2023 International Conference on Smart Applications, Communications and Networking (SmartNets) | 979-8-3503-0252-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/SmartNets58706.2023.10216042

Intellitell: A Web-based Storytelling Platform for Emotion Recognition with Machine Learning

Aaron Jay V. Argel[1], Madeleine M. Ragmac[1], Markus Elija Thomas E. Lindo[1], Nestor Michael C. Tiglao[1,2]
[1] Electrical and Electronics Engineering Institute, University of the Philippines Diliman, Quezon City, Philippines 1101
[2] Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742, USA
aaron.jay.argel@eee.upd.edu.ph, madeleine.ragmac@eee.upd.edu.ph
elija.lindo@eee.upd.edu.ph, ntiglao@umd.edu

***Abstract*—Remote learning is relatively new for the education sector and is proving to be a challenge. Preschoolers are in a stage where they are highly active, and making sure that they are immersed in their educational activities is important. However, surveys and studies have shown that many students are not engaged in module-based distance learning methods. Higher engagement has been associated with higher emotional investment and higher chances of receiving positive learning outcomes. The goal of this paper was to develop a web-based storytelling platform, Intellitell, that gathered data on the student's engagement. The paper assessed a preschooler's engagement by gathering their emotional response using faceAPI.js during a storytelling activity using emotion match accuracy as a defining metric. The results showed that the emotion match accuracy of the students using the Intellitell webapp was 80.16% for defined data points. Further analysis showed that the ground truth is susceptible to distractions due to the low sample size, and classifying emotions that are similar like surprised and fearful is inconsistent without prior context to stimuli.**

***Index Terms*—emotion recognition, machine learning, faceAPI.js, FER-2013.**

I. Introduction

Because of the COVID-19 threat, schools opted to do online classes for their students to continue their education. However, in a survey regarding online classes by Pulse Asia [1], of the 63% of respondents who were parents with children in basic education, only 46% said that their child was learning and 25% said that their child was not learning.

In light of this, the International Society for Technology in Education (ISTE) Standards called for educators to design learner-centered materials that used technology to personalize learning experiences [2]. This showed that there is space for new approaches in the online setting.

The online setting particularly affected young children, who were still only learning emotional expression and behavior at their age. Proper emotional expression is part of proper development [3] [4]. By expressing appropriate reactions when faced with a situation, students show that their attention is focused on what they are seeing and are therefore more likely to be learning. Trowler and Trowler's [5] report on student engagement found that studies support the idea that student engagement led to positive educational outcomes.

The main focus of this paper was to develop a web app with basic emotion classification and engagement computation functionalities, with features that can be accessed depending on whether the user is a teacher or a student. The study did not gather bio-signals or other data from sensors.

The storage of the students' video after they watch their assigned storytelling video was not implemented for several reasons:

- Sending the video requires keeping a certain modal open, and participants might exit before the process is finished.
- Keeping video recordings of the student is intended as a fallback in case problems with face capturing happens. It is a feature that is not a main focus of the study.
- Uploading the captured video to the cloud storage may vary in time depending on internet speeds and may even fail to upload. Analyzing the live webcam feed and only sending the data is much faster and consistent.

Furthermore, this paper's motivation was to pivot from a researcher-focused platform into a student-focused web application where its primary goal is to provide a platform for students to learn in an alternative environment in a non-traditional teaching method.

In consideration of these factors, this study set out to create an accessible tool that can help teachers in easily assessing preschool students' responses and engagement towards different stimuli or teaching methods. The main contributions of this paper are the following:

- We formulated a quantifiable metric for relating emotion response to stimuli and engagement. This would be an additional metric that educators may use when evaluating emotional development and focus level in young children.
- We conducted a face detection performance analysis of browser-based facial landmark detection in real-world scenarios. This can serve as additional data when conducting further research into the subject.
- Developed an online platform that can aid in research for remote and alternative learning methods for the education sector. This serves as a proof of concept for the study.

II. Related Work

Using technology to gauge student engagement has been an interest even before the pandemic. Several works have tackled this issue with varying tools, levels of control, and methods.

