Child Attention Detection through Facial Expression Recognition using SVM Algorithm

Aika Patricia Baldovino
School of Information Technology
Mapua University
Makati City, Philippines
+639999073838
aprbaldovino@mymail.mapua.edu.ph

Frances Neele Vergonio
School of Information Technology
Mapua University
Makati City, Philippines
+639176723400
fnbvergonio@mymail.mapua.edu.ph

John Paul Tomas
School of Information Technology
Mapua University
Makati City, Philippines
+639178695257
jpqtomas@mapua.edu.ph

ABSTRACT
Determining the ability of children to focus in a very young age is something that is very important for adults to know for them to understand the child's learning capability. The development of their attention skills during their younger years affects how much excellence they can perform during their adolescence stage. Since most attention detection researches are frequently done through eye gaze detection, this research is focused on detecting the attention of a child through facial expressions. The proposed system shows that basing from a child's facial expression, it can determine their attention skills which has given accurate results. A total of forty (40) grade one (1) students took part in this research. The data gathered was in the form of a recorded video obtained from the web camera. Each video was processed frame by frame to extract necessary facial features that is needed in determining the facial expression through OpenFace application. SVM algorithm was used in training and testing the model's validity. The model is written in a Java Programming Language and has an output of a subtitle file which will be imported into the recorded video. From there, the subtitle file has a label of the student's facial expression, thus determines their attention. To determine the predictive power of the model, K-fold cross validation method was used.

CCS Concepts
• **Computing methodologies→Computer vision • Computing methodologies → Image and video acquisition**

Keywords
Attention Recognition; Expression; Recognition; Child Attention, Computer Vision; Image Processing; Support Vector Machine

1. INTRODUCTION
Attention is considered essential for learning. Educators often discuss attention as a general mental state where the mind focuses on a particular feature of the environment [1]. There is good evidence that attention skills firmly predict educational achievement for all students, not just those suffering from ADHD [2]. An effective attention system has effects that are beneficial which can be noticed very early in development.

The attention skills of a child can predict his or her intelligence as an adolescent creating a difference of up to 20 IQ points in combination with environmental differences [3]. In several academic institutions, students are now required to have their own laptops or tablets, stable internet connection, and most instructors prefer to use e-learning tools in discussing their lessons and administer assessment of student understanding. Navigating on the internet can result in cognitive overload for an average user, and for a student with ADD/ADHD it may be impossible to stay focused and remain on-task [4]. The analysis of facial expression is quickly starting to become a scope of great interest in the society of human computer interaction design and computer science. The most demonstrative and suggestive way humans show their emotions is through their facial expressions. According to the child psychologist that the researchers have consulted [5], examining the child's facial expression will help determine the child's emotional state and then identify how these emotional states are linked to the child's comprehension and understanding of the subject.

The combination or set of algorithms applied in previous studies [9] [11] [12] [18] relating to attention detection and facial expression recognition focused on the area of the eyes, particularly the eye gaze of the person. This paper gives an opportunity to do further research by adding more facial features such as the mouth and brows hoping to contribute to the progressing research on modern developments of more accurate models for attention detection system. In addition to this, it provides an opportunity to contribute studies that are focused on students, especially children, as their subject of interest in a e-learning environment.

2. REVIEW OF RELATED LITERATURE
Children in a young age need skill to attain success and achievement at school. For these children to direct themselves to school settings effectively, they should have the capability of focusing their attention to their teacher and completing their tasks even with some distractions surrounding them. Working memory and attention are the strongest predictors of the academic growth of the children for both reading comprehension and math. Changes in working memory and attention predicted an increase in the academic skills of the children who advanced through their early elementary grades. Since the children are paying attention, their academic achievements and competence belief values become more stable as they advance through higher grades of elementary [6]. An engaged student is not only open to gaining knowledge but is more likely to retain the gained information as well. However, there is no single guide in determining a person's emotional state of mind [7]. In estimating attention, there is no standard method. To solve the problem, many approaches have been developed. A

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ITCC 2019, August 16–18, 2019, Singapore, Singapore
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-7228-2/19/08…$15.00
DOI: https://doi.org/10.1145/3355402.3355411

thorough overview of subjects like brainwave readings, body temperature and taking in measurements are some methods involved. While these measurements are thorough and extensive, they are also intrusive. Intrusive measures can often be uncomfortable for the user, which in turn can lead to errors in the collected data. Additionally, they often require more physical setup, adding an additional cost to be considered. As a result, other researchers have turned toward less-intrusive approaches instead. These measures require less active thought from users but can still generate valuable results.

