offered monetary compensation in exchange for their participation. Recruited participants were between the age of 18 and 30 and had similar education, standard or corrected-to-normal vision and no neurological or psychological disorders. One participant had to be excluded since he never completed the active task (*i.e.* pressing the space bar) during the experiment.

2.2. Material

One hundred twenty video clips were used as humorous stimuli. These video clips were selected in two steps. First, undergrad volunteers selected short portions of humorous and neutral videos from movies, short clips and video compilations. A total of 50 neutral videos and 100 humorous videos were selected. The videos were cropped to have a length between 8 to 12 seconds (mean of 10 seconds). Furthermore, black outlines and the sound, when they were present, were removed from the clip. Second, a preliminary study was conducted in order to validate the selected stimuli. Forty participants watched and rated every video on the following scales: arousal, valence and funniness. Those ratings were used to confirm that there is enough variability in the selected videos. We performed a K-Means clustering with three clusters on arousal, pleasantness and funniness. For the three clusters of funniness, corresponding to neutral, funny and very funny, we selected the best 40 videos of each as stimuli for this study.

2.3. Procedure

Participants arrived at the Functional Neuroimaging Unit and were inquired to read, understand, and sign the consent form. Participants were seated comfortably in a Faraday Cage. There was a screen, a mouse, and a keyboard in front of them to perform the experiment. The task used in this study was created with Psychopy 3 [15] and consisted of four blocks with 30 trials each. Each bloc was designed with a pseudo-randomized order and included ten neutral videos, ten funny videos and ten very funny videos. We made sure that there were at most three videos of the same type in a row. A single trial consists of a fixation cross (2-3 seconds), followed by a video (8-12 seconds), another fixation cross (3 seconds) and a single question ("how funny was this video"). The question was on a scale of 1 (not funny) to 100 (very funny) and was answered with the mouse. An active task was used while watching the video to keep the participant engaged in the task. The participant was asked to press the spacebar on the keyboard when he thinks the video was funny.

The first part of the experiment consisted of a practice block where the participant got familiar with the trial design in the experimenter's presence. The participant was free to ask any question about the trial and the experimenter made sure that the task was understood. The practice block was followed by an emotional questionnaire where the participant evaluated the presence of 20 emotions on a Likert scale of 1-7. Then, a resting state (6 minutes) was measured. During the resting state, the participant was asked to look at the cross in the screen's center and stay neutral. The participant was then ready to start the four blocks of the experiment. Once the participant was ready, he could press

the keyboard's space button to start the block. The experimenter went inside the room between each block and ensured the participant was still in good shape to proceed with the task. It was recommended to take a couple of minutes to relax between each block. After all the blocks were completed, the participant was presented with another resting state and the same emotional questionnaire.

2.4. EEG recording

The Emotiv Epoc headset was used to collect electrical activity during the task. EEGs were recorded from 14 electrodes (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) with two reference nodes located behind the ears. The generated data are in μVolt with a sampling frequency of 128 samples per second. Electrodes were moisturized with a saline solution to maintain electrode impedance under the software's required level. Impedance was checked during the initial installation, followed by a rechecked before the start of each block.

2.5. Data Preparation

2.5.1. Participant Evaluation of Amusement

Only humorous videos were used to develop the model. Since the rating interval differs between the participants, we scaled the rating between 0 and 1 for each participant. The participant's lowest value was converted to 1, his highest value to 100 and every value in between was scaled proportionately. We computed the user's amusement rating level by dividing the rating scale into three intervals: [0..0.25[for low amusement, [0.25..0.50[for medium amusement, [0.50..0.75[for high amusement and [0.75..1[for very high amusement.

2.5.2. Time of Interest

Funniness appears mostly at the videos' end [16], leading us to choose the video's end as the time of interest for humorous videos (Figure 1). More precisely, if the participant pressed the spacebar to indicate that it is indeed a funny clip, we only used EEG data between the button press and the end of the video and assigned the user's reported amusement rating. On the other hand, if no button was pressed, particularly for less funny videos, we assumed that the reported funniness was stable across the video and used EEG data associated with the full video's length.