A. Facial Expression, Engagement, and Learning

According to Teixeira et al. [6], emotion regulation works recursively around stimuli, emotion, attention, and behavioral response. A subject is exposed to a stimulus, if their facial and emotional response to the stimulus is 'typical' as to how most people would react to the stimulus, then they choose to stay exposed to it and is viewed in the eyes of another human as engaged. In a classroom, observing a child's physical (especially facial) response to the classroom discussion is one of the most basic tools in a teacher's repertoire. Empirical research has shown that student engagement is correlated with experiencing desirable outcomes such as critical thinking, cognitive development, psycho-social development, and many more [5] [7].

B. Recognition of Student Engagement using Machine Learning

Several studies have been conducted to measure a student's engagement, usually using surveys and observational assessment. Many of these studies started automating the observation process due to distance learning and the need for teachers to monitor students' reactions as close to real-time as possible.

Zhang et al. were able to assess learning engagement by gathering video data from the camera and recording mouse activity [8]. Their results showed that the dataset that used the mouse activity as reference had a higher recognition rate which is 94.60% unlike the 91.51% rate of the other set which was labeled by the students themselves. Aslan et al.'s [9] research analyzed screen capture, eye tracking, depth analyzers, and body posture to gauge engagement. They were able to detect perceived student engagement at 85% accuracy. Monkaresi et al. [10] used a camera recording of the subjects, ECG electrodes on the wrists of the subjects, self-reports during the writing activity, and self-reports one week after the activity. They used WEKA for their machine learning tool and used classifiers such as Updateable Naïve Bayes, Bayes Net, and Logical regression, among others. These studies require sensors which are not available to typical students.

The next two studies require no sensors or biosignal detectors and mostly depend on videos of students' face during an activity. Whitehill et al. [11] took videos of students interacting with an iPad and took three (3) approaches to labelling them. They applied linear Support Vector Machines (SVMs) and Gabor features to process the videos and images. They found that human judgements and automatic engagement judgements correlate with task performance. Alyuz et al. [12] had their own study aiming to detect student engagement while consuming educational content. They obtained 55.79%-90.89% accuracy from their different settings, with the 'Personal' model being most accurate.

C. Face Recognition

The expression of emotions have been shown to be related to the attention and engagement that a child dedicates to stimuli. The first step to identifying a child's expression is to detect their face in a video. Facial recognition technology was used to automatically identify faces in the webcam live feed.

1) face-api.js: Face-api.js [13] is a javascript module that aims to bring face recognition tools that are optimized for the browser. The module is able to detect faces and face landmarks. The tiny version of the 68 point face landmark detection model is only around 80kb which makes it suitable for lightweight browser applications.

The face detection models used by the module are the SSD Mobilenet V1 and the Tiny Face Detector. The SSD Mobilenet V1 model uses a Single Shot Multibox Detector based on Mobilenet V1. The model was trained with the WIDER FACE dataset which consists of 393,703 labeled faces.

The Tiny Face Detector model is a faster model than the SSD Mobilenet V1 that outputs face detection in real-time. It performs better on mobile devices and is less resource-intensive, making it a perfect choice for real-time web-applications. This model is similar to a variation of YOLO [14] called Tiny YOLO v2 but it uses depthwise separable convolutions instead of regular convolutions thus allowing it to be only 190kb.

D. Emotion Classification

After the face has been detected, the final step was to classify the detected facial features' emotion. FER2013 [15] is one of the most popular datasets and they label their facial expression using seven emotions, namely happiness, neutral, sadness, anger, surprise, disgust and fear. Microsoft's FER+ [16] was able to further improve the FER2013 dataset by labeling the data with 10 taggers thus having an emotion probability distribution per face. This enables output of statistical distribution for each prediction.

1) Perception for Autonomous Systems: Perception for Autonomous Systems (PAZ) is a hierarchical perception library in Python where the system has three tiered API levels that a developer can utilize [17]. The high-level API are specifically used directly with the application as a series of processing steps to get the desired output.

DetectMiniXceptionFER is one of the high-level functions of interest as it is an emotion classification and detection pipeline. The function is segmented into three parts: detection, classification and drawing. DetectMiniXceptionFer also uses another high-level function called HaarCascadeFrontalFace. This function detects faces using a model from OpenCV [18].

