

Detecting the Confusion of Students in Massive Open Online Courses Using EEG

Xiuping Men, Xia Li*

School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, China

*Corresponding author: muzixia0212@126.com

Abstract: Confusion among students hinders learning and contributes to demotivation and disinterest in the course materials. However, it takes a lot of time and resources to identify confused pupils in extensive courses. Using LSTM and Attention, we suggest a deep learning model for monitoring students' confusion by EEG signals from students when they watching MOOC videos. The model obtained an accuracy of 0.82 on the EEG data, exceeding the previous experimental results for this dataset. Experiments show that the attention mechanism picks up on the significance of various features on prediction results. It can effectively solve the overfitting problem and improve the model classification effect.

Keywords: LSTM, BiLSTM, Attention, Confusion, EEG.

1. Introduction

Since the global outbreak of the Covid-19 pandemic, Nearly 400 million verified occurrences of New Crown pneumonia have been reported globally. Schools were closed all around the world as a result of the epidemic's continuous expansion. Over 1.2 billion children worldwide attend schools outside of the classroom in 186 nations. [1]. Because of the distinctive growth of e-learning, where instruction is delivered online and through digital platforms, education has undergone significant transformation. Online learning has been demonstrated to improve retention of information and reduces the time required, suggesting that the alterations brought on by the coronavirus may be long-lasting.

Online teaching ensured that universities were able to carry out their teaching activities during the epidemic, but they also have many shortcomings. Students who enroll in such courses frequently complain about the lack of interaction and feedback in video tutorials [2]. In-classroom education, the teacher or professor can have the chance to assess the students' level of misunderstanding by observing their body language such as facial expressions, body postures, scratching the head, etc. When it comes to on-line teaching, it would be more difficult to observe student's confusion [3].

AI systems that can forecast the needs, emotions, and abilities of their users and then tailor their interactions with them are becoming increasingly popular. This involves recognizing and responding to a user's emotional condition. Confusion is one such condition that is especially crucial to user experience when engaging with complicated interfaces since when people is confused, their pleasure and performance suffer [4]. When a system can recognize its user's bewilderment, it gets awareness that may be used to deliver relevant interventions to help them overcome their problems. Because data visualizations are already ubiquitous in our everyday lives, detecting and resolving misunderstanding is becoming increasingly important in helping people interacting with Information Visualizations.

This issue has attracted many scholars and has led to some important research results. The most basic approach of measuring confusion is through participants' own reports after learning [5]. Facial expressions are an obvious way to identify

other people's emotions. Because of the simple and uncomplicated processing, Facial expression extraction methods are widely used for confusion recognition [6]. Eye tracking, according to some experts, could be useful in determining the level of confusion encountered by students. The eye tracker records the positions and lengths of the eye's visual gaze on the screen, which are assumed to indicate attention allocation [7]. Facial electromyography (EMG) is another technology that goes beyond visual monitoring of the face to objectively determine facial expressions of confusion. With electrodes attached on the skin's surface, EMG is used to analyze the electrical activity of contracted muscles so as to identify different emotions, i.e., facial EMG can reflect a person's expressions even though they are hardly observable on people's face [8]. The electroencephalogram (EEG) is a technique for recording the activity of neurons in the brain. Because EEG reacts to the biological activity of brain tissues, it can be used to determine the brain's functioning status. EEG signals collected from various parts of the brain represent a wide range of information [9]. Furthermore, a study of students learning from video tutorials in MOOC found that utilizing EEG to identify learners' confusion was demonstrated the same as using education experts who assessed confusion through observing students' body language [10].

This study focuses on proposing the best deep learning algorithms to detect the confusion in students while watching MOOCs by recorded EEG data. We propose an attention model based on a hybrid of LSTM/BiLSTM and Attention which achieves 82% classification accuracy, surpassing previous classification results on this dataset.

