# Detecting Distracted Students in an Educational VR Environment Utilizing Machine Learning on EEG and Eye-Gaze Data

Sarker M. Asish *  Arun K Kulshreshth †  Christoph W Borst ‡

University of Louisiana at Lafayette, Lafayette, Louisiana, United States

## ABSTRACT

Virtual Reality (VR) is frequently used in various educational contexts since it could improve knowledge retention compared to traditional learning methods. However, distraction is an unavoidable problem in the educational VR environment due to stress, mind-wandering, unwanted noise/sounds, irrelevant stimuli, etc. We explored the combination of EEG and eye gaze data to detect student distractions in an educational VR environment. We designed an educational VR environment and trained three machine learning models (CNN-LSTM, Random Forest and SVM) to detect distracted students. Our preliminary study results show that Random Forest and CNN-LSTM provide better accuracy (98%) compared to SVM.

**Keywords:** Machine learning; Virtual Reality; Distraction; EEG; Eye-tracking; Education.

**Index Terms:** Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality; Computing methodologies—Machine Learning;

## 1 INTRODUCTION

Using VR for educational applications has many potential benefits including increased engagement and motivation of students, better communication of size and spatial relationships of modeled objects, and stronger memories of the experience. In a real classroom, teachers have a sense of the student's engagement and actions from cues such as body movements, eye gaze, and facial expressions. This awareness is significantly reduced in a VR environment because a teacher can't see students directly. Additionally, students could get distracted in VR due to reasons involving stress, mind-wandering, unwanted noise, external alerts, etc. Thus, to help teachers manage their VR-based classes, an automated distraction detection is required.

Distractions could be classified as either external (due to the environment) or internal (due to internal thoughts). Eye tracking data has been used in the past [2] to detect external distractions such as noise or sounds played in the VR environment. Internal distraction could occur when students have internal stress, mind-wandering, or other thoughts in mind. However, eye-gaze data can not measure internal distractions. It is quite possible that the user may be looking at the educational content and may be thinking about something else. EEG data has been used in the past [1] to detect these internal distractions. In this work, we explore the combination of EEG and gaze data to detect both kinds of distractions (internal and external) using machine learning approaches. We designed an educational VR environment (see Figure 1) and collected EEG and eye gaze data from 10 participants. The data set was then used to train three machine learning models (RF, SVM, and CNN-LSTM) to classify if there were internal/external distractions or not.

This is an author formatted version. Original publication: S. M. Asish, A. K. Kulshreshth and C. W. Borst, "Detecting Distracted Students in an Educational VR Environment Utilizing Machine Learning on EEG and Eye-Gaze Data," 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Shanghai, China, 2023, pp. 703-704, doi: 10.1109/VRW58643.2023.00194.



Figure 1: Virtual solar field environment explains how solar power is generated. An avatar explains different components using audio, animations, and text slides.

## 2 METHOD

We chose a Virtual Energy Center (see Figure 1) as our educational VR environment. Our experiment (duration of 20-25 minutes) had 10 participants (7 males and 3 females, age range: 19 to 30). The participants practiced the tasks using a different educational VR environment (an educational VR presentation on biological cell structure). Our experiment had two phases: 1) Phase-I with internal distractions tasks, and 2) Phase-II with external distractors. In both phases, we collected EEG (using a OpenBCI electrode cap) and eye-tracking data from four different educational sessions (in random order with session times ranging from 1.5 to 4 minutes) related to solar field displayed on the HTC Vive Pro Eye. For each participant, we calibrated the eye-tracker and made sure that the EEG cap was working correctly. The two phases are described below:

**Phase-I:** An internal distraction (see Table 1) appears randomly once during the educational sessions after 50 to 60 seconds. A doorbell sound (1 sec duration) was played following the task description indicating the participants to start the internal distraction task. At the end of each session, we asked participants about what they thought or counted for Q1 and Q2 (Table 1) and if they were distracted. If they mentioned distraction then we labeled the data as internally distracted for a 20 second window, 5 sec after the doorbell sound.

