# 590U UbiComp Project
# Confused or Not: Analyzing difficulty in understanding videos

**Saranya Krishnakumar**
UMass Amherst
saranyakrish@umass.edu

**Chesta Singh**
UMass Amherst
csingh@umass.edu

## ABSTRACT

Unlike classroom education, immediate feedback from the student is less accessible in Massive Open Online Courses (MOOC). Using a 14 channel headset to detect a students mental state can be valuable to improve the content of the course materials. We conducted a study on 9 participants while watching MOOC video content and collected the raw EEG data. We trained various machine learning models for classification such as XGBoost, Logistic regression, LSTM, etc. and obtained a maximum accuracy of 81% with XGBoost. We also compared the relative difficulty of MOOC videos available with subtitles with respect to the videos available in english language.

## General Terms

EEG signals, Emotiv Epoc+, Brain Confusion, EEGLAB

## INTRODUCTION

MOOCs (Massive Open Online Course) are the biggest development in distance education by allowing unlimited participation and open access via the web. According to the study done by Class Central for 2017 [2], the MOOC landscape has grown to include 9,400 courses and more than a dozen graduate degrees. The total number of MOOC learners has increased from 58 million in 2016 [1] to 81 million in 2017 [2]. However, there are certain drawbacks with this form of learning. Lack of immediate feedback to the lecturer when students encounter difficulty is one of the major deterrent to learning. In an in-class environment, a lecturer can identify from the body language and gestures of the students (e.g., furrowed brow, head scratching, etc.) when students are confused and this is absent in online education. This is mitigated in some degree by using discussion forums where students can raise doubts after lectures. Additionally more than 75% of MOOC courses are presented in the English language, when, native English speakers are a minority among the world's population [9]. Research studies show that some English Language Learners (ELL) prefer to take MOOCs in English, despite the language challenges, as it promotes their goals of Economic, Social, and Geographic mobility [10]. The ELLs struggle when video content lacks corresponding visual supporting materials [8] (e.g., an instructor narrating instruction without text support in the background), or due to their own hesitation to participate in MOOC discussion forums [3]. In such scenarios, discussion forums also prove inefficient to resolve or detect the confusion in student about course content. In this project we address this limitation by using electroencephalography (EEG) from a commercially available device to measure a students mental state to detect confusion.

The EEG signal is a voltage signal that can be measured on the surface of the scalp, arising from large areas of coordinated neural activity manifested as synchronization (groups of neurons firing at the same rate) [5]. This neural activity varies as a function of development, mental state, and cognitive activity, and the EEG signal can measurably detect such variation. The availability of simple, relatively cheaper, portable EEG monitoring devices now makes it possible to take this technology to the schools. Emotiv Epoc+ is an example of this and we used it in our study.

The Emotiv Epoc+ has 14 EEG channels for accurate spatial resolution. A previous study [11] has been performed using MindSet which is a cheaper device but has a single channel EEG sensor. The high cost and difficulty in setting up may be a deterrent to conduct this study for a large number of students in an open environment. However, the MOOC content providers (such as Stanford University, Peking University etc) can conduct a study on a select group of students using EEG devices to get feedback on their course materials. The business model for MOOC has evolved and made it a monetary profitable enterprise. This makes it a relevant problem to help in user engagement.

Our project consists of collecting EEG data from different participants while watching MOOC videos and training a classifier for predicting confusion. Our specific contributions in this regard are :

- We used a 14 channel EEG headset which is much more accurate when compared to the state-of-the art data as collected by Wang et al. [11].
- Our dataset contains an equal number of subtitled content to address the feedback limitation obtained from English Language Learners students.

*UbiComp '13*, UbiComp Conference

- We performed an experiment to determine if the subtitled content was more confusing for the students as compared to the content in the English language (in which they are very proficient). This experiment justifies our hypothesis that more feedback is needed to make the MOOC content equally accessible to non-native English speakers.

## RELATED WORK

There are many researchers applying machine learning methods to EEG data to accomplish different tasks. Yeo et al. [12] used Support Vector Machines (SVMs) to detect the drowsiness of car drivers. Their results showed that extracting features from four EEG frequency bands achieved 99.3% accuracy. Besides drowsiness, Subashi et al. [7] applied SVM classifiers to predict if EEG signals represented epileptic seizures and achieved 100% accuracy. Wang et al. [11] showed the possibility of using EEG data to detect the confusion of students when they watch MOOC videos. This EEG data was collected by the single channel Mindset headset. They analyzed the EEG data using Gaussian Naive Bayes classifiers. Gaussian Naive Bayes classifiers achieved a classification accuracy of 57%.