For classification, another high-level function called MiniXceptionFER which is used to classify emotions from RGB faces exists. This function uses a model called Mini-Xception that can be trained using different datasets. The Mini-Xception model was based on Chollet's Xception model [19] which highlights the depth-wise separable convolutions that reduces computation around eight or nine times [20] and utilizes residual modules [21].

2) EmoPy: EmoPy is an open-source Python toolkit that stands among the best Facial Expression Recognition (FER) systems available [22]. It includes several modules that can help build a trained FER prediction model from different

choices for neural network architectures and the flexibility of datasets.

There are five neural network architectures in EmoPy: ConvolutionalNN, TimeDelayConvNN, ConvolutionalLstmNN, TransferLearningNN and ConvolutionalNNDropout. These neural network architectures are an array of layers that feed outputs to each other sequentially in order to create a model When detecting emotions for new images, the ConvolutionalNN proved to be the best-performing [22]. For contrasting emotions, the ConvolutionalNN was correct nine out of ten times. For a very similar emotion set like ”anger, fear, and surprise”, the ConvolutionNN was able to correctly detect the emotions almost seven out of ten.

Both PAZ and EmoPy have features that allow them to augment any dataset that further increase the performance of the models [17]. The strength of EmoPy over PAZ is the number of neural network architectures for classifying emotions. You can pick and train EmoPy models depending on the dataset that you have. PAZ has only one emotion classifier model which is MiniXception but the model can also be retrained with other datasets. What PAZ lacks compared to EmoPy is made up by the robustness of the hierarchical API that makes development easier in all levels.

III. Methodology

The execution of the study is split between the preparatory work and the development of the Intellitell web app.

A. Preparatory Work

The preparatory work is meant to gauge the feasibility of the study and to prepare data that would be needed for the actual web app. The primary emotion classification technology, face-api.js, was tested and no breaking issues were found. A call for participants was also held in order to gather a ground truth data, which will then be used to compare to user data that will be gathered in the completed web app. A simple version of the web app was created in order to gather this ground truth data from 6 participants. A webcam feed of the students watching *Three Little Pigs - kids story — Fairy Tales — MNOP Stories for Kids* from the YouTube channel *MNOP Kids* was recorded and later manually labeled with emotion per time instance of 1 second. The results were used to directly compare emotion match accuracy in the next section.

B. Emotion Analysis

An example of using facial emotion recognition to assist online learning is when the ESG Business School in France [23] used emotion recognition in order to measure and keep student engagement. However, this study’s objective is to use emotion recognition as a tool for educators of young children. Thus, the method for emotion analysis was done via a web application that utilises facial expression recognition using javascript. A web app was chosen in order to decrease complexity in the interface that the young students will be using. This is because all that is required from the user is a browser, a camera, and an internet connection. The emotions detected were classified into

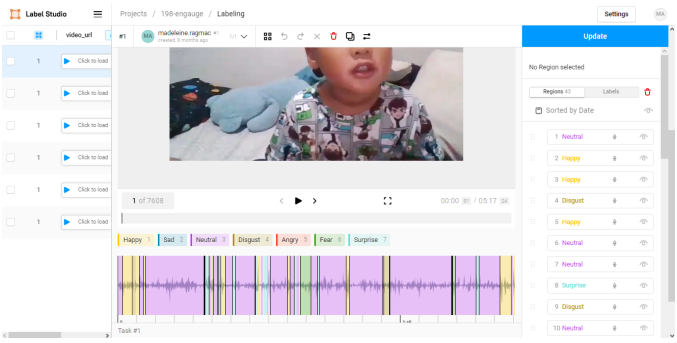

Fig. 1. Label Studio labeling screen

seven emotions following the FER2013 dataset [16] and each emotion was given a confidence score by face-api.js based on how confident the model was in its analysis. (e.g. happy 0.41, neutral 0.1, surprised 0.07, etc.).