**Phase-II:** We created five external distractions for the participants. Our four external sound/noise distractions were: (1) a pre-recorded mobile ringtone, (2) an pre-recorded external audio conversation between two people, (3) a dialogue unrelated to the educational content played randomly, and (4) door closing and opening sounds. The duration was 10 seconds for (1) and (2), and 3 seconds for (3) and (4). Our fifth distraction was a graphical glitch where we simulated the camera glitch effect to distort the screen for 10 seconds. For each session in Phase II, these distractions appear every 40 to 50 seconds randomly.

For segmenting the distracted data, we needed to pick a time window (starting from the time when the distraction event started). Based on results from a prior work [1] which compared 20 and 30 second window, a window length of 20 seconds was selected since it provided a better accuracy and we believe that distraction does

Table 1: Internal distraction tasks appeared randomly in the VR scene followed by a doorbell sound indicating them to start the task.

| Internal task to create distractions | |
|---|---|
| Q1 | Count how many times the avatar moves his hand. |
| Q2 | Count the total number of words starting with ”S”. |
| Q3 | Think about the activities for the upcoming weekend. |
| Q4 | Think about the activities during the last weekend. |

not last long. All the data points in this window were labeled as distracted.

**Sensor Data Collected:** Raw eye gaze data recorded includes timestamps, eye diameter, eye openness, eye wideness, gaze position, gaze direction, and gaze origin value (one 3d vector for each eye). Similarly, raw EEG data was collected from the frontal, central, occipital and parietal regions throughout the sessions using the OpenBCI electrode cap kit. It had 16 channels and after discarding invalid/unstable channels, we ended up with only 8 channels for all the participants. The electrode locations of these 8 channels were FP1, FP2, C3, C4, P3, P4, O1 and O2 placed according to the international 10-20 system. The notch filter was applied at the power frequency of 60 Hz, the bandpass filter was 1 to 50 Hz and smoothing was turned on from the OpenBCI tool settings while recording the EEG data. To clean the EEG data, we used the independent component analysis (ICA) to remove artifacts by MNE-Python library, though it is impossible to remove all the artifacts [3]. The eye gaze and EEG data sampling rate was 120 Hz and 125 Hz. We used early fusion approach. Thus, to be temporally compatible, we down-sampled the EEG data to 120 Hz to match with the eye-tracker data sampling rate.

Based on previous research on multimodal data classification [3, 4], we chose Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network with Long Short Term Memory (CNN-LSTM) ( [2]) to classify distraction using the EEG and gaze data. We used these models for our person-independent classification task in which we split the dataset into two parts: 70% training (134,400 data samples)and 30% testing (57,600 data samples). We used default parameters to implement these ML classifiers using Python (version 3.8.8) with the sklearn library.

We considered two cases: (1) three classes (internally-distracted, externally-distracted, and not-distracted), and (2) two classes (distracted and not distracted). For the first case, we found that the number of data points associated with each class was vastly different. The data was biased more towards ND class. To avoid the bias, we took equal number of slices for the three classes to train the models. For second case, we consider two classes (D and ND) and we combined the data from ID and ED labels into D class. To avoid bias and overfitting with the training data, we applied a combination of under-sampling and Synthetic Minority Over-sampling (SMOTE) so that classifiers can learn from the dataset perturbed by ”SMOTING” the minority class and under-sampling the majority class.

## 3 Preliminary Study Results

The accuracy for the three models is shown in Figure 2. The highest accuracy was 98% provided by RF and CNN-LSTM, and the lowest accuracy was 76% for SVM. Since accuracy does not provide class-wise results to understand individual class level performance, we also evaluated precision, recall, and F1-scores shown in Tables 2 and 3. We found that the best precision and recall scores achieved by RF and CNN-LSTM are very similar for both two class and three class classification tasks. However, SVM had the lowest precision, recall and F1 scores for both two class and three class classification results. CNN-LSTM takes account of both temporal and spatial features and performs well. However, RF and SVM only take account of spatial features. RF still performs better due to smaller dataset.