2. Related Works

In recent years, experts have looked into several facets of students' perplexed feelings. Asking participants to describe their level of perplexity during or after learning exercises is the simplest technique of measuring [11]. Geller proposes a new method for identifying confusion in online education platform based on students' self-reported emotional states (as expressed through a set of pre-defined hashtags) [12]. When this supplementary work interferes with the core learning task and affects academic achievement, self-reported emotional

impacts are unsatisfactory. Another issue with self-report is its lack of sensitivity, and participants are sometimes unable to appropriately name their feelings.

Sims proposed an architecture that detects user confusion with eye-tracking data using RNN and CNN sub-models in tandem, leading to a 22% increase in comprehensive accuracies for confused and not confused class [13]. The EMG system correctly identified bewilderment in 21 subjects (87.5%), demonstrating its efficacy in detecting confusion [14]. Because of the basic and uncomplicated processing, which includes lower cost and improved accuracy, several approaches based on facial expression extraction have been used for confusion detection. For example, a recent study by Shi found that combining two classifiers, Convolutional Neural Network and Support Vector Machine (CNN-SVM), gave good results in recognizing perplexity for academic 82 students, with an accuracy of 93.8 percent [15].

The Electroencephalogram (EEG) system, or EEG for short, employs electrodes on the scalp to collect data on numerous aspects of human cognition, actions, and emotions, and is utilized in brain research, health, emotion, and mood monitoring.

Deep learning not only captures expert knowledge and incorporates it into the system, but it also performs feature extraction for detecting brain patterns that occur during seizures, as well as simulated artifacts, eye movements, and background noise, resulting in a sensitivity of over 90% and a false alarm rate of less than 5%, which meets the requirements for clinical applications [16].

Ni employed "Bidirectional LSTM Recurrent Neural Networks" to identify students' bewilderment through EEG data. The findings demonstrated that it is feasible to construct a model that can determine whether a person is confused or not and evaluate continuous data. In that investigation, 73.3 percent accuracy was attained [17]. Another project described by Zhou used an EEG device to identify bewilderment. The use of an EEG-based Brain Computer Interface was the first step towards observing and interfering with the learning process. They obtained and analyzed EEG data from 16 subjects. Raw EEG data from the Emotiv headset was displayed in less than 15 seconds. The accuracy rate was determined to be 71.36% [18].

3. Methodology

3.1. RNN

RNN is a neural network composed of self-feedback neurons with loops to preserve information. RNNs are cyclic networks where each element in a sequence is performed the same operation, and the output element depends on the previous element or state. RNN takes the output of the previous hidden layer to the next hidden layer and trains it together each time, and that is the main difference from traditional neural networks. RNNs are typically employed in deep learning tasks such as sentiment classification, image classification, image acquisition, machine translation, video splitting, etc. RNNs can learn the contextual information of samples. In this paper, we mainly use the variant LSTM and BiLSTM of RNN.

LSTM is a variant of RNN that solves the long-term dependency problem of RNN by controlling the circulation of features and anytime through three gates [19]. In short, LSTM is better at handling longer sequences than ordinary RNN. Figure 1 shows its structure.

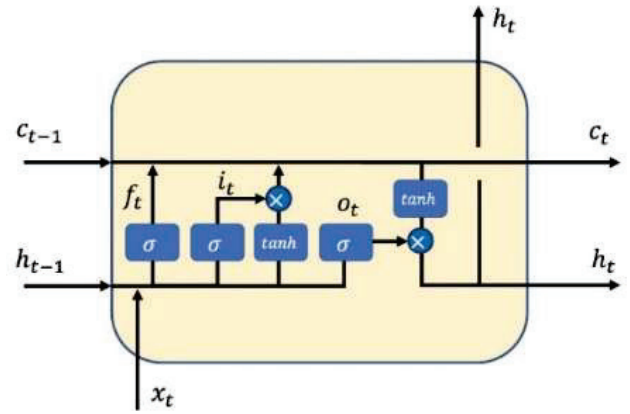


Figure 1. LSTM Unit Architecture

The first step is to choose at what level to discard irrelevant information from the cell state through the forgetting gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

The next step will be to decide what new information we will save in the unit state. It has two parallel layers, the tanh layer creates a cell state update value (\tilde{C}_t) and the "input gate layer" controls which features of this vector (\tilde{C}_t) are used to update C_t .

