DETECTION SYSTEM *ENGAGEMENT* STUDENTS IN AN E-ENVIRONMENT LEARNING WITH CNN BASED OPENCV TECHNOLOGY

Livia Ellen
1606887560
Department of Electrical Engineering
University of Indonesia
Depok, Indonesia
livia.ellen@ui.ac.id

***Abstract*— In this digital era, learning with methods *e-learning* be a general solution implemented in distance education. Disadvantages of the method *e-learning* This is the lack of teacher information regarding student enthusiasm and level of participation in learning. This problem can be solved with a system that is capable of detection *engagement* student. Level *engagement* students on *e-learning*g can be determined from students' views and facial expressions during learning. Detection system *engagement* students work by detecting the direction of students' eyes and facial expressions using OpenCV technology with the CNN method (*convolutional neural networks*) on input *files* in the form of a video or *webcams* in a way *real-time.* The system will provide output in the form of value *engagement* student "*engaged"* based on the duration of the student's eyes staring at the screen and the student's facial expression in the form of a neutral or positive expression. The system will provide output in the form of student attendance scores "*disengaged"* based on the duration of the student's eyes not looking at the screen and the student's facial expression shows a negative expression. The system analyzes students' emotional reactions which are represented in the percentage value parameters of neutral, positive and negative reactions using the FER-2013 dataset. Detection system *engagement* students can measure presence, status *attendance* students pay attention to the screen, emotions, impressions and status *engagement* students with an accuracy rate of 83.33%, precision of 100%, *recalls* amounted to 66.67% and *f1 score* of 80.00%.**

Keywords—*OpenCV, student engagement, Convolutional neural network, e-learning*

I. PINTRODUCTION

In the current era of globalization, learning with methods *e-learning* becoming a common solution for distance education. Many universities and private institutions offer distance education services.

Distance education can be defined as an organized educational process that bridges the separation between students and educators and is mediated by the use of technology and minimal face-to-face meetings. Distance education is offered across space and time so that students gain the flexibility to study in different times and places, and use a variety of learning resources. Distance education is the right solution to solve educational problems that are hampered by geographic location which makes conventional learning impossible.

Based on Law no. 12 of 2012 concerning Higher Education, distance education is a teaching and learning process carried out remotely through the use of various communication media.[1]

Distance education aims to:

a) provide Higher Education services to community groups who cannot participate in face-to-face or regular education; And

b) expanding access and simplifying Higher Education services in education and learning

Disadvantages of distance education methods (*elearning)* This is the lack of teacher information regarding student enthusiasm and level of participation in learning. This can be solved with a system that is capable of detection *engagement* students use *webcams* on students' devices. This system works by scanning students' faces and eye direction using OpenCV technology with the CNN method *(convolutional neural networks).*

Based on *Handelsman et al* (2005), *student engagement* Good performance can be seen from four things, namely: behavior in training one's abilities, positive emotional response during the learning process, active participation in the learning process, and learning performance.[2]

Therefore, the author developed a detection system *engagement* students to be able to measure 2 of the 4 things that indicate a good level of student engagement. Detection system *engagement* students determine *engagement* based on the level of focus of the student's gaze on the screen and the student's facial expression. The system will provide output in the form of values *engagement* student "*engaged"* based on the duration of the student's eyes staring at the screen and the student's facial expression in the form of a neutral or positive expression. The system will provide output in the form of student attendance scores "*disengaged"* based on the duration of the student's eyes not looking at the screen and the student's facial expression shows a negative expression.

The system analyzes students' emotional reactions which are represented in the percentage value parameters of neutral, positive and negative reactions using the FER-2013 dataset. Detection system *engagement* students can measure presence, status *attendance* students pay attention to the screen, emotions, impressions and status *engagement* student.

Engagement students are predictors of good learning as well as predictors of effective teaching (*Handelsman et al,* 2005).[2] *Reeve* (2005) suggests that the higher the level *engagement* a learner, the better the learning process will be.[3] Status *engagement* output results of the detection system program *engagement* these students can provide *feedback* for teachers to find out responses from students. The output from this program can help teachers find out how motivated students are in learning. By understanding status *engagement* from students, teachers can evaluating and modifying the teacher's delivery method in order to maximize the teaching and learning process in the environment *e-learning.*

A. Learning Process

Learning activities relate to learning and teaching activities. Teaching and learning activities allow interaction between students and educators. According to Slameto, educators are one of the human components in the teaching and learning process, which plays a role in efforts to form potential human resources in the field of development. (Slameto., Bina Literacy).[4] The learning objective is so that students can achieve the ability or level of mastery that is expected to be achieved after following a learning process. To achieve these learning objectives, it is necessary to choose appropriate methods in learning.

B. Student Engagement

Engagement students are a display or manifestation of motivation that can be found through actions, namely behavior, emotions and cognition displayed by students in academic activities (Connell and Welborn (1991).[5] Reeve (2005) provides a definition of *student engagement* namely, the intensity of behavior, emotional quality, and personal effort of students' active involvement in learning activities.[3]

Engagement students are predictors of good learners as well as predictors of effective teaching (*Handelsman et al,* 2005).[2] *Reeve* (2005) suggests that the higher the level *engagement* a learner, the better the learning process will be.[3]

C. Emotional Expression

Every human being must have emotions. Emotions are a form of transition, response *neurophysiological* to a stimulus that arouses the components of the coordination system, the response informs us about our relationship to the stimulus, and prepares us to relate to the emotion in some way. (Matsumoto and Juang, 2008: 198).[6]

Ekman and Izard (in Matsumoto, 2008: 137) emphasize the universality of expressions of anger, disgust (*disgusted*), Afraid (*scared),* like *(happiness),* sad *(sadness),* and surprised *(surprise).*[7]

Gunarsa (in Safaria & Saputra, 2009) believes that emotional expression is a form of communication through changes in facial expressions and *gestures* that accompanies emotions, as an outburst of emotions, expressing, conveying feelings to others, and determining how others feel.[8]

Hude (2006) also believes that forms of human emotional expression that appear in reality are generally displayed through: (a) facial expressions, (b) voice expressions, (c) expressions of attitudes and behavior and (d) other expressions.[9]

So, to determine a person's emotional expression, one of the parameters that can be used is facial expression. Facial expressions are emotional expressions that are easy to recognize because they change shape The physical things that are most clearly visible when certain emotions appear include changes in the forehead, eyebrows, eyelids, nose, cheeks, mouth and lips.

D. Digital Image

Digital image processing *(digital image processing)* is a scientific discipline that studies image processing techniques for still images (photos) and moving images (video). The meaning of digital in digital image processing itself means that image processing is done digitally using a computer.[10] One library that is widely used in digital image processing is OpenCV.

E. OpenCV

OpenCV *(Open Source Computer Vision)* is *libraries* of programming functions for *real-time* computer vision [11]. OpenCV uses the BSD license and is free for both academic and commercial use. OpenCV can be used in C, C++, Python, Java programming languages, and so on. OpenCV can be used on Linux, Mac OS, Windows, Android and IOA operating systems. In this research, the author used OpenCV 4.1.2 on Mac OS using the Python programming language.

The features contained in OpenCV include (Gary, 2008):[11]

1. Data manipulation *image* (allocation, release, duplication, settings, conversions),
2. *Image* and video I/O (file and camera based input, image/video file output),
3. Structural analysis (connected components, contour processing, distance transformation, moment variations, *Hough transformation*, estimation *polygonal*, adjust the line, *delay triangulation*),
4. Manipulation of matrices and vectors as well as linear algebra (product, solution, *eigenvalues*, SVD),
5. Camera calibration (finding and tracing calibration patterns, calibration, basis of matrix estimation, homography estimation, stereo correspondence),
6. Various dynamic data structures (list, row, graph),
7. Basic image processing (filters, edge detection, corner detection, sampling and interpolation, color conversion, morphological operations, histograms),

8. *Basic Graphical User Interface* or GUI (display *image* And *videos*, handling *mouse, scroll-bars* And *keyboards*),

9. Labeling *image* (lines, polygons, text images),

10. Movement analysis (*optical flow*, motion segmentation, search),

11. Object recognition (eigen method, HMM).

The modules contained in OpenCV include:

- *CV* –main functions of OpenCV,
- *cvaux*–OpenCV helper functions,
- *cxcore*–supporting data structures and linear algebra,
- *highgui*–GUI function.

Figure 1 OpenCV logo

F. Haar-like Feature

*Haar-like feature*e is a method *feature extraction* And *classification* which was first introduced by Paul Viola and Michael Jones.[12] *Haar-like feature* is *rectangular features*, which can provide specific indications on an image or *image*. *Haar-like feature* used to recognize objects based on the simple value of a feature, not the pixel values contained in the object's image.

Training data *image* on *haha* requires 2 types of object images in the training process, namely:

1. *Positive samples,* contains an image of the object you want to detect. If you want to detect a face, these positive samples contain an image of a face, as well as other objects you want to recognize.
2. *Negative samples*, contains images of objects other than the image you want to recognize, generally in the form of background images (walls, views, floors, etc.). The resolution for negative sample images is recommended to have the same resolution as the resolution of the camera used.

G. Python Programming Language

Python is a high-level programming language that is interpretive, interactive, *object oriented*, and can operate on almost any platform: Mac, Linux, and Windows. Python is a programming language