## 4 Conclusion and Future Work

We collected EEG and eye-tracking data during an educational presentation with multiple distraction events. We tested three different ML models (RF, SVM, and CNN-LSTM) to classify the data into two classes (Internally-distracted and not-distracted) and three classes (Internally-distracted, Externally-distracted and not-distracted). Our results show that RF and CNN-LSTM provides the best results compared to SVM for both two class and three class classification. In the future, we would like to consider intermediate and late feature fusion of multimodal data for detecting internal and external distractions. We also need to test our models by extracting the common EEG features such as alpha, theta, delta, beta and gamma wave.

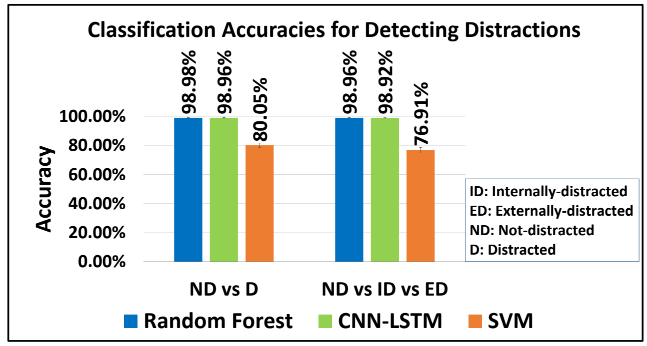


Figure 2: Mean Accuracy of the models

Table 2: Precision, recall and F1-score of the models for three classes

| Name | Class | precision % | recall % | F1-score % |
|---|---|---|---|---|
| SVM | ND | 0.62 | 0.94 | 0.75 |
| | ID | 0.90 | 0.50 | 0.64 |
| | ED | 0.94 | 0.89 | 0.92 |
| RF | ND | 0.97 | 0.99 | 0.98 |
| | ID | 0.99 | 1.0 | 0.99 |
| | ED | 1.0 | 0.99 | 0.99 |
| CNN-LSTM | ND | 0.96 | 0.98 | 0.98 |
| | ID | 0.99 | 0.98 | 0.98 |
| | ED | 1.0 | 0.98 | 0.99 |

Table 3: Precision, recall and F1-score of the models for two classes

| Name | Class | precision % | recall % | F1-score % |
|---|---|---|---|---|
| SVM | ND | 0.72 | 0.69 | 0.71 |
| | D | 0.84 | 0.86 | 0.85 |
| RF | ND | 1.0 | 0.99 | 0.99 |
| | D | 1.0 | 1.0 | 1.0 |
| CNN-LSTM | ND | 1.0 | 0.99 | 0.99 |
| | D | 1.0 | 1.0 | 1.0 |

## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1815976 and by the Louisiana Board of Regents under contract No. LEQSF(2022-25)-RD-A-24.

## References

[1] S. M. Asish, A. Kulshreshth, and C. Borst. Detecting internal distraction in an educational vr environment using eeg data. In *Symposium on Spatial User Interaction*, pages 1–2, 2022.

[2] S. M. Asish, A. K. Kulshreshth, and C. W. Borst. Detecting distracted students in educational vr environments using machine learning on eye gaze data. *Computers & Graphics*, 109:75–87, 2022.

[3] D. Jeong, S. Yoo, and J. Yun. Cybersickness analysis with eeg using deep learning algorithms. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)*, pages 827–835. IEEE, 2019.

[4] L.-M. Vortmann, S. Ceh, and F. Putze. Multimodal eeg and eye tracking feature fusion approaches for attention classification in hybrid bcis. *Recognizing the State of Emotion, Cognition and Action from Physiological and Behavioural Signals*, 2022.