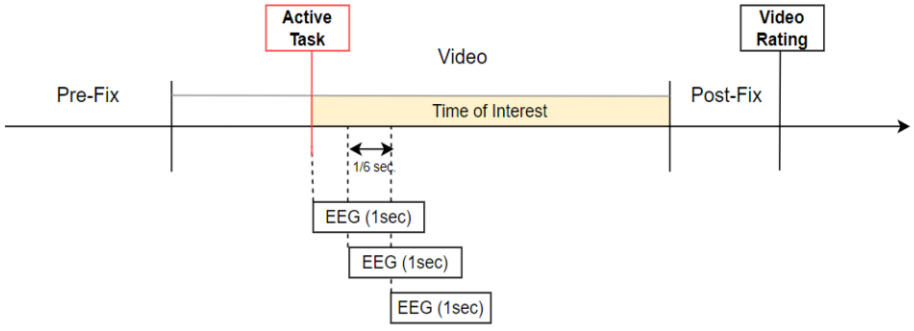

Figure 1. Task performed by the participants.

2.5.3. EEG Data Cleaning

EEG data collected via Emotiv were cleaned via the python's MNE library. Emotiv has a sampling frequency of 128hz per second. The first cleaning part is done automatically by the Emotiv Software, where it uses a 5th-order digital Sinc to filter between 0.2 Hz and 45 Hz. Plus, it uses a Notch filter at 60Hz since it is the frequency band for North America's electricity. It also removes most of the eye's blinks and heartbeat from the signal. Additionally, we complemented the cleaning from Emotiv Software with an additional process done with python's library MNE. To validate that all eyes and cardiac artifacts were well removed, we decomposed the EEG signal using an Independent Component Analysis (ICA). ICAs that were strongly correlated with either eye blinks or heartbeat were removed from the signal. Finally, we manually observed the signal of each participant and annotated then noisy parts of the signal. Epochs with those annotated parts were not used in further analysis.

2.5.4. Data Selection

Our model will attempt to predict the user's amusement rating from a 1 second EEG data from all electrodes. We used a data matrix of shape 128x14 which holds 1 second of recording for each of the 14 electrodes (Figure 1). This second of recording was associated with the rating of the participant for this specific video. For the length of the trial's time of interest, we move the data matrix 1/6 second in time and assign the participant's rating to the data matrix.

2.6. Model Training and validation

Deep-Learning models learn data representations within their hidden layer at multiple levels of abstraction [17]. We have constructed an LSTM model with 5 hidden layers (four layers of 14* 128 neurons and the fifth layer of 128 neurons), with a batch size of 256 and an input layer of 14*128 (see Figure 2). The Network output layer has four nodes representing the amusement level.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 14, 128)	131584
lstm_1 (LSTM)	(None, 14, 128)	131584
lstm_2 (LSTM)	(None, 14, 128)	131584
lstm_3 (LSTM)	(None, 14, 128)	131584
lstm_4 (LSTM)	(None, 14, 128)	131584
lstm_5 (LSTM)	(None, 128)	131584
dropout (Dropout)	(None, 128)	0
dense (Dense)	(None, 4)	516

Figure 2. Neural network summary

To make sure our model can generalize on new content, we extracted the data of one video from all participants as our final test of the generated model. To ensure our validation accuracy is unbiased by the chosen video, we repeated our model training with 16 different videos and discussed the results below. Furthermore, to ensure the model is

usable on an unseen participant, we used a leave-one-group-out (LOGO) technique during the training and testing phase (Figure 3). More precisely, we trained the algorithm on all 8 participants and tested it on the last one not previously seen by the algorithm. We repeat this procedure so that each participant is used as the test set once. The mean accuracy of all algorithms can describe the algorithms' performance.