An example of an intrusive approach is done in [8], whereas the study is conducted by researcher who proposed a novel approach to automatic estimation of attention of students during lectures in the classroom. The approach uses 2D and 3D features obtained by the Kinect One sensor characterizing both facial and body properties of a student, including gaze point and body posture. Machine learning algorithms are used to train attention model, providing classifiers which estimate attention level of individual student. Human encoding of attention level is used as a training set data. The Kinect One sensor was set up to observe three (3) students acting as test persons. The software used to record the incoming data stream to the hard drive was the Kinect Studio. The recording data rate went up to 120 MBps. Matlab scripts were used for data analyzing, which is provided by Kin2 Toolbox for Matlab, which encapsulates the Microsoft Kinect 2 SDK. The first real-time pass of analysis was intended to capture the video and skeletons data and store them on the disk drive. Then, the offline analysis of extracted data was performed by Matlab scripts. Three (3) Features were used to for the automatic student attention measurement system: *Body Posture, Face Gaze Point, and Facial Features.* Kinect SDK provides an estimation of the head gaze. Using the head position in the 3D camera space, formulas and the Kinect sensor position in the world space *K'*, the researchers have calculated projection of the head gaze onto the x – y plane in the world coordinates, resulting in the 2D world gaze point coordinates. Facial features are derived from the 17 Animation Units computed from detailed 3D face model, which are used to characterize observable behaviors such as yawning and writing. Logistic function is used to preprocess the original values.

1. *Eyes Closed: is computed from the RightEyeClosed indicator, and corresponds to writing and observing notes.*
2. *Mouth Open: is computed from the JawOpen indicator and corresponds to yawning.*
3. *Face deformation: is computed from the LeftcheckPuff indicator, and corresponds to supporting head with the left hand.*

The researchers have compared their overall accuracy in estimating attention level using 10-fold cross validation. Linear discriminant classifier and Simple Tree classfier resulted in accuracy of 75%. Best accuracy was achieved using Bagged Trees, ranging from 85.0 to 86.9% depending on method paramaters.

In [9], a real time non-obstructive system of tracking attention is proposed by Narayanan et al., using a normal web camera. This system is uniform to rotation and scale and has tolerance to classifications of false attention states. Classification of student attention states are divided into three types: disappeared state, sleepy state and attentive state. A geometric model for the detection of the corners of the eyes is also proposed. Experiments of passive and active tracking of attention is done by a fifty-four-minute video footage presenting as the e-learning lecture. From the live video stream, face region is detected using Haar-Cascade Classifier trained for eye detection. In the formula that the researchers have used, the distance between the inner corner and outer corner of an eye is one-third the inter-eye distance. Once the inner and outer corner points are detected in a frame, the corner points will be tracked in the subsequent frames using Lucas-Kanade tracker. The ROI image is morphologically opened to remove the highlights present in the eye due to the computer monitor and other light sources. The morphologically opened image is subtracted from the ROI image. Attention is measured as the average intensity in the ROI region after the subtraction operations. The attention tracking system is real time, scale and rotation invariant and tolerant to blink related false accusations. Logitech Webcam Pro 9000 is used in the experiment where the video is captured at the resolution of 480x640 and 30FPS. The attention tracking system is prototype in Matlab and implemented using OpenCV, where the execution time of the proposed algorithm for key frames (including face detection, eye detection and corner detection) is 36 milliseconds and for other frames (which involves Lucas-Kanade tracker) is 10 milliseconds.