1) Intellitell Web Application: The Intellitell Web App landing page is at `www.intellitell.org`. ReactJS for front-end and NodeJS with ExpressJS for back-end were chosen because they were among the web development frameworks with easily-accessible documentation, tutorials, and extensions. The technology used for emotion recognition is the Javascript library, FaceAPI.js, chosen for its lightweight design that best fits the web app format.

In the web application, ’student’ accounts will have a list of videos that have been made available to them by a ’teacher’ account. Upon playing a video, the system starts reading emotion data from the webcam feed. When the video ends, the modal closed by itself when it has finished sending the data it gathered to the database. A ’teacher’ account can add students, assign videos to students, and view the plot of the student’s facial expressions and their corresponding confidence score against the ground truth. All videos must be curated by ’teacher’ accounts for the students.

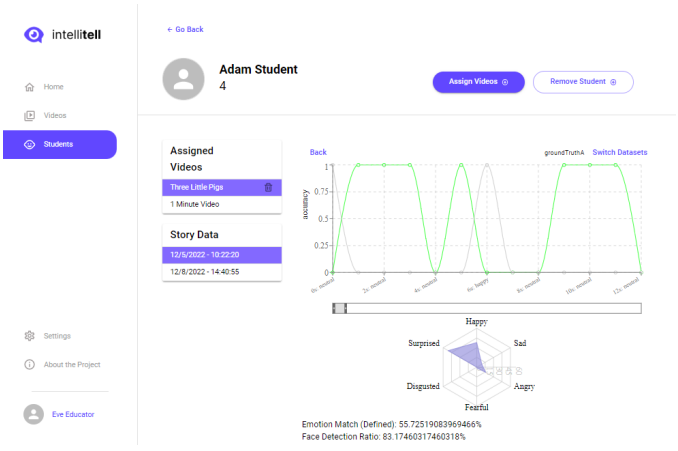

Fig. 2. Teacher’s view of student’s videos and watch instances

2) Database Design: Intellitell’s back-end uses REST API and is hosted in Firebase functions. There are thirty-six API calls made to complete Intellitell’s basic functions. The data

that these API calls interact with are stored in a Firebase Firestore database and in Firebase Storage (for profile pictures). Intellitell's Firestore Database is made up of five collections: users, educators, students, stories, and typesense_sync.

IV. Results and Analysis

To determine the performance of the system, participants were asked to use the completed web application in order to simulate actual users. This validation phase was able to gather 6 students to use the application. The only story available to these students was the same *Three Little Pigs* video that the ground truth was gathered from. Table I shows a snippet of the confidence scores from the application for Student 1 from 8-11 seconds.

TABLE I
SNIPPET OF DATA FOR STUDENT 1 FROM $t(8)$ TO $t(11)$

Emotion	t(8)	t(9)	t(10)	t(11)
Angry	0	0	2.00e-04	1.00e-04
Disgusted	0	0	0	0
Fearful	0	0	0	0
Happy	0	0.9119	0.0015	0.0019
Neutral	1	0.0878	0.9980	0.9980
Sad	0	3.00e-04	1.00e-04	0
Surprised	0	0	1.00e-04	0

A. Individual Analysis

To compare it to the ground truth data, the emotion with the highest confidence score is determined using the $[M, I] = max(A)$ function, where I is the index where the confidence score is the highest. Using this, the data was transformed to represent nominal values where each emotion had an assigned value. In instances that the system did not detect the face - where all emotions' confidence scores were set to 0, the emotion was undefined and was pruned from the dataset. In sub-optimal confidence scores, a minimum threshold of 90% was considered as a valid emotion since relatively low confidence data points were not adequate representations of the set and therefore had questionable accuracy. Otherwise, the emotion for that time instance is said to be undefined.