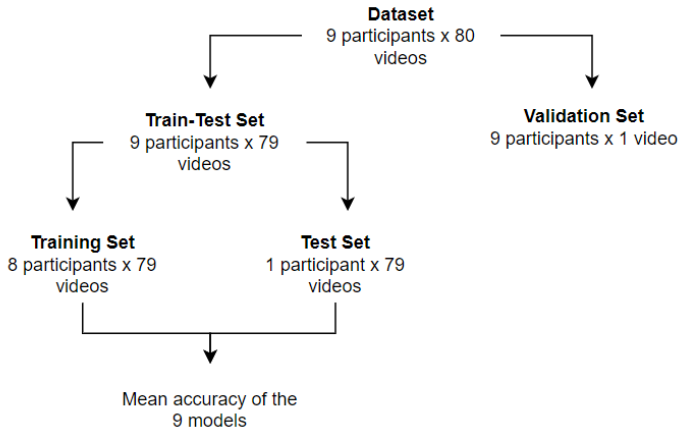

Figure 3. Leave-One-Group-Out cross-validation method

Initially, we set the number of training epochs to 100 by cross-validation. We have used early stopping techniques to prevent model overfitting (Figure 4). We set model monitoring on validation loss metric during training. Early stopping was used to evaluate different learning rate values for the model. The weights of the best model were recorded with minimum validation loss.

```
    earlyStopping = EarlyStopping(monitor='val_loss', patience=10,
verbose=0, mode='min')

    mcp_save = ModelCheckpoint('SavedModels/mdl_clf_wts.hdf5',
save_best_only=True, monitor='val_loss', mode='min')

    reduce_lr_loss = ReduceLROnPlateau(monitor='val_loss', factor=0.1,
patience=7, verbose=1, epsilon=1e-4, mode='min')
```

Figure 4. Early stopping code snapshot

3. Results and Discussion

3.1. Base Model

3.1.1. Generalization of the model

The validation accuracy of the base model, when tested on an unseen video can be found in Table 1 under *validation accuracy*. Taken together, our models predict an

unseen video with an average of 64.2% (std=14.7%) with a maximum accuracy of 88.9% (model #1) and a minimum of 32.9% (model #6). Since there is high variability in the validation accuracy, we cannot conclude that this specific algorithm can yet be transferable to other content. On the other hand, when the algorithm is tested on an unseen participant during the training phase, accuracy is more stable. The column Mean Accuracy Training (STD) of Table 1 shows the model's mean accuracy when tested on each of the unseen participants (n=9). We obtain a mean accuracy of 74.9% (std=3.8%) with an accuracy as high as 87.5% (model #1) and as low as 71.5% (model #9). This high accuracy and low standard deviation show that our model can predict the amusement level based on a participant's brain activity that it has never seen before.

While looking at each model's confusion matrix, we saw that the fourth class, namely *very-high amusement*, is well represented in none of the models (see example in Figure 5), which may cause the high variability observed in the validation accuracy. Furthermore, the good results obtained might be due to overfitting on those unbalanced categories.

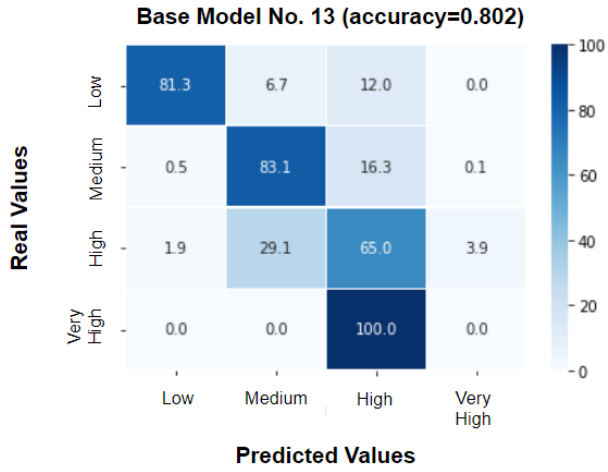

Figure 5. Example of a confusion matrix during validation phase where very high amusement is unwell represented

3.2. Model with Class Weight

3.2.1. Weight Classes

To overcome the fact that our classes are unbalanced, we assign each class a weight during the training phase. We used an automatic function that looks at the distribution of labels and produces weights to equally penalize under or over-represented classes in the training set. While each training model had different weights, the weight was very similar between models. A mean weight of 1.339 (std=0.014) was assign to low amusement, (std=0.009) to medium amusement, 0.591 (std=0.004) to high amusement and 6.284 (std=0.298) to very high amusement.

3.2.2. Generalization of the model

Out of the 16 models we trained, 10 of them saw their validation accuracy drop after the model was adjusted with weights (Table 1 under Validation Accuracy Gain/Loss). This confirms that our base model was overfitting on most of the models. We obtained a mean validation accuracy of 57.3% (std=13.7%). Like our base model, this high variability across the accuracy confirms that our model is not yet able to transfer perfectly to unseen videos. Our best model obtains an accuracy of 82.4% and our lowest is at 31.1%.