Recently, deep learning has shown its promise on many classification related tasks when compared with traditional machine learning approaches. On the same dataset as used by Wang et al. [11], Zhaoheng Ni et al. [4] detected confusion which is a symptom of Brain Fog. They applied many deep learning models including Bidirectional LSTM Recurrent Neural Networks to the EEG data and achieved state-of-art performance with Bi-LSTM. Their classification accuracy is 73.3%. They also concluded that Gamma wave signal is more important in detecting confusion. For our experiment with only 80 data points, we performed experiments only on shallow networks with fewer hidden layers.

## USER STUDY AND DATA COLLECTION

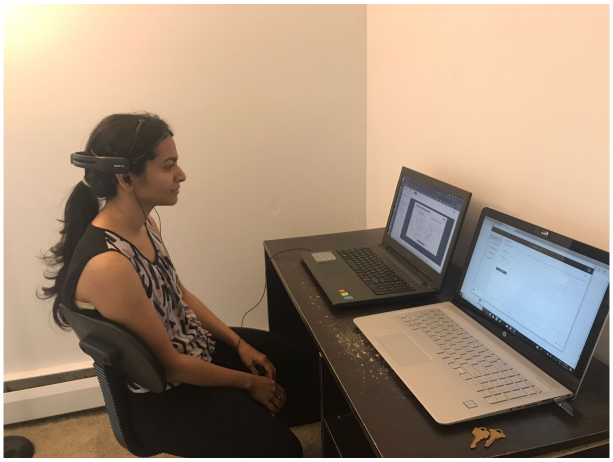


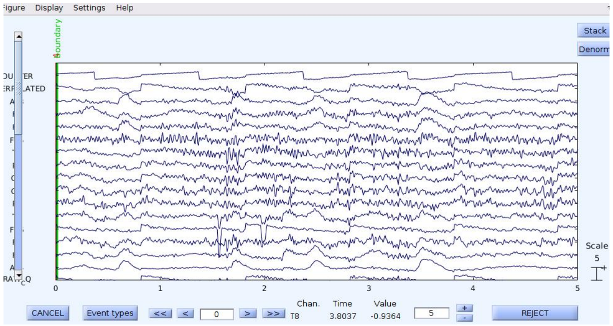
Figure 1. Data collection setup

### Apparatus

We used Emotiv Epoc+ with 14 channels : AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 and two reference nodes at P3/P4. The different electrodes were saturated with saline and placed on the scalp of the participant. Careful adjusting of the reference nodes allowed us to get 100% contact quality. The Emotiv Epoc+ provides a free SDK which shows the contact quality while recording so the headset can be adjusted again if contact quality drops in the experiment mid way.

We conducted this study on a total of 9 participants. *Figure 1* shows the setup used for collecting EEG data. All of these participants were proficient in English so we picked videos from other languages such as Arabic, French, Chinese, etc. with subtitles in English. We believe this is a good heuristic to imitate the scenario where non English speaking students watch MOOC videos in English language subtitled with their native language. All the participants were graduate students in STEM. To facilitate confusion and not completely lose their attention, we picked educational content from STEM courses such as fluid engineering, sciences, maths, etc. We collected 80 video clips of 2 minutes duration from Coursera, edX, MIT ocw and youtube. Four kind of video clips are collected: easy videos in English, confusing videos in English, easy videos with subtitles, confusing videos with subtitles. Easy clips are usually picked from introductory undergraduate classes when the lecturer introduces a topic. Confusing clips are from graduate level courses and often from the middle of a lecture. After the EEG data for each video was collected, the participants were asked to rate if they found the video easy or confusing. This is used as **ground truth** for our experiment.