1) Emotion Match Accuracy: The chosen metric for engagement is emotion match per time instance. To compute for emotion match accuracy, we used the formula below where $n_{defined}$ is the number of time instances where emotion is defined and n_{match} is the number of time instances where the defined measured data matches the ground truth. This metric was used to determine how close the measured data follows the ground truth.

$$EM = \frac{n_{match}}{n_{defined}} \tag{1}$$

2) Face Detection Ratio: An accompanying metric, face detection ratio, was also computed below where $n_{defined}$ was the number of time instances where emotion was defined and n_{total} was the total seconds the measured data was recorded. The face detection ratio showed the performance of the system in detecting the face, and giving optimal confidence scores. The face detection ratio was also affected by external factors like participants going out of the frame of the camera, sub-optimal lighting conditions, head-turning away from the camera, and a distracting environment. Even with guidelines given, these factors were considered in interpreting the data as the experiment was performed in a limited uncontrolled testing environment.

$$FD = \frac{n_{define}}{n_{total}} \tag{2}$$

3) Exponentially Weighted Average: An exponentially moving weighted average was also plotted with $\alpha = 0.095$ for a window size of 20 seconds to observe the general movement of emotion and still be sensitive to shifts with small time periods. It is computed using the formula below where r_t is the current value of measured data.

$$\text{EWMA}_t = \alpha * r_t + (1 - \alpha) * \text{EWMA}_{t-1} \tag{3}$$

Figure 3 shows Student 1's measured data plotted with the ground truth and EWMA. With a face detection ratio of 53.00%, the emotion match was accurate 91.07% of the time. We saw that the EWMA tends to stick and approach along the neutral emotion which shows that Student 1's non-neutral reactions are only for short periods of time as we can see in the plot.

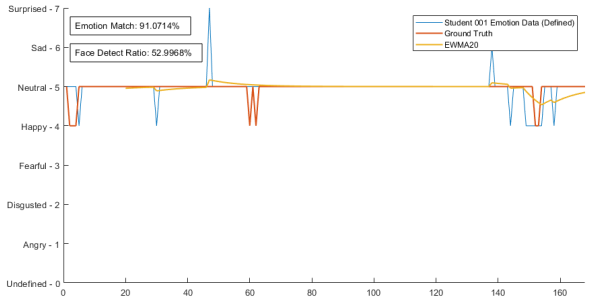

Fig. 3. Results - Student 1 Analysis

The breakdown of all students' performances are found in Table II. The face detection ratio was the percentage of time that faceAPI was able to identify the child's whole face and the emotion they were displaying. The data returned when a face is not detected was undefined and so did not contain any emotion data. Here we can see the sub-optimal performance of the system in terms of face detection with an average of 42.14% across all students.

From the six participants, the third column from Table II are computations of the emotion match accuracy including all observations of undefined values and sub-optimal confidence scores. To provide a better comparison for the results, the fourth column shows computations of the emotion match accuracy that were defined and passed the threshold of 90%. We observed that four out of six participants have an emotion match accuracy higher than 87%. Similar to Student 1, the

TABLE II
BREAKDOWN OF METRIC FOR THE VALIDATION PHASE

Student	FD	EM - All	EM - Defined
001	53.00%	51.25%	91.07%
002	43.06%	45.91%	88.43%
003	41.32%	47.50%	87.02%
004	39.12%	32.45%	66.94%
005	49.53%	44.41%	52.23%
006	26.81%	28.04%	95.29%
Average	42.14%	41.43%	80.16%

Students 2, 3, and 6 showed similar behavior wherein their EWMA tended to stick to the neutral emotion most of the time since their non-neutral emotions were only recorded for short periods of time.

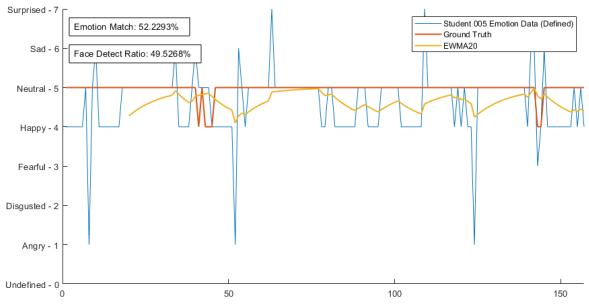

Fig. 4. Results - Student 5 Analysis

Students 4 and 5, however, were divergent from the population: they had more non-neutral emotion than their counterparts. They specifically had longer time periods wherein they evoked the emotion of 'happy' than others. As we observed the EWMA of these students, they showed more movement towards other non-neutral emotions compared to the other students.