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

The following operation is the update of C_t :

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

Eventually, the output of LSTM is obtained:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

In order to solve the problem, two LSTM networks are designed for forward and directional LSTMs, which are called bi-directional LSTM [20]. The idea is to plug the same input sequence into the forward and successive LSTMs, and then connect the implicit layers of the two networks together, and jointly access to the output layer for prediction. The structure of the LSTM and BiLSTM models used in this paper is shown in the figure 2 and Figure 3 respectively.

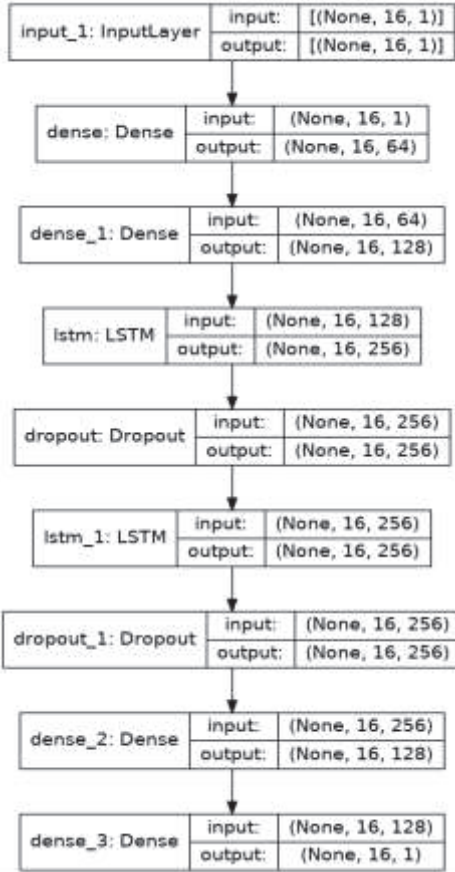


Figure 2. The Architecture of LSTM Model

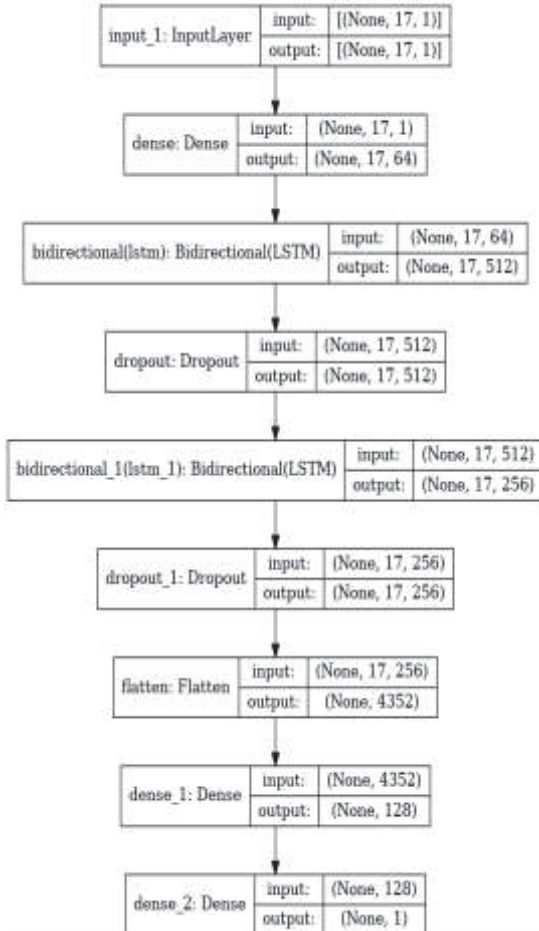


Figure 3. The Architecture of BiLSTM Model in This Paper

3.2. Attention

Attention is a strategy in neural networks that simulates cognitive attention [21]. The effect improves some parts of the input data while detracting from others, with the goal of the network focusing more on the little but crucial bits of the data.