In [10], the researchers have conducted a non-intrusive approach that focuses on the gaze data obtained through an eye-tracker which has a goal of creating a predictive model of students' attention in a classroom setting. The eye-tracker used to collect the eye-gaze data is known as Tobii Eyetracker, model 4c. The participants were from undergraduate classes in the field of mechanical engineering which are composed of Junior level students and had approximately 25 students in attendance. The process of eye gaze data measuring is especially significant because it collects measurements in a non-intrusive way while also being easy and simple to set up and use. In this study, data acquired from an eye tracker can certainly be used to estimate and predict the attention of a student as measure of affect over the whole period of a class. The set-up of the classroom was in the style of computer laboratory room. The students were provided with each of their own desktop computers to use at their designated desks with PowerPoints as lectures. The eye trackers were secured at the bottom of each of the monitors. The eye trackers were carefully placed in the position where it won't be a distraction to the students and won't be a factor in change of behavior or attention. From their research, 77% of accuracy has been achieved using the technique of Extreme Gradient Boosting of machine learning. The result had indicated that eye gaze can certainly be used as a basis for establishing a predictive model.

In [11], the researchers developed a system that monitors the students' behavior and estimate how effective the teachings approaches are. It aims to monitor the emotional status of each student and provide information about student's mood to the teacher in real time and save collected data for later analysis. Their goals were to develop an attention system that can measure the level of student's attention and increasing the precision of identifying the student's concentration. The thresholds of fidgeting levels are defined according to the usual movement behavior of students. A range of fidgeting levels supports classifying more precisely the hypothetical emotional model of students. Regarding a process of monitoring face expressions, body movements and level of noise, the research suggests five behavioral models concerning attention (concentrated, calm, bored, excited and sleepy). The researchers suppose the facial expression to be of primary importance together with body movement and noise level. Concentrated model is described with neutral face facial expression and low levels of noise and fidgeting. Student behavior is characterized as calm when face expression is defined as neutral, happy or sad while a fidgeting level is low, and the noise level is low or medium. Both models bored and excited present more meaningful face expressions and medium or higher levels of fidgeting and noise. Defining a model sleepy relies on detecting sleepy eyes, low level of noise and low or not available. Absent is adopted for the proper functioning of the system in case of one or both student's absence (see Table 1). The researchers used a CK+ dataset that consists of 448 images of different people in eight emotional conditions*: angry, happy, disgust, fear, neutral, sadness, contempt* and *surprised*.

Table 1. Description of Behavioral Models

Behavior Model	Face Expression	Noise Level	Fidgeting Level
CONCENTRATED	Neutral	low	low
CALM	Neutral/ happy/ sad	low/ medium	low
BORED	Happy/ angry/ disgust	medium/ high	medium/ high
EXCITED	Happy/ surprise/ disgust	high	medium/ high
ABSENT	NA	Low/ medium	NA
SLEEPY	sleepy	low	low

In [12], a new feature-based approach for facial expression recognition was done by the researchers which provides a fully automatic solution to identify human expressions as well as overcoming facial expressions variation and intensity problems. Facial features were detected namely*: eyes, eyebrows and mouth* using vertical and horizontal projection. Active contour was then applied to segment these facial features with a reason that it gives more close and natural representation of the detected feature shape. After segmenting the facial features needed, the relevant facial features points were extracted which define the prominent landmarks surrounding facial components. The points of interest were identified in order to compute relative distances between facial features was also measured. Some of the most challenging problems while recognizing facial expressions are the unlimited number of facial expressions as well as the intensity variation while giving voice to one's emotion. SIPINA algorithm was used though the use of fusion operator and sensitive measurement between features to reduce the drawback of tree methods. The researchers have detected feature points of the eye, brows and mouth. A total of seven (7) facial expressions were described: JOY, SMILE, SURPRISE, DISGUST, ANGER, SADNESS AND FEAR.

