Supervised vs Unsupervised Learning on Gaze Data to Classify Student Distraction Level in an Educational VR Environment

Sarker M. Asish
University of Louisiana at Lafayette
Lafayette, LA, United States
asish.sust@gmail.com

Arun K. Kulshreshth
University of Louisiana at Lafayette
Lafayette, LA, United States
arunkul@louisiana.edu

Christoph W. Borst
University of Louisiana at Lafayette
Lafayette, LA, United States
cwborst@gmail.com

ABSTRACT

Educational VR may help students by being more engaging or improving retention compared to traditional learning methods. However, a student can get distracted in a VR environment due to stress, mind-wandering, unwanted noise, external alerts, etc. Student eye gaze can be useful for detecting these distraction. We explore deep-learning-based approaches to detect distractions from gaze data. We designed an educational VR environment and trained three deep learning models (CNN, LSTM, and CNN-LSTM) to gauge a student's distraction level from gaze data, using both supervised and unsupervised learning methods. Our results show that supervised learning provided better test accuracy compared to unsupervised learning methods.

CCS CONCEPTS

• **Computing methodologies** → *Deep learning*; **Virtual reality**; • **Applied computing** → *Education.*

KEYWORDS

Deep Learning, Virtual Reality, Distraction, Education

ACM Reference Format:
Sarker M. Asish, Arun K. Kulshreshth, and Christoph W. Borst. 2021. Supervised vs Unsupervised Learning on Gaze Data to Classify Student Distraction Level in an Educational VR Environment. In *Symposium on Spatial User Interaction (SUI '21), November 9–10, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3485279.3488283

1 INTRODUCTION

Potential benefits of VR for education include 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 audience'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 reasons involving stress, mind-wandering, unwanted noise, external alerts, etc.

Gaze visualizations have been explored in the past to detect distracted students [3]. However, this approach is not feasible for a large class due to increased cognitive load of the teacher. We need automated distraction detection for an educational VR interface such as in [2]. In a previous work [1], we designed an education VR environment (see Figure 1) and collected eye gaze data of students from it. The data set was then used to train three supervised learning models (CNN, LSTM, and CNN-LSTM) to classify distraction level of a student on a per-session basis. However, it is still unclear if the supervised data labeling is the best approach for the gaze data. In this work, we are exploring several supervised and unsupervised learning methods to find the best data labeling approach for classifying distracted students based on the gaze data.

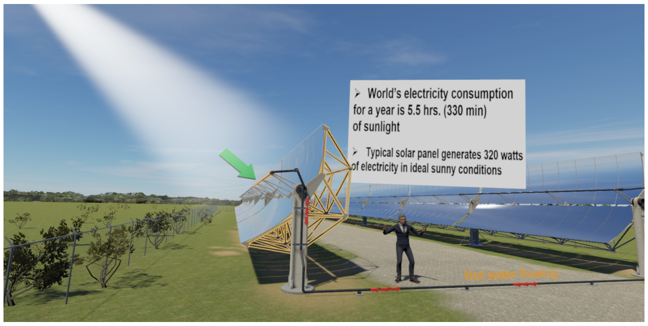
➢ World's electricity consumption for a year is 5.5 hrs. (330 min) of sunlight
➢ Typical solar panel generates 320 watts of electricity in ideal sunny conditions

Figure 1: Educational VR environment to explain how a solar field generates power. An avatar explains different components using audio, animations and text slides.

2 METHOD

Our VR environment was a Virtual Energy Center (see Figure 1) used for virtual field trips. Our experiment (duration 45-60 minutes) with 21 participants (16 male and 5 female, age range : 19 to 35) collected gaze data in two phases (appeared in random order): phase I with no external distractions and phase 2 with external distractions such as social media notifications, mobile ringtones, and external conversations/sounds. For each session containing distractions, these distractions appeared every 45 seconds. Each phase was divided into small sessions with a quiz in the end. The purpose of the quiz questions was to help gauge if the participant was distracted, under the assumption of some correlation between correct quiz answers and attention. The performance in the quiz was then used during data labeling in the supervised training models (see [1] for more details). Raw gaze data collected (sampling rate of 120Hz) throughout the sessions included timestamps, eye diameter, eye openness, eye wideness, gaze position, gaze direction, and a distance value (calculated as the distance between the Vive Eye's reported gaze origin and the highlighted object's position).

SUI '21, November 9–10, 2021, Virtual Event, USA

We split the dataset into training (70%) and test (30%) sets. The training set was used to train the classifiers and the test set was used to test a classifier's accuracy. Using the same training data, we trained three machine learning models (CNN, LSTM and CNN-LSTM) for both supervised and unsupervised data labeling methods. The unsupervised methods used K-means clustering for data labeling. The elbow method on our data shows a kink at k=2 and k=3, which indicates that we should consider two or three clusters. We chose three cluster to compare our results with our supervised models with three classes corresponding to low, medium and high distraction levels.