Attentional mechanisms usually require the following computational steps:

(1) A query vector $q = s_{t-1}$ is first constructed, and then a scoring function calculates the correlation between the query vector and each corresponding key, resulting in a score. This value can usually be obtained by the dot product operation:

$$S_{q,k_i} = q \cdot k_i \quad (7)$$

(2) A SoftMax function is used to numerically transform the attention scores to highlight the weights of important elements:

$$\alpha_{q,k_i} = \text{softmax}(S_{q,k_i}) \quad (8)$$

(3) The generalized attention is then computed by a weighted sum of the value vectors, V_{k_i} :

$$\text{att}(q, K, V) = \sum_i \alpha_{q,k_i} V_{k_i} \quad (9)$$

We apply attention methods borrow from attention-based encoder-decoder architecture [22]. Attention weights are SoftMax of

$$V * (W * S_{t-1} + U * h) \quad (10)$$

Where, S_{t-1} is hidden state at $\langle t-1 \rangle$ of decoder and h is hidden state outputs of the encoder.

3.3. Models based on LSTM and Attention

Attention mechanisms have achieved remarkable results in many text and image classification tasks [23]. In this research, we combine attention with LSTM to propose a hybrid neural network model, as shown in Figure 4. The contextual information is obtained by the LSTM or BiLSTM module. To eliminate the redundancy of features, we add the attention mechanism after the LSTM/BiLSTM model.

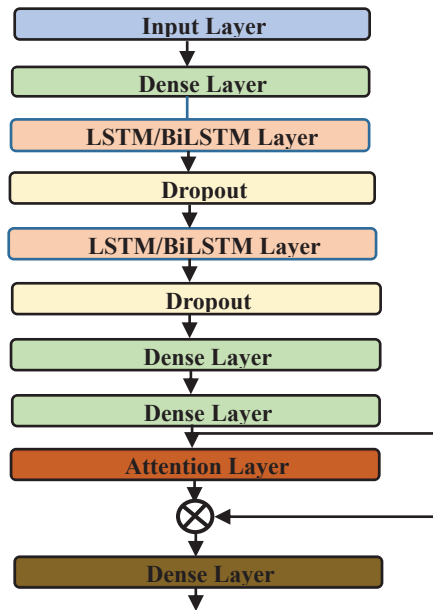


Figure 4. The Serial Model for LSTM/BiLSTM+Attention

4. Experiment Analysis

4.1. DataSet

The data used in this paper come from the EEG signals collected from 10 college students while they were watching a MOOC video clip [24]. By gathering data from the brain, the EEG signal can assist in detecting students' mental states. These ten pupils were given 20 movies to watch, 10 of which were deemed "easy" and 10 of which were designated "tough." To begin, various online education videos were chosen that would not be confusing to learners. Each video lasted roughly two minutes. In order to make the students more perplexed when watching the video, the middle part of the "difficult" video was randomly deleted for two minutes.

Each student's dataset was created by picking 5 films at random from each category. There are 12811 samples in the dataset. For each data point, there are 120 samples on average. There are 100 data points total, with 10 from each student.

Each student watched the video wearing a wireless mind-machine that measured frontal lobe activity. At the end of each video, students will be given a confusion rating on a scale of 1 to 7, with 1 being the least confused and 7 being the most confused. For simplicity, the 7 categories are converted into binary classification problems.

4.2. Data Processing

we need to remove several characteristics that might significantly skew the findings. Some values in meditation and attention are zero, which makes no sense. These values could be attributable to data collecting issues, according to the

dataset's author. i.e., the student's erroneous placement of the gadget on the head. There may be some features in the dataset that cause the results to be significantly skewed. SubjectID and VideoID are two of these functionalities. The SubjectID and VideoID fields contain information that is unrelated to the EEG brainwave. The "predefinedlabel" label indicates which confusion state was supposed to be detected by the experiment conductor prior to the test. We don't need this feature. Our target is the "user-definedlabel" since it is the label that indicates if a signal is correlated to a confusion state.