Table 2. Feature Distance Definitions

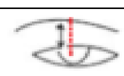
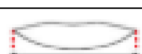
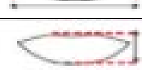
D0		Distance between the center of the eye and the horizontal axis of the detected eyebrows
D1		Distance between the upper and lower eyelid
D2		Distance between the inner corner of the eye and the inner corner of the eyebrow
D3		Mouth width
D4		Mouth opening height
D5		Distance between the corner of the mouth and the corresponding outer corner of the eye
D6		Distance between the corner of the mouth and the corresponding eye center

Table 2 shows the six (6) characteristic distances from (D_0 to D_6) where D_0 is defined as the distance between the center of the eye and the horizontal axis of the detected eyebrows, and D_1 is defined as the distance between upper and lower eyelid. D_2 is defined as the distance between the inner corner of the eye and the inner corner of the eyebrow, and D_3 is the mouth width, while D_4 is the mouth opening height. D_5 is the distance between the corner of the mouth and the corresponding outer corner of the eye and lastly, D_6 is the distance between the corner of the mouth and the corresponding eye center. The researchers have considered the intensity of JOY expression, therefore defined the expression SMILE as the lowest intensity of JOY.

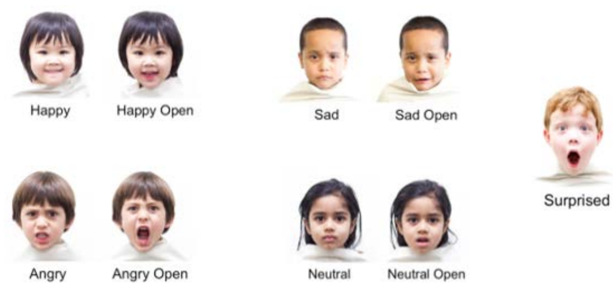


Figure 1. Basic Facial Expressions

Images introduced above (Figure 1) are some examples from the Child Affective Facial Expression set (CAFE) — a new stimulus set of emotional facial expressions into the domain of research on emotional development. The CAFE set features photographs of a racially and ethnically diverse group of two to eight-year-old children [13]. Facial expressions such as happy, angry, sad and surprised are the basic expressions that children from ages six to seven years old express through their faces. When a child is watching animations that show basic expressions and is also reflected to their own facial expression while watching, it is a sign that they are focused, and their attention is diverted to the animation. It also shows how empathetic the child is. A child may be focused and attentive on the animation but do not reflect the expression on their faces. This is because they lack empathy towards the situation being presented.

3. MODEL DESIGN AND EXPERIMENT

3.1 Model Design

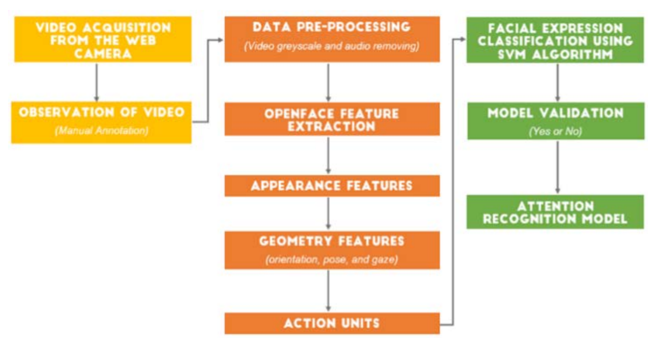


Figure 2. Model Design

Figure 2 shows the model design that this study used. The data was gathered through recording the faces of the students while they watch an e-learning video. The recorded video of the student's face while watching an e-learning video should undergo through processes to eliminate unnecessary collected data and identify those that are significant. Total of forty (40) students took part in the experiment. The training data set is made up of twenty (20) student values, and the remaining twenty (20) were left for testing.