### Data Processing



Chan. T8 Time 3.8037 Value -0.9364 5

Figure 2. Cleaned EEG signal

We trained a number of baseline models such as SVM which require clean EEG signals for higher accuracy. We used EEGLAB, a matlab tool developed by Swartz Center of Computational Neuroscience (SSCN) . There is not one standard pipeline to remove noise from EEG data . We adapted the steps recommended by Makoto Miyakoshi of SSCN for our project. The EEG data from the headset is of the format .EDF. This data is then imported to the EEGLAB and is passed through the data processing pipeline. An added step while using Emotiv Epoc+ is to remove the DC offset. However, this step is managed when we remove the baseline. After removing the baseline drift, we remove and then interpolate bad channels. An alternative step could be to not use interpolation and this is recommended for highly sensitive data. However, our models include deep learning models with fixed feature length so we chose interpolation. Then, we performed two steps for removing artifacts :

- We performed ASR (Artifact Subspace Reconstruction) on our EEG data for removing occasional large-amplitude noise

- We used ICA (Independent Component Analysis) for decomposing constant fixed-source noise.

Usually, brain-generated Independent Components (IC) are well modeled using one equivalent dipole or, in the case of IC scalp maps that appear bilaterally symmetric, with two position-symmetric dipoles [6] As suggested in the pipeline, we estimated single equivalent dipoles and then symmetrically bilateral dipoles. After these steps, our data is ready for training and is exported in the .mat format.
*Figure 2*, shows the cleaned EEG signal in the EEGLAB window. *Figure 3* shows the channel spectra of a student while watching easy and confusing videos. The is plotted for a duration of 25 seconds and the scalp maps are shown after 6, 10 and 22 seconds. As we can see that the EEG signal has more intensity for difficult videos.

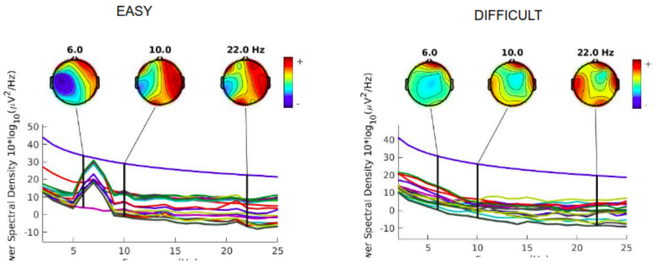


**Figure 3. Channel spectra of a student for Easy(left) and Confusing(right) video**

### Train, validation and test datasets

We collected 160 minutes of EEG data. We used the following split for our train, test and validation datasets :

1. Train set (total duration : 128 minutes)
   - 32 english videos
   - 32 subtitled videos
2. Validation set (total duration : 16 minutes)
   - 4 english videos
   - 4 subtitled videos
3. Test set : (total duration : 16 minutes)
   - 4 english videos
   - 4 subtitled videos

We ensured that an equal number of easy and confusing data is present in all the datasets.

## FEATURE ENGINEERING

We performed our experiments on two different kind of models. We, therefore, perform feature extraction in two ways,

- Extract hand engineered features from raw EEG data by windowing

- Automatically learn features from raw EEG using deep neural networks

The hand engineered features are used for training our baseline models including SVM, KNN, Decision Tree etc. The raw EEG data is fed to the CNN and LSTM models and these deep architectures learn the features and perform classification. We used 1-D convolutional filter for analyzing EEG data and attached a softmax layer in the output for classification.
We used overlapping windows for extracting hand crafted features. Each window contains data of 2.5 seconds and there is an overlap of 1.25 seconds with the next window. This overlap ensures local peaks in EEG data are not lost. The hand engineered features extracted are shown in the table.

| Features Extracted within Window |
|---|
| mean EEG signal |
| min EEG signal |
| max EEG signal |
| standard deviation of EEG signals |
| number of peaks in EEG signal |
| mean peak EEG signal |
| peak at 25 percentile |
| peak at 50 percentile |
| peak at 75 percentile |

When a person is confused, there are more number of peaks in EEG data as compared to when he is not confused. Similarly the maximum EEG signal also varies. Hence these features are a good representation of the raw EEG data, and can be used for training our baseline models.

## METHODS

We follow a supervised learning approach, since our problem is a binary classification problem. We train both traditional machine learning models and deep learning models for predicting whether a given student is confused or not.

### Baseline Models

This table consists of all the baseline models that we trained.

| Baseline Models trained |
|---|
| SVM with sigmoid kernel |
| SVM with rbf kernel |
| K Nearest Neighbors Classifier |
| Decision Tree Classifier |
| Random Forest Classifier |
| Logistic Regression Classifier |

We used a 5 fold cross-validation to evaluate our models. For SVM classifier with different kernels, we found the best parameters of the margin were $C$=10. For KNN classifier we experimented with different number of neighbors and found taking 5 neighbors is best. For decision tree classifier, we trained decision trees of depth 2. For random forest classifier, we trained random forest trees of depth 5 and number of sub estimators (sub trees for forming forest) is set as 4. We are using *L2* loss for logistic regression classifier.