After cleaning the dataset, the EEG data were first merged with the demographic data of the students, then the data is normalized and standardized.

4.3. Experiment Configuration

In order to test the effectiveness of attention mechanism, LSTM and BiLSTM were adopted as basic models to compare the results of attention and no attention respectively.

Since it is a binary classification problem, the model is built using a binary cross-entropy loss function. The Adam optimizer is employed, and the model's accuracy is monitored over epochs. We performed experiments on a Intel Core i7 Processor of 2.8 GHz with 16GB RAM, along with a Tesla T4 GPU, with batch size of 32 and number of epochs of 100. Initial learning rate was set to $1e-5$. To optimize the training process, a dynamic adjustment strategy for the learning rate is needed in the experiment. We use an adaptive adjustment of the learning rate based on the callback function ReduceLROnPlateau. To speed up the training process, the Early Stopping callback function of the keras.callbacks module is used to monitor the change in the loss function for each epoch and stop training if there is no reduction for five consecutive epochs. To lessen volatility in the results caused by the random selection of folds, 10 runs of 10-fold cross validation are used to evaluate all models.

4.4. Result Analysis

At first, we use two basic models, LSTM and BiLSTM, to train on the EEG dataset and compare their classification results separately. We plotted the training process curves in Figure 5.

Both LSTM and BiLSTM stops training within 40 epochs, where LSTM does not converge. LSTM suffers from greater overfitting than BiLSTM, However, the BiLSTM model does not learn the features of the data well on the training set. The accuracy of the three models in the validation set (averaged over the last 10 epochs) 0.75 for LSTM, and 0.70 for BiLSTM, respectively.

Neither model met our expectations. This shows that it is difficult to obtain better training results with a single model. Naturally, we combine different models to make up for the shortcomings of a single model.

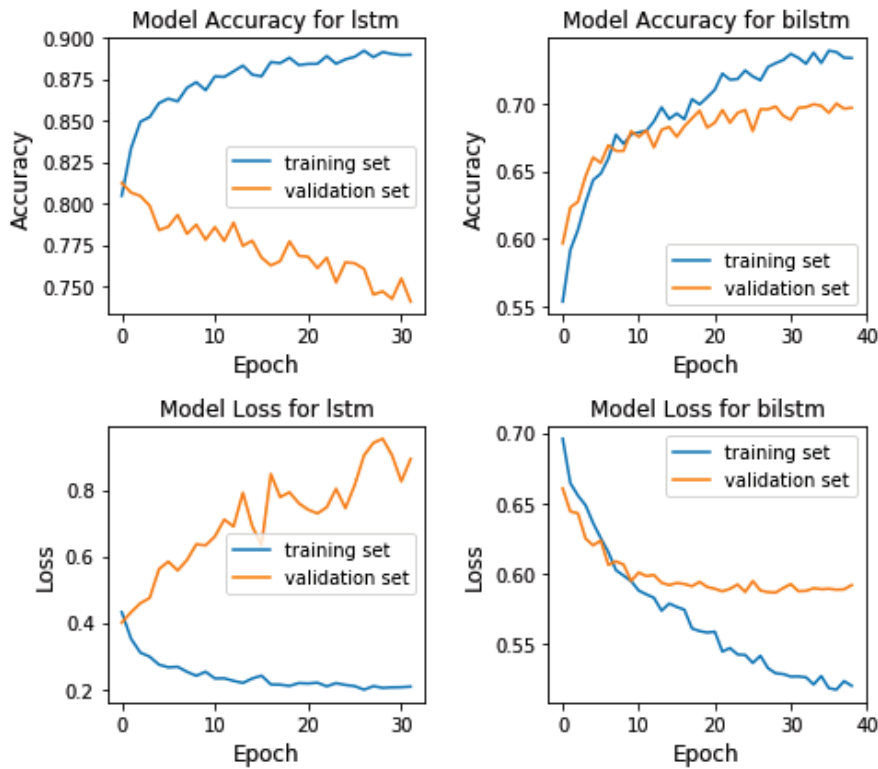


Figure 5. Test Set Performance of LSTM and BiLSTM

Then, the attention mechanism was embedded into the basic model respectively, and the experimental results are

shown in the Figure 6.