After the video has been acquitted from the web camera, each video file is manually annotated by the child psychologist. Each video file was pre-processed by deleting unnecessary parts of the video, muting the audio, and turning it into greyscale. This means that parts of the video where students do not show any kind of facial expression: *being happy, sad, surprised, angry and neutral* was removed from the video. This study does not present a real-time attention detection system given that the videos underwent processing after it has been recorded. It was not designed with the purpose of delivering a real-time detecting tool, but to evaluate the potential of the methodology. Once pre-processed, the OpenFace application tool was used for extracting the facial attributes needed which are the eye gaze, action units, head pose, 2D and 3D landmarks of the student's face. The OpenFace application has an output of .csv file of each recorded video which contains the coordinate values of the extracted facial features that are necessarily needed to identify which facial expression does the student expressed. This study used a software tool called RapidMiner Studio to conduct the testing and training of data using a machine learning algorithm called Support Vector Machine (SVM). When data training had finished, testing of the remaining twenty (20) visual data gathered would commence. Data gathered were divided into subsets in which twenty (20) student data sets were used as training data, while the remaining twenty (20) student data sets were used for testing and were not included in training the data. The data is validated through Accuracy, Recall and Precision with the use of K-fold cross validation. The model of this study is made using Java programming language. The model will output a subtitle file which will be imported in the recorded video, whereas that subtitle file will have a label of "attentive*/facial expression*". The output facial expression depends on what the expression does the student showed. The model's detecting power is decided by a confusion matrix based on the perceptual evaluation by the Child Psychologist that serves as the baseline of the action whether the student is attentive or not. The model's output was compared to the perceptual evaluations to give a reason for whether the model is accurate for detecting the attention of the students based on their facial expressions when they watch e-learning videos.

4. METHODS

4.1 Data Gathering

The researchers conducted their data gathering at an Elementary School with a total of forty (40) children from first grade students who came from the same section. Figure 3 shows the sample system set-up where the subject was seated in front of a laptop where the web camera was mounted. The web camera was placed in the angle where the entire head of the subject can be completely recorded. During the time of recording, the subject was asked to watch some educational videos.

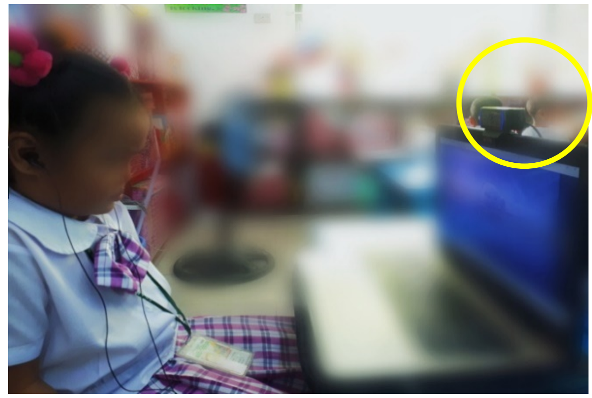


Figure 3. Sample System Set-up

4.1.1 Data Pre-processing

Once the web camera has accumulated the student's face and the visual data has been gathered, all recorded videos will undergo a pre-processing method. Since the web camera is the only tool that were used to collect the data, auditory cues are not necessary for detecting the attention of the student, thus the researchers only focused on visual media. Additionally, the recorded videos were converted into greyscale to reduce the memory space which allows the model to process the data easier and could probably produce faster results. This pre-processing allows the model to function efficiently and possibly, could bring more accurate results.

4.2 Feature Extraction

Figure 4. OpenFace Facial Recognition Process

The recorded videos were processed to extract facial landmarks, eye gaze and action units from the student. The process of detecting attention was done through eye gaze tracking, action unit detection, head pose orientation, 2D and 3D landmarks. In Figure 4, the *OpenFace_2.0.5_win_x64* application was used as a tool for feature extraction the appearance features, geometry features and action units of the student's face. The OpenFace software is a facial behavior analysis toolkit developed through an integrated implementation of the Multicomp Group, Technologies Institute at the Carnegie Mellon University and Rainbow Group, Computer Laboratory, University of Cambridge [14]. It is a toolkit that contains a feature of recognizing the action units of a person's face, facial landmarks tracking, eye gaze tracking, and head pose tracking that generated a .csv file. The .csv file contains the values of the facial attributes of the students that were used to determine facial expression features. Convolutional Experts Constrained Local Model (CE-CLM) algorithm was used for landmark detection. The columns (gaze_0_x, gaze_0_y, gaze_0_z, gaze_1_x, gaze_1_y, gaze_1_z, gaze_angle_x, gaze_angle_y, eye_lmk_x_0 to eye_lmk_x_55, eye_lmk_y_0 to eye_lmk_y_55, eye_lmk_z_0 to eye_lmk_z_55, pose_Tx, pose_Ty, pose_Tz, pose_Rx, pose_Ry, pose_Rz, x_0 to x_67, y_0 to y_67, X_0 to X_38) were obtained from the output of the .csv file which contains the approximate values of the position of the accumulated facial attributes of the students.