### XGBoost

XGBoost is an implementation of gradient boosted decision trees. The main advantages of XGBoost classifier when compared to traditional models are :

- Faster Execution time
- Sparse awareness i.e handles missing values
- Supports parallelization in tree construction
- Continued Training

### CNN and Dilated CNN Model

We created two CNN architectures from scratch and trained them on raw EEG signals. One of the architectures is created using dilated convolutional filters. Dilated convolutions increase receptive field size using fewer parameters and it can be trained faster. The other architecture contains a 1-D convolutional filter. For both these architectures we used a softmax layer for classification with cross entropy loss:

$$L_i = -\log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$$

**Figure 4. Cross entropy loss**

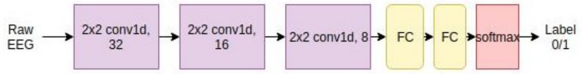


**Figure 5. CNN architecture**

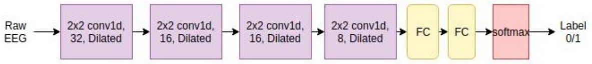


**Figure 6. Dilated CNN architecture**

### LSTM Model

Our raw EEG data is a time series data, hence training a LSTM model with EEG data should give better performance than the shallow CNN architectures. It can incorporate context information across time and improve performance. In CNN architectures, the prediction for current input does not depend on previous signal values. Long Short-Term Memory (LSTM) model consists of memory units which preserves error signal so that it is large enough to be back propagated through time. The prediction of current time depends on mentioned number of previous time steps. For our model, each data point consists of 4000 time steps and the final prediction depends on previous 4000 time steps. The memory units also learn how much memory is to be transferred to the next hidden state based on current input. Our model is a shallow LSTM network with one hidden layer.

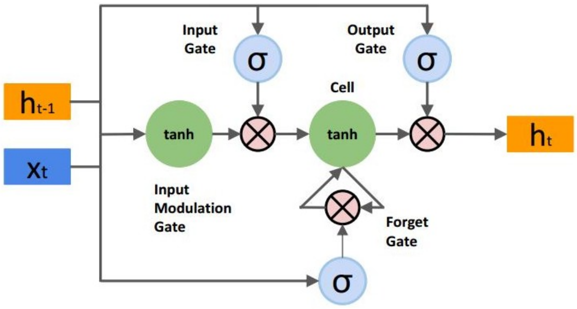


**Figure 7. LSTM architecture taken from [4]**

$$\begin{pmatrix} \tilde{f_t} \\ \tilde{i_t} \\ \tilde{o_t} \\ \tilde{g_t} \end{pmatrix} = W_h h_{t-1} + W_x x_t + b,$$

$$c_t = \sigma(\tilde{f_t}) \odot c_{t-1} + \sigma(\tilde{i_t}) \odot tanh(\tilde{g_t}),$$

$$h_t = \sigma(\tilde{o_t}) \odot tanh(c_t),$$

**Figure 8. LSTM architecture can be defined by these equations taken from [4]**

## RESULTS

### Baseline Models

From the plots, *figure 9*, we can see that, some classifiers perform well with some features and not so well with other features. Decision tree and random forest classifier generally classifies better when trained with standard deviation EEG, max EEG, min EEG and mean EEG features. Logistic regression classifier performs better when trained with mean EEG peaks, percentile 25 EEG peaks, percentile 50 EEG peaks and percentile 75 EEG peaks features. SVM classifier with rbf kernel performs best when trained with number of EEG peaks feature. SVM classifier with rbf kernel generally performs better than SVM with sigmoid kernel.