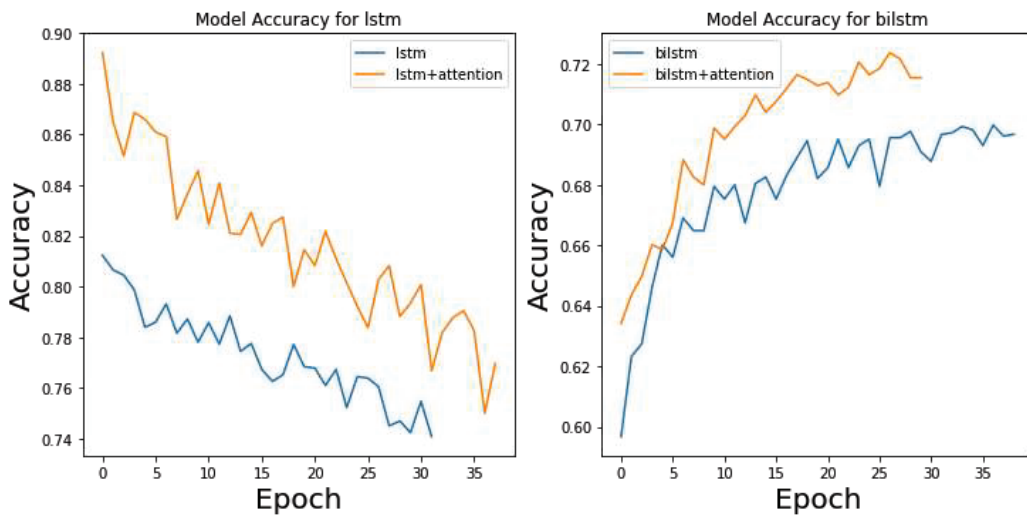


Figure 6. Test Set Performance of LSTM/BiLSTM Models with Attention

From the experimental result curves, it can be found that the accuracy of the models on the test set is significantly improved when the attention mechanism is added for both LSTM models. And the overfitting problem has been greatly alleviated.

The accuracy of LSTM with Attention improved by 3 percentage points over the LSTM model to 0.78. The accuracy of the BiLSTM with Attention model improved by 2 percentage points over the BiLSTM model, reaching 0.72. It shows that the Attention mechanism can effectively weight the irrelevant features learned by the basic model, which effectively reduces the impact of these features on the model performance.

The average accuracy of each model for all epochs is presented in Table 1. It can be found that the accuracy of the

LSTM and BiLSTM models differ greatly. The gap between the performance of LSTM and BiLSTM models also appears in their results when combined with Attention. This indicates that the improvement of the performance of the model by the Attention mechanism receives the limitation of the performance of the model itself.

Table 1. Test set performance of all models

Model	Accuracy
LSTM	0.773
BiLSTM	0.679
LSTM+Attention	0.817
BiLSTM+Attention	0.697

5. Conclusion and Discussion

In this paper, we propose an Attention model based on BiLSTM and LSTM for detecting students' confusion when watching MOOC learning videos. It is found that LSTM and BiLSTM are able to learn contextual features. However, LSTM/BiLSTM model also learns some irrelevant features, which leads to a serious overfitting of the model, and the perfect effect of the model on the training set fails on the validation set. To solve this problem, we add the attention module to the model, through which the weight of the features learned by the model is adjusted to highlight the features that are relevant to the classification results and weaken the features that are not relevant to the classification results in this row.

The attention used in this paper is relatively simple, and we will further employ other attentions in the future to enhance the model classification results. In addition, the dataset used in this paper is only from 10 experimental subjects, and the diversity of data has some limitations, and we will adopt a larger dataset in the future to verify the effectiveness of our method.