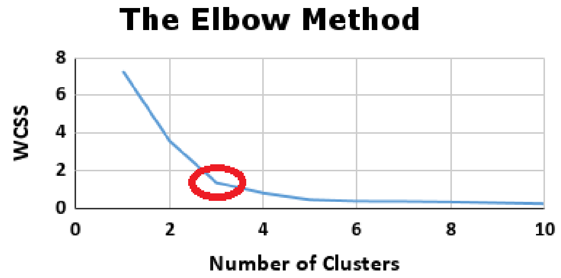

Figure 2: The relationship between the number of clusters and Within Cluster Sum of Squares (WCSS)(elbow method)

3 PRELIMINARY STUDY RESULTS

The accuracy for the three models for both unsupervised and supervised learning is shown in Figure 3. The overall accuracy for all models are very close to each other for both unsupervised and supervised learning models. However, the accuracy is significantly lower for the unsupervised models. The precision, recall and F1-scores are shown in Table 1 for the supervised models and in Table 2 for the unsupervised models. We found that accuracy was better for the low distraction class with the unsupervised learning. However, it had significantly lower accuracy for medium and high distraction classes compared to the supervised learning models.

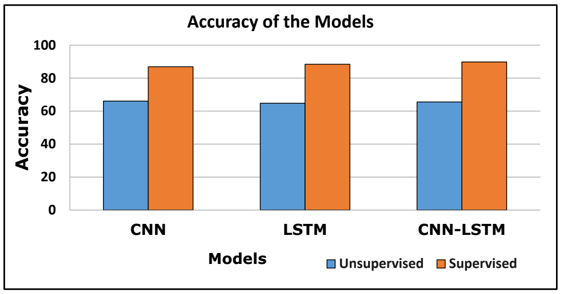

Figure 3: Average accuracy of the models

4 CONCLUSION AND FUTURE WORK

We compared supervised and unsupervised learning models on eye gaze data to classify student distraction level. Our results show that supervised learning is significantly more accurate then unsupervised learning for the three models (CNN, LSTM, and CNN-LSTM) we tested. Thus, we do not recommend these unsupervised learning models based on K-means clustering for student distraction level detection. Furthermore, the clusters we got after K-means may not necessarily represent distraction level and may correspond to some other aspect of the student experience.

Distraction level cannot be measured merely from eye gaze, as there are other factors involved (like physical and mental well being) that could affect distraction level. In the future, we would like to consider more metrics and sensor data (EEG, heart rate, skin conductance, etc.) for detecting distraction. Additionally, it is important to develop real-time detection methods to work in a wider range of VR environments.

Table 1: Precision, recall and F1-score of the models with supervised learning

Name	Class	precision %	recall %	F1-score %
CNN	low	0.88	0.85	0.86
	mid	0.87	0.88	0.87
	high	0.85	0.89	0.87
LSTM	low	0.91	0.85	0.88
	mid	0.88	0.90	0.89
	high	0.85	0.91	0.88
CNN-LSTM	low	0.90	0.89	0.90
	mid	0.91	0.89	0.90
	high	0.88	0.91	0.90

Table 2: Precision, recall and F1-score of the models with unsupervised learning

Name	Class	precision %	recall %	F1-score %
CNN	low	0.92	1.00	0.96
	mid	0.38	0.51	0.43
	high	0.63	0.45	0.52
LSTM	low	0.91	1.00	0.95
	mid	0.38	0.49	0.43
	high	0.59	0.43	0.50
CNN-LSTM	low	0.91	1.00	0.95
	mid	0.36	0.50	0.42
	high	0.64	0.44	0.52

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(2019-20)-ENH-DE-22.

REFERENCES

[1] Sarker Monojit Asish, Ekram Hossain, Arun K. Kulshreshth, and Christoph W. Borst. 2021. Deep Learning on Eye Gaze Data to Classify Student Distraction Level in an Educational VR Environment. In *ICAT-EGVE 2021 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, Jason Orlosky, Dirk Reiners, and Benjamin Weyers (Eds.). The Eurographics Association. https://doi.org/10.2312/egve.20211326

[2] David M Broussard, Yitoshee Rahman, Arun K Kulshreshth, and Christoph W Borst. 2021. An Interface for Enhanced Teacher Awareness of Student Actions and Attention in a VR Classroom. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 284–290. https://doi.org/10.1109/VRW52623.2021.00058

[3] Yitoshee Rahman, Sarker M Asish, Nicholas P Fisher, Ethan C Bruce, Arun K Kulshreshth, and Christoph W Borst. 2020. Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 868–877.