4.3 Process

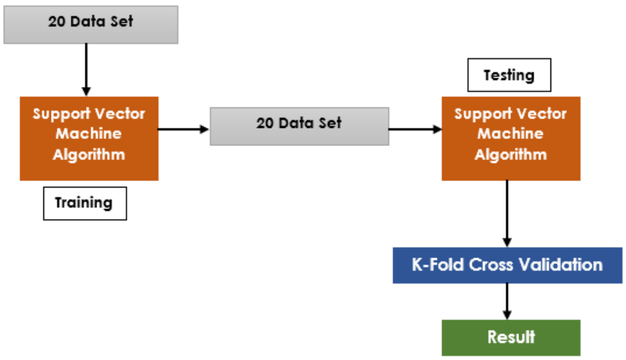


Figure 5. Model Using Support Vector Algorithm

A perceptual evaluation of the accumulated student's videos was performed by the Child Psychologist as per request, to manually assess if the student is attentive or not, also it is served as the baseline for this study to assure the accuracy of the model in detecting child's attention. Since the manual annotation served as a baseline, the model's output was compared to the annotation done by the Child Psychologist. Figure 5 depicts the flow of training and testing the data that were gathered. To develop the attention detection model, the researchers used a data of twenty (20) students for training and twenty (20) students, as well, for testing. The data were trained and tested using the RapidMiner Studio which is a world-leading open-source system for data mining which is easier to use for validating the results of data [15]. Support Vector Machine (SVM), a machine learning algorithm that analyzes the data for classification and regression analysis was applied. It utilized a method called the kernel trick to change information and after which depending on these changes finds an ideal limit between probable results, at that point the algorithm makes sense of how to separate the information based on the labels, results are characterized. K-Fold method was used for cross validation to avoid bias with the divisions of the data.

Table 3. Manual Annotation

CSV TIMESTAMP	ATTENTIVE EXPRESSION
0 – 14.982	NEUTRAL
15.015 - 20.988	HAPPY
23.023 – 30.998	HAPPY
31.031 – 32.966	SURPRISED
33 – 59.993	NEUTRAL
60.027 – 79.98	HAPPY
80.013 – 83.984	SURPRISED
84.017 – 90.991	HAPPY
91.024 – 97.998	SAD
98.031 - 99.967	SURPRISED
100 – 125.993	SAD
126.026 – 139.973	NEUTRAL
140.007 – 171.972	SAD
172.005 – 184.985	NEUTRAL
187.02 – 228.995	NEUTRAL
246.012 – 280.981	NEUTRAL
281.014 – 285.986	SAD
286.019 – 324.991	HAPPY
325.025 – 355.989	NEUTRAL
356.022 – 386.987	SAD
387.02 – 392.993	NEUTRAL
393.026 – 402.969	HAPPY
403.003 – 434.968	NEUTRAL

Table 3 shows the manual annotation of the Child Psychologist. After pre-processing method, all the recorded videos were manually annotated by the Child Psychologist with the help of the researchers which is used as the baseline for this study, to identify whether the model is accurate in detecting child's attention when they watch an educational video.

4.4 Model Validation

A model is developed which uses the Support Vector Machine (SVM) as the machine learning algorithm. The model is validated through k-Fold cross validation *(k = 20)* because by using the K-fold method for validating the data, it can evaluate how well the model would perform if the values being implemented was not learned by the model. K-Fold Cross Validation (CV) provides a solution to this problem by dividing the data into folds and ensuring that each fold is used as a testing set at some point [16]. In this study, the researchers used twenty (20) subsets for training the model and one (1) subset which is for testing. With the K-Fold method, the process is repeated twenty (20) times, wherein different subset of student's data is used as the testing data per time. This method is used because it proved to have better confidence in prediction accuracy. To analyze the data gathered, the researchers have used the *Accuracy, Recall, and Precision* as the methods for final validation [17].