The accuracy values for our baseline models can be in *table 1*. The best features with which the models given good accuracy are :

- max EEG signal
- min EEG signal
- mean EEG signal
- number of EEG peaks
- percentile at 75 EEG peaks

### XGBoost

When compared to traditional machine learning models we can see that XGBoost performs better than all models. Also it gives good accuracy with any feature selected for training.

| XGBoost Classifier Accuracy | |
|---|---|
| Train Acc | **83%** |
| Test Acc | **81.9%** |

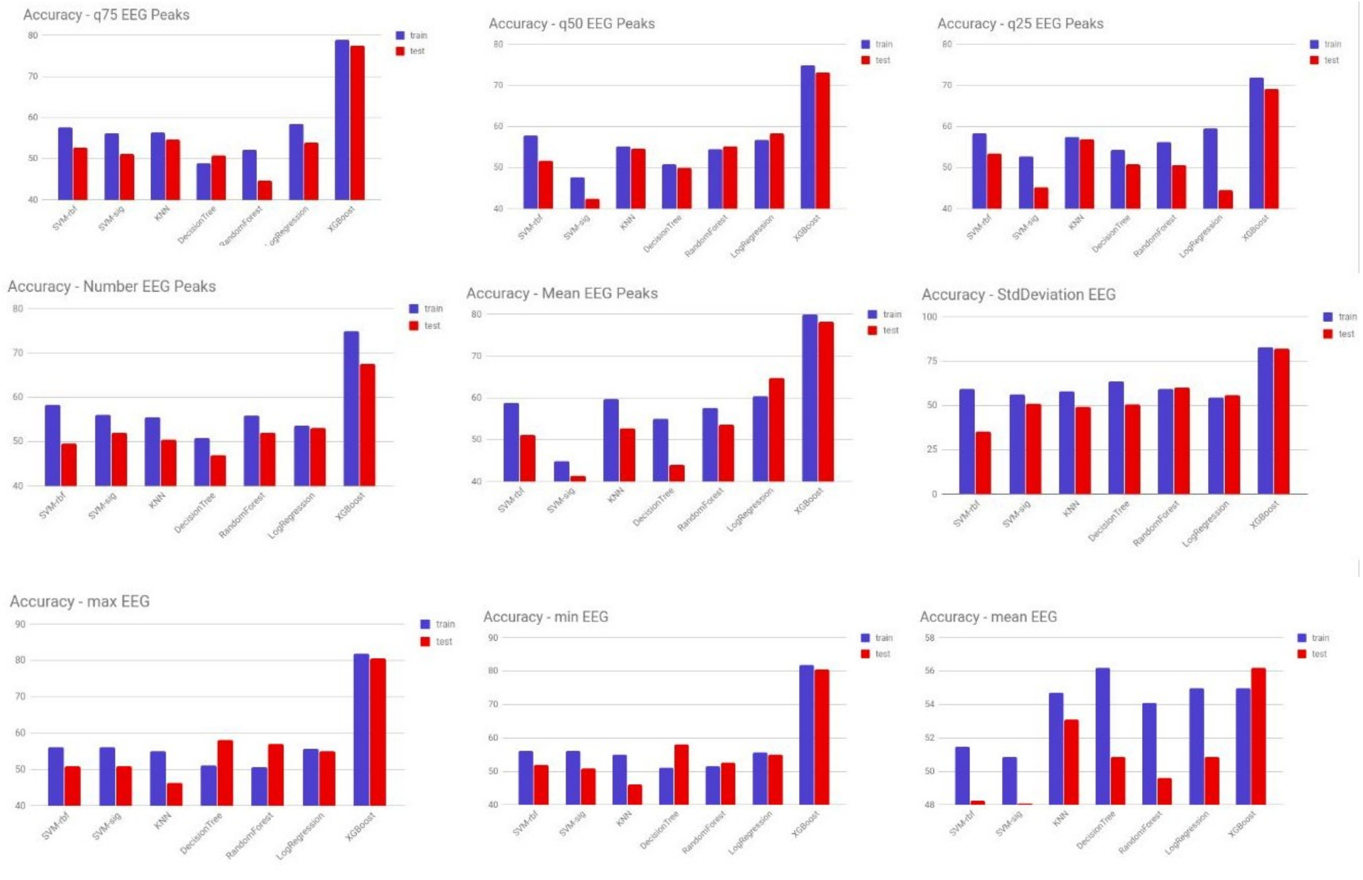


**Figure 9. Plots for train and val accuracies for various ML models**

| Model | Best Feature | Train Acc | Test Acc |
|---|---|---|---|
| SVM with sigmoid kernel | mean EEG peak | 58.4% | 53.5% |
| SVM with rbf kernel | percentile 25 EEG peak | 56.51% | 51.1% |
| K Nearest Neighbors Classifier | percentile 75 EEG peak | 59.31% | 54% |
| Decision Tree Classifier | std deviation EEG peak | 64% | 58% |
| Random Forest Classifier | std deviation EEG peak | 63% | 60% |
| Logistic Regression Classifier | percentile 75 EEG peak | 62% | **65%** |