Training and testing accuracy with the adjusted weights also dropped for all models. During the training and testing phase. We obtained a mean accuracy of 63.1% (std=3.0%) where our best model has an accuracy of 72.6% (std=13.3%) and our worst model has an accuracy of 59.6% (std=15.5%). The mean accuracy is still above the theoretical chance level for four classes (chance=25%) and constant across our model. This low accuracy variability supports that our LSTM model can transfer across unseen participant's brain activity.

Table 1. Model Generalization results

	Base Model		Models with Weight		
Validation Video	**Validation Accuracy**	**Mean Accuracy Training**	**Validation Accuracy**	**Validation Accuracy Gain/Loss**	**Mean Accuracy Training**
1	0.3602	0.7336 (0.143)	0.3115	-0.0487	0.634 (0.152)
2	0.5584	0.7489 (0.140)	0.5121	-0.0463	0.611 (0.157)
3	0.5339	**0.7912 (0.135)**	0.4899	-0.0440	0.602 (0.154)
4	0.5588	0.7420 (0.133)	0.6110	0.0522	0.616 (0.154)
5	**0.7443**	0.7264 (0.133)	0.6018	-0.1425	**0.726 (0.133)**
6	0.3289	0.7328 (0.139)	0.3537	0.0248	0.629 (0.159)
7	0.5169	0.7466 (0.132)	0.5560	0.0391	0.622 (0.149)
8	0.5086	0.7229 (0.126)	0.5328	0.0242	0.620 (0.152)
9	0.4874	0.7155 (0.136)	0.3923	-0.0951	0.596 (0.155)
10	0.6858	0.7286 (0.131)	0.6344	-0.0514	0.643 (0.137)
11	**0.7596**	0.7372 (0.135)	0.6795	-0.0801	0.606 (0.155)
12	**0.7401**	0.7339 (0.134)	0.6450	-0.0951	0.645 (0.150)
13	**0.8015**	0.7290 (0.133)	**0.8241**	0.0226	0.632 (0.140)
14	0.6978	**0.7577 (0.137)**	**0.7326**	0.0348	0.634 (0.144)
15	**0.7840**	**0.7714 (0.149)**	0.6228	-0.1612	0.616 (0.153)
16	0.6786	0.7251 (0.134)	0.6687	-0.0099	0.667 (0.161)

3.2.3. Best model

Across our models, only one seems to both transfer across unseen videos and unseen participants. Our best model (model #13) can accurately predict the amusement level (low, medium, high, high amusement) of an unseen video based on the participant's brain activity with 82.41% accuracy. This high accuracy suggests that brain activity collected with a commercial headset can be used to predict amusement.

While our model reaches a high accuracy level, we can see from the confusion matrix (Figure 6) that our model still has difficulty distinguishing between high and very high amusement. Our model can accurately predict the low and medium levels of amusement, but high and very high amusement are still inadequately predicted. It is

possible that the model cannot classify between high and very high because the brain activity is more alike in those two categories than in low and medium amusement.

Inspired by Liu [8], we believe that this problem could be resolved by first creating a model that classifies the data between three types of funniness: low, medium and high (where high is a combination of high and very high amusement). This would be followed by a second model trained to classify especially between high and very high amusement, thus increasing our model prediction.

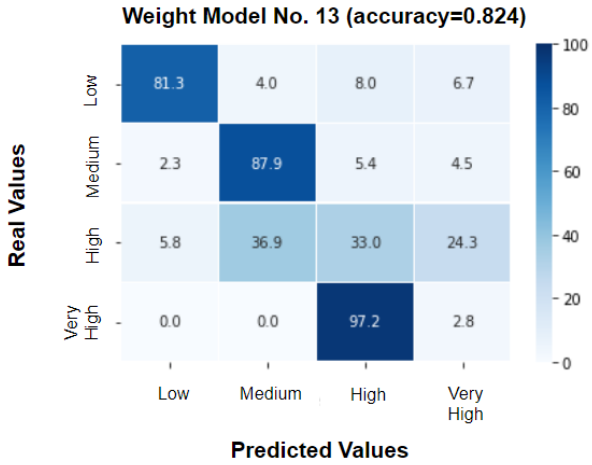

Figure 6. Confusion matrix of best weighted model during validation phase

4. Conclusion

In this study, we aimed to develop an algorithm that predicts amusement based on EEG data from a commercial headset. The objective of this paper was to develop a model that can predict amusement with high accuracy while ensuring that it is transferable across both new participants and new contents. Using an LSTM algorithm, we were able to obtain a model that can predict amusement with an accuracy of 82.4%. This high accuracy confirms that brain activity can accurately predict amusement experienced by the subject. While our model had, on average, a low variability when testing on unseen participants, models tested on unseen videos were more variable. This lets us believe that we can still improve our model. Classification of amusement in four-level (low, medium, high and very high) is our first step into creating a deep learning model that can predict amusement. In this study, we support both the use of EEG data and LSTM to predict amusement.