5. RESULTS

After gathering the data, pre-processing method, extraction of the facial features needed, and applying the Support Vector Machine (SVM) algorithm, the student set used for testing the data has a video length of five (5) minutes and twenty (20) seconds. The facial

expressions served as the baseline in detecting the attention of the student in this study. The facial expressions that were detected are as follows: **HAPPY, SAD, ANGRY, SURPRISED AND NEUTRAL**. The researchers have considered neutral as part of the facial expression detected because most students have shown neutral expression throughout the experiment.

The result of the Accuracy, Precision, Recall and Classification error are stated in the statistical table (Table 4).

Table 4. Statistical Table

PARAMETER	PERCENTAGE VALUE
Accuracy	74.49%
Precision	79.20%
Recall	73.55%
Classification Error	25.51%

Table 4 depicts the statistical table of the model's results. The researchers have a 25.51% classification error that occurred during the testing of the model.

However, the result of the accuracy peaked at 74.49%, the precision at 79.20%, and the recall at 73.55%. These result shows that even though the model has a classification error of 25.51%, the accuracy, precision and recall is still higher than the misclassification of the facial expressions that were detected. The results from this study signify that facial expressions *(a combination of facial landmarks, eye gaze, and action units)* provide higher detection power for the specific data that was used for testing.

	true DISTRA...	true HAPPY	true NEUTRAL	true SAD	true ANGRY	true SURPRI...	class precisi...
pred. DISTR...	771	17	33	23	3	38	87.12%
pred. HAPPY	89	1395	61	7	20	12	88.07%
pred. NEUTR...	441	43	1576	42	3	258	66.69%
pred. SAD	135	9	32	1227	11	43	84.21%
pred. ANGRY	243	31	198	201	1063	175	55.63%
pred. SURP...	21	5	0	0	0	374	93.50%
class recall	45.35%	93.00%	82.95%	81.80%	96.64%	41.56%	

Figure 6. Performance Vector of the Model

Figure 6 shows the performance vector of the result of the model. Seen above are the calculated true positive, true negative, false positive, and false negative. The predicted 'distracted' has a result of 771 frames, while the predicted 'happy' has a result of 1395 frames. The predicted 'neutral' expression has a result of 1576 frames and the predicted 'sad' has a result of 1227 frames. On the other hand, the predicted 'angry' has a result of 1063 frames and lastly, the last expression which is a predicted 'surprised' has a result of 374 frames, making it the least of them all. The model has detected few angry expressions as well because according to the researchers after they had an interview with the students, the e-learning clips provided does not make them get mad that's why the predicted angry expression is relatively low, making it second to the lowest. With that said, the researchers can identify the attention of the students based on their facial expressions because the model has shown a good number of accuracy in detecting their facial expression.

6. CONCLUSION

This study shows evidence that facial expressions can be a means to detect and recognize attention through children of younger ages. The more likely they react and express themselves through their faces, the possibility of them being attentive is higher. The subjects that were detected with multiple expressive faces resulting them to be very attentive, were also the same subjects that had higher scores during the test and interview.

The accuracy of detecting the facial expression of the subjects through the given regions of interests (brows, eyes, mouth) and necessary data – appearance features, geometry features and action units of the student's face is limited to the capability of the software that does the feature extraction. Constant head movements, side view faces and positions where the entire region of the face cannot be completely seen or detected by the software decreases the accuracy of determining the facial expression. Upon acquiring the features, the model acquired a 74.49% for accuracy in detecting attention.

The researchers recommend using a longer video for the subjects to watch since the researchers had only used a 10 – 15-minute video. This would also give the opportunity for them to compare their expressions if they are still that expressive during the first few minutes and during the last few minutes. The facial expressions in this study were limited to happy, sad, angry and surprised, it would be better to add more facial expressions for future researchers. And lastly, trying this in different age groups to see if children from younger or older ages can still use facial expression to detect attention.