Acknowledgment

This research work was supported by School-level general teaching research topics of Anhui University of Finance and Economics (Project acjy2020030).

References

- [1] Saeed, B., Ullah, A., & Khan, M. A. (2021). Attitude of university students on online teaching under corona virus pandemic situation in Pakistan. *Liberal Arts and Social Sciences International Journal*, 5(1), pp.28-40.
- [2] Ahn, J. W., Ku, Y., & Kim, H. C. (2019). A novel wearable EEG and ECG recording system for stress assessment. *Sensors*, 19(9), 1991.
- [3] Anala, V. A. S. M., & Bhumireddy, G. (2022). Comparison of Machine Learning algorithms on detecting the confusion of students while watching MOOCs.
- [4] Sucheta Nadkarni and Reetika Gupta. 2007. A Task-based Model of Perceived Website Complexity. *Mis Quarterly*, pp.501-524.
- [5] Lehman, B., D'Mello, S., & Graesser, A. (2012). Confusion and complex learning during interactions with computer learning environments. *Internet and Higher Education*, 15(3), pp.184-194. doi:10.1016/J.Iheduc.2012.01.002.
- [6] Grafsgaard, Joseph F., Kristy Elizabeth Boyer, and James C. Lester. "Predicting facial indicators of confusion with hidden Markov models." *International Conference on Affective computing and intelligent interaction*. Springer, Berlin, Heidelberg, 2011.
- [7] Sims, S. D., & Conati, C. (2020, October). A neural architecture for detecting user confusion in eye-tracking data. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 15-23).
- [8] Durso, F. T., Geldbach, K. M., & Corballis, P. (2012). Detecting confusion using facial electromyography. *Human factors*, 54(1), 60-69.
- [9] Jun, G., & Smitha, K. G. (2016). EEG based stress level identification. In *2016 IEEE international conference on systems, man, and cybernetics (SMC)* (pp. 003270-003274). IEEE.
- [10] Wang, H., Li, Y., Hu, X., Yang, Y., Meng, Z., & Chang, K. M. (2013). Using EEG to Improve Massive Open Online Courses Feedback Interaction. In *AIED Workshops*.
- [11] D'Mello, S., & Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, 151, 106-116. doi:10.1016/J.Actpsy.2014.06.005
- [12] Geller, Shay A., et al. "Confused and beyond: detecting confusion in course forums using students' hashtags." *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 2020.
- [13] Sims, S. D., & Conati, C. (2020). A neural architecture for detecting user confusion in eye-tracking data. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 15-23).
- [14] Durso, Francis T., Kaitlin M. Geldbach, and Paul Corballis. (2012). Detecting confusion using facial electromyography. *Human factors*, 54.1: 60-69.
- [15] Shi, Zheng, et al. "Automatic Academic Confusion Recognition In Online Learning Based On Facial Expressions." *2019 14th International Conference on Computer Science & Education (ICCSE)*. IEEE, 2019.
- [16] Isaev, D. Y., Tchapyjnikov, D., Cotten, C. M., Tanaka, D., Martinez, N., Bertran, M., & Carlson, D. (2020). Attention-based network for weak labels in neonatal seizure detection. *Proceedings of machine learning research*, 126, 479.
- [17] Ni, Zhaoheng, et al. "Confused or not confused? disentangling brain activity from eeg data using bidirectional lstm recurrent neural networks." *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2017.
- [18] Zhou, Yun, et al. "Confusion State Induction and EEG-based Detection in Learning." *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018.
- [19] Graves, A. (2012). Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pp.37-45.
- [20] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks*, 18(5-6), 602-610.
- [21] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- [22] Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural Machine Translation By Jointly Learning To Align and Translate. In *ICLR*, pp. 1-15.
- [23] Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November). Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 606-615.
- [24] Wang, H., Li, Y., Hu, X., Yang, Y., Meng, Z., & Chang, K. M. (2013, June). Using EEG to Improve Massive Open Online Courses Feedback Interaction. In *AIED Workshops*.