B. Overall Performance

Further analysis was performed to see how well the ground truth correlated with the overall behavior of the measured subjects.

1) Mean of Measured Data: We calculated for the mean M of the measured data using the formula below where n is the number of participants and c_i is the principal emotion detected per time t. This formula was also used to compute for the ground truth mean in the nominal scale.

2) Confidence Interval: To show the variance of the possible true values of the measured data, the confidence interval was calculated. First, we needed to calculate for the standard error of the mean SEM, where σ was the standard deviation of the measured data, and n was the number of students. T-score was also generated using MATLAB Statistics Toolbox's tinv function. The formula for the confidence interval was found below where M was the mean of the measured data. This gave us an upper and lower bounds where the possible true values of the measured data will fall in. The 50%, 60%, 70%, 80%, and 90% Confidence Intervals were calculated. A snippet is shown in Figure 5.

$$SE = \frac{\sigma}{\sqrt{n}} \tag{4}$$

$$ts_{80\%} = \text{tinv}([0.1\ 0.9],\ \text{n} - 1) \tag{5}$$

$$CI = M + ts * SE \tag{6}$$

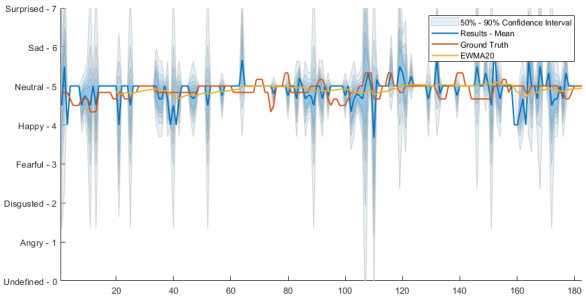

Fig. 5. Snippet of Plot from 0 - 190 seconds with the Undefined Emotion Readings Removed

Figure 5 shows the plots of the ground truth and the mean data of the measured subjects. Also plotted is the Exponentially Weighted Moving Average (EWMA) with a window size of 20 to lessen the unrealistic and abrupt change of values between emotions in a short amount of time. The y-axis is based on a nominal system named after emotions, but it is still possible to see where the values tended towards on this scale.

At first glance, we observed that both the ground truth and the mean of the measured data settled around neutral which is expected as we have observed from each individual participant having neutral as the dominant emotion for the length of the measured data.

In the ground truth plot which was mostly neutral, it was noted that there were specific points where the 'happy' emotion was more dominant. In context, these points correspond to when the storytelling video was at the channel jingle, the building of the straw house, and the ending. Two out of three ground truth children who responded to the channel jingle also displayed non-neutral emotions more often. Others, while not expressing non-neutral emotions as often, looked engaged to a human observer because of how they seem focused on the screen, possibly explaining the divergence from the expected 'happy' time slices.

Due to the low sample size, the ground truth is susceptible to distractions. At around 210 seconds, one subject was distracted and seemed to have been bitten by an insect. This pulled the ground truth downward towards disgusted. In another instance, one child made a happy face but it was in response to them making fun of someone off-camera. More participants would have been able to even out small distractions because each subject would have less of a pull on the ground truth mean.

At around 250 seconds, there seemed to be a discrepancy where the measured subjects' mean both rise towards surprise and pull down towards fearful. In context, this part of the story corresponded to the wolf being unable to blow the brick house down and deciding to climb the chimney. FaceAPI and the manual labeler, who had no context to what the subject would have been reacting to, may have labeled the faces at this time to be either surprised or fear, since both would share similarities in how they are displayed, especially when the subjects were not trying to make a point to differentiate between them. While context was removed during labelling to prevent bias due to expecting an emotion from the children, there might have to be exceptions made for similar-looking emotions to be able to interpret them properly.

V. Conclusion and Future Work

This study on gauging student engagement using a web app platform was able to get a fairly accurate reading of a student's engagement when compared to the perception of a human observer. Further development can help increase accuracy while still being accessible to many people. This line of study can help teachers maintain student engagement and improve learning in distance learning and even face-to-face setups.