7. ACKNOWLEDGMENTS

The researchers would like to thank Mapua University for allowing them to conduct and work on this research. They would also like to express their gratitude to the faculty of School of Information Technology for providing insights about the study. They also would like to express their appreciation to Mr. John Paul Tomas, their adviser in this study, for providing them guidance throughout the whole time in conducing this research and giving constructive feedback that helped improved their research paper. They also give thanks to their panel, Mr. Ariel Kelly Balan, Mr. Joel De Goma and Ms. Mary Jane Samonte, for also giving constructive criticism to improve the research paper. They are also grateful to Ms. Apryl Mae Parcon for giving time to discuss to the researchers the background and information regarding children behaviors and attention related subject matters. In extension, their most gratitude goes to the children and parents who allowed the children to take part in this research study.

8. REFERENCES

[1] Piontkowski, D., & Calfee, R. (1979). Attention in the Classroom. Attention and Cognitive Development, 297–329. doi:10.1007/978-1-4613-2985-5_11]

[2] Breslau, N., Breslau, J., Peterson, E., Miller, E., Lucia, V. C., Bohnert, K., & Nigg, J. 2010. Change in Teachers' Ratings of Attention Problems and Subsequent Change in Academic

Achievement: A Prospective Analysis. Psychological Medicine, 40(1). 2010. 159-166.

[3] Sigman, M., Cohen, S. E., & Beckwith, L. (1997). Why does infant attention predict adolescent intelligence? Infant Behavior and Development, 20(2), 133-140.

[4] Assiter, K. (2008). Attention and learning in the connected classroom. Journal of Computing Sciences in Colleges, 24(1), 219-226.

[5] Parcon, Apryl Mae. (2018, October 24). Interview.

[6] Stipek, D., & Valentino, R. A. (2015). Early childhood memory and attention as predictors of academic growth trajectories. Journal of Educational Psychology, 107(3), 77

[7] B. Kort, B. Reilly, and R. Picard. 2001. An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy - Building A Learning Companion. In Advanced Learning Technologies, 2001. Proceedings. IEEE International Conference on. IEEE, Madison, USA, pp. 43–46

[8] Zaletelj, J. (2017, September). Estimation of students' attention in the classroom from kinect features. In Image and Signal Processing and Analysis (ISPA), 2017 10th International Symposium on (pp. 220-224). IEEE.

[9] Narayanan, S. A., Prasanth, M., Mohan, P., Kaimal, M. R., & Bijlani, K. (2012, January). Attention analysis in e-learning environment using a simple web camera. In Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on (pp. 1-4). IEEE.

[10] Veliyath, N., De, P., Allen, A. A., Hodges, C. B., & Mitra, A. (2019, April). Modeling Students' Attention in the Classroom using Eyetrackers. In Proceedings of the 2019 ACM Southeast Conference (pp. 2-9). ACM.

[11] Savov, T., Terzieva, V., & Todorova, K. (2018, September). Computer Vision and Internet of Things: Attention System in Educational Context. In Proceedings of the 19th International Conference on Computer Systems and Technologies (pp. 171-177). ACM.

[12] Mliki, H., Fourati, N., Smaoui, S., & Hammami, M. (2013, May). Automatic Facial Expression Recognition System. In 2013 ACS International Conference on Computer Systems and Applications (AICCSA) (pp. 1-4). IEEE.

[13] LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set validity and reliability from untrained adults. Frontiers in psychology, 5, 1532.

[14] OpenFace: an open source facial behaviour analysis toolkit: 2018. https://github.com/TadasBaltrusaitis/OpenFace. Accessed: 2019-02-13.

[15] Rapid Miner. (2019, May 31). Key Features of RapidMiner Studio. Retrieved from https://rapidminer.com/products/studio/feature-list/

[16] Magoosh Data Science Blog. (2017, December 06). What is K-Fold Cross Validation?. Retrieve from https://magoosh.com/data-science/k-fold-cross-validation/

[17] Classification Accuracy is Not Enough: More Performance Measures You Can Use: 2014. https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/. Accessed: 2018-11-06.

[18] De Castro, M. J. C., De Goma, J. C., Devaraj, M., Lopez, J. P. G., & Medina, J. R. E. (2018, September). Distraction Detection through Facial Attributes of Transport Network Vehicle Service Drivers. In Proceedings of the 2018 International Conference on Information Hiding and Image Processing (pp. 112-118). ACM.