**Table 1. Accuracy Values for traditional ML Models**

### CNN and Dilated CNN Model

The test accuracy obtained by neural network architectures, *figure 10*, is not as good as that given by XGBoost classifier. This could be because, large amount of data is required for training neural networks. Since our dataset size is small, these models get overfitted and give higher train accuracy but lower test accuracy.

| Accuracy for Neural Network Architectures | | |
|---|---|---|
| | CNN | Dilated CNN |
| Train Acc | **86%** | **88%** |
| Val Acc | **77%** | **68%** |
| Test Acc | **74%** | **63.2%** |

### LSTM Model

The LSTM model performs better than CNN models but not as good as XGBoost classifier. This again could be because of small dataset sizes.

| Accuracy for LSTM | |
|---|---|
| Train Acc | **75%** |
| Test Acc | **70%** |

### EXPERIMENT: RELATIVE DIFFICULTY OF SUBTITLED VIDEOS

We performed an experiment to find the relative difficulty of subtitled videos when compared with English language videos. For this experiment, our model was trained only on the data obtained on English language videos. This model was tested against the EEG data obtained while watching subtitled video.

Since we are using only english videos for training, the number of data points for training are halved. Hence deep learning methods will not give good results for our experiment. Also from the above experiments, we can conclude that XGBoost classifier gave best performance among all the models, so we

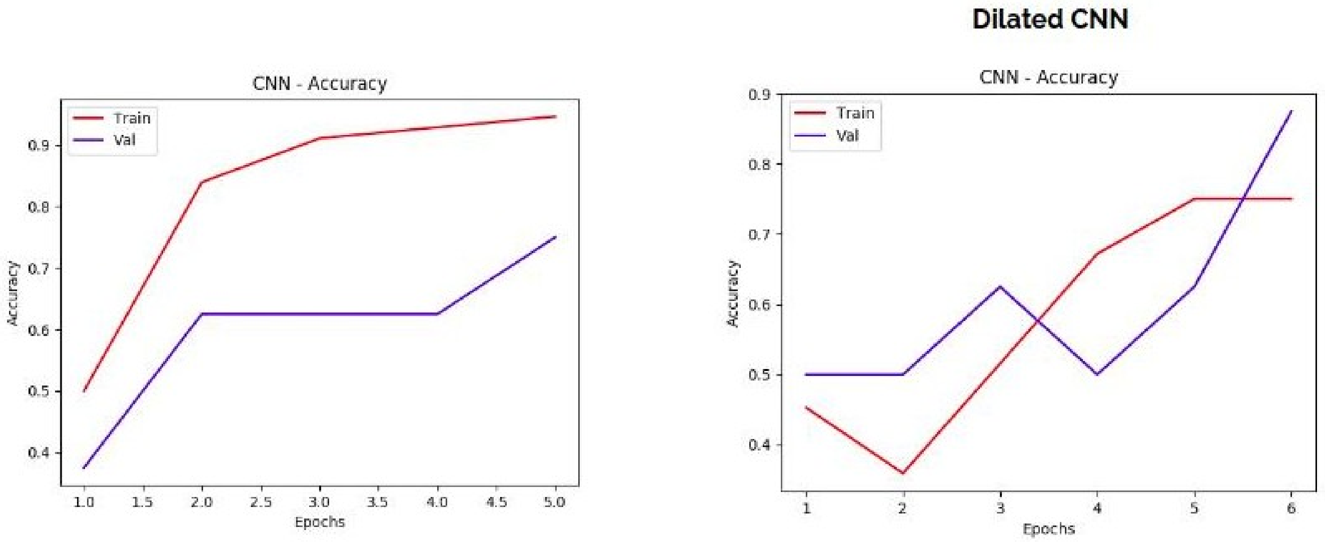
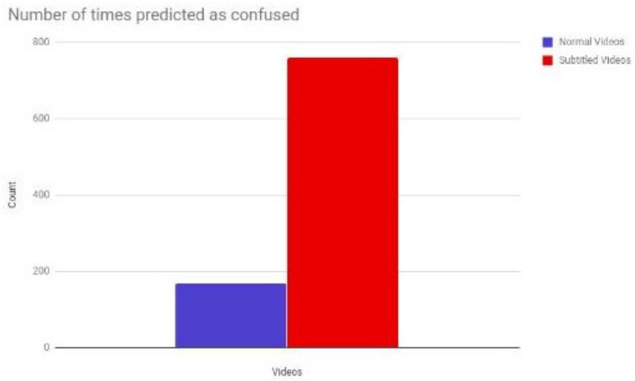


**Figure 10. Accuracy plots for neural network architecture**

performed our experiment only using it.