We recommend the following activities for future work. First, gathering a bigger sample size is imperative to further strengthen and validate the model. Second, limiting the external factors on the testing environments (such as sub-optimal lighting conditions, distracting environments etc.) during data collection. Third, saving the video recordings as a fallback if the emotion recognition pipeline fails to return useful data. Lastly, future research on the improvement of accuracy through the following ways:

- Looking into using training data that better represent the age, culture, and skin color of the subjects
- Recognizing whether or not a head is turned towards or focused on the screen, or if the eyes of the participant are focused on the screen.
- Calculating the system's precision, recall, and F1 scores.

References

[1] J. Ismael, "Survey: Distance learning not working."

[2] H. Morgan, "Best Practices for Implementing Remote Learning during a Pandemic," *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, vol. 93, no. 3, pp. 135–141, 2020. [Online]. Available: https://doi.org/10.1080/00098655.2020.1751480

[3] T. Yates, M. M. Ostrosky, G. A. Cheatham, A. Fettig, L. Shaffer, and R. M. Santos, "Research synthesis on screening and assessing social-emotional competence," 2008.

[4] H. S. Han and K. M. Kemple, "Components of social competence and strategies of support: Considering what to teach and how," *Early Childhood Education Journal*, vol. 34, pp. 241–246, 12 2006.

[5] V. Trowler and P. Trowler, "Student engagement evidence summary," 2010.

[6] T. Teixeira, M. Wedel, and R. Pieters, "Emotion-induced engagement in Internet video advertisements," *Journal of Marketing Research*, vol. 49, no. 2, pp. 144–159, 2012.

[7] C. W. Sandra L. Christenson, Amy L. Reschly, *Handbook of Research on Student Engagement*, 2012.

[8] Z. Zhang, Z. Li, H. Liu, T. Cao, and S. Liu, "Data-driven Online Learning Engagement Detection via Facial Expression and Mouse Behavior Recognition Technology," *Journal of Educational Computing Research*, vol. 58, no. 1, pp. 63–86, 2020.

[9] S. Aslan, Z. Cataltepe, I. Diner, O. Dundar, A. A. Esme, R. Ferens, G. Kamhi, E. Oktay, C. Soysal, and M. Yener, "Learner engagement measurement and classification in 1:1 learning," in *Proceedings - 2014 13th International Conference on Machine Learning and Applications, ICMLA 2014*. Institute of Electrical and Electronics Engineers Inc., 2 2014, pp. 545–552.

[10] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Transactions on Affective Computing*, vol. 8, pp. 15–28, 1 2017.

[11] J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, pp. 86–98, 2014.

[12] N. Alyuz, E. Okur, E. Oktay, U. Genc, S. Aslan, S. E. Mete, D. Stanhill, B. Arnrich, and A. A. Esme, "Towards an emotional engagement model: Can affective states of a learner be automatically detected in a 1:1 learning scenario?" 2016.

[13] V. Muhler. (2018) face-api.js. [Online]. Available: https://github.com/justadudewhohacks/face-api.js

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 779–788, 2016.

[15] P.-L. Carrier and A. Courville. (2018, 5) The Facial Expression Recognition 2013 (FER-2013) Dataset. Wolfram Research. [Online]. Available: https://datarepository.wolframcloud.com/resources/FER-2013

[16] E. Barsoum, C. Zhang, C. Canton Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *ACM International Conference on Multimodal Interaction (ICMI)*, 2016.

[17] O. Arriaga, M. Valdenegro-Toro, M. Muthuraja, S. Devaramani, and F. Kirchner, "Perception for autonomous systems (paz)," 2020.

[18] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1800–1807, 2017.

[20] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," 2017. [Online]. Available: http://arxiv.org/abs/1704.04861

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, 2016.

[22] A. Perez. (2018, 9) EmoPy: A Machine Learning Toolkit For Emotional Expression. ThoughtWorks. [Online]. Available: https://thoughtworksarts.io/blog/emopy-emotional-expression-toolkit/

[23] C. Garcia-Montero, "L'ia rappelle à l'ordre les étudiants," *Alliancy le mag*, 2017. [Online]. Available: https://www.alliancy.fr/lia-rappelle-a-lordre-les-etudiants