In our futures research, we want to improve our classification model by first creating a model that classifies the data between three types of funniness: low, medium, and high (where high is a combination of high and very high amusement). A second model would then be trained to classify between high and very high amusement. Furthermore, using the same nested-cross-validation, we will train an LSTM algorithm to predict a value between 0 (not funny) and 1 (very funny) on a continuous scale.

5. Acknowledgments

We acknowledge NSERC-CRD (National Science and Engineering Research Council Cooperative Research Development), Prompt, and BMU (Beam Me Up) for funding this work.

References

[1] Martin, R.A. (2007). The Social Psychology of Humor. In The Psychology of Humor, (Elsevier), pp. 113–152.
[2] Vrticka, P., Black, J. M., & Reiss, A. L. (2013). The neural basis of humour processing. Nature Reviews Neuroscience, 14(12), 860.
[3] Gelkopf, M., Kreitler, S., & Sigal, M. (1993). Laughter in a psychiatric ward. Somatic, emotional, social and clinical influences on schizophrenic patients. Journal of Nervous and Mental Disease, 181(5), 283–289.
[4] Hirosaki, M., Ohira, T., Kajiura, M., Kiyama, M., Kitamura, A., Sato, S., & Iso, H. (2013). Effects of a laughter and exercise program on physiological and psychological health among community‐dwelling elderly in Japan: Randomized controlled trial. Geriatrics & gerontology international, 13(1), 152-160.
[5] Zhao, J., Yin, H., Wang, X., Zhang, G., Jia, Y., Shang, B., ... & Chen, L. (2020). Effect of humour intervention programme on depression, anxiety, subjective well‐being, cognitive function and sleep quality in Chinese nursing home residents. Journal of Advanced Nursing, 76(10), 2709-2718.
[6] Savage, B. M., Lujan, H. L., Thipparthi, R. R., & DiCarlo, S. E. (2017). Humor, laughter, learning, and health! A brief review. Advances in Physiology Education.
[7] Herring, D. R., Burleson, M. H., Roberts, N. A., & Devine, M. J. (2011). Coherent with laughter: Subjective experience, behavior, and physiological responses during amusement and joy. International Journal of Psychophysiology, 79(2), 211-218.
[8] Liu, Y. J., Yu, M., Zhao, G., Song, J., Ge, Y., & Shi, Y. (2017). Real-time movie-induced discrete emotion recognition from EEG signals. IEEE Transactions on Affective Computing, 9(4), 550-562.
[9] Kim, J. H., Kim, B. G., Roy, P. P., & Jeong, D. M. (2019). Efficient facial expression recognition algorithm based on hierarchical deep neural network structure. IEEE Access, 7, 41273-41285.
[10] Aydın, S. (2019). Deep Learning Classification of Neuro-Emotional Phase Domain Complexity Levels Induced by Affective Video Film Clips. IEEE Journal of Biomedical and Health Informatics, 24(6), 1695-1702.
[11] Stöckli, S.; Schulte-Mecklenbeck, M.; Borer, S. & Samson, A.C. (2018) Facial expression analysis with AFFDEX and FACET: A validation study. Behavior Research Methods, 50 (4), 1446-1460.
[12] Duan, R. N., Zhu, J. Y., & Lu, B. L. (2013, November). Differential entropy feature for EEG-based emotion classification. In 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER) (pp. 81-84). IEEE.
[13] Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: a review. Journal of neural engineering, 16(3), 031001.
[14] Alhagry, S., Fahmy, A. A., & El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. Emotion, 8(10), 355-358.
[15] Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. Behavior Research Methods. 10.3758/s13428-018-01193
[16] Barral, O., Kosunen, I., & Jacucci, G. (2017). No need to laugh out loud: Predicting humor appraisal of comic strips based on physiological signals in a realistic environment. ACM Transactions on Computer-Human Interaction (TOCHI), 24(6), 1-29.
[17] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. 521(7553): p. 436-444.