**. Figure 11. Num of times videos classified as confusing**

From the *figure 11*, we can see that, the number of times subtitled videos are getting classified as confusing (red) is more than that of normal english videos(blue).

## CONCLUSION

In this project, we collected EEG dataset for detecting confusion using a 14 channel EEG headset. Our experiments demonstrate that EEG signals can be used to train models to predict confusion in a student. This provides a good mechanism to provide feedback of video content available via web.Among our models, XGBoost classifier gave the highest performance with accuracy of 81.9%.

We also performed an experiment to test if the relative difficulty of subtitled videos is greater than normal english language video. From our results, we can conclude that EEG data of subtitled videos are generally classified as confusing.

## FUTURE WORK

We would like to do a larger user study with more number of participants. More data would allow us to get better performance on deep learning methods such as shallow ConvNets and Bi-LSTMs. Along with that, we would enlarge our handcrafted features by including power spectrum features from beta, alpha, theta and gamma waves.

We trained our baseline models individually with one feature at a time, further work can be done on analyzing the performance by training with combination of features, also combination of different baseline models can be tried to arrive at a more powerful model.

## ACKNOWLEDGMENTS

We thank Prof. Tauhidur Rahman for providing us with guidance and support for completing our project. We are also thankful to be allowed access to the Emotiv Epoc+ headset. We are also grateful to all the participants of our user study for helping us to collect our data.

## REFERENCES

1. https://www.class-central.com/report/mooc-stats-2016/.
2. https://www.class-central.com/report/moocs-stats-and-trends-2017/.
3. Kulkarni, C., Cambre, J., Kotturi, Y., Bernstein, M. S., and Klemmer, S. Talkabout: Making distance matter with small groups in massive classes. In *Design Thinking Research*. Springer, 2016, 67–92.
4. Ni, Z., Yuksel, A. C., Ni, X., Mandel, M. I., and Xie, L. Confused or not confused?: Disentangling brain activity from eeg data using bidirectional lstm recurrent neural networks. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, ACM (2017), 241–246.
5. Niedermeyer, E., and da Silva, F. L. *Electroencephalography: basic principles, clinical*

*applications, and related fields*. Lippincott Williams & Wilkins, 2005.

6. Piazza, C., Miyakoshi, M., Akalin-Acar, Z., Cantiani, C., Reni, G., Bianchi, A. M., and Makeig, S. An automated function for identifying eeg independent components representing bilateral source activity. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, Springer (2016), 105–109.

7. Subasi, A., and Gursoy, M. I. Eeg signal classification using pca, ica, lda and support vector machines. *Expert Systems with Applications 37*, 12 (2010), 8659–8666.

8. Uchidiuno, J., Hammer, J., Yarzebinski, E., Koedinger, K. R., and Ogan, A. Characterizing ell students' behavior during mooc videos using content type. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, ACM (2017), 185–188.

9. Uchidiuno, J., Ogan, A., Koedinger, K. R., Yarzebinski, E., and Hammer, J. Browser language preferences as a metric for identifying esl speakers in moocs. In *L@S* (2016).

10. Uchidiuno, J., Ogan, A., Yarzebinski, E., and Hammer, J. Understanding esl students' motivations to increase mooc accessibility. In *Proceedings of the third (2016) ACM conference on Learning@ Scale*, ACM (2016), 169–172.

11. Wang, H., Li, Y., Hu, X., Yang, Y., Meng, Z., and Chang, K.-m. Using eeg to improve massive open online courses feedback interaction. In *AIED Workshops* (2013).

12. Yeo, M. V., Li, X., Shen, K., and Wilder-Smith, E. P. Can svm be used for automatic eeg detection of drowsiness during car driving? *Safety Science 47*, 1 (2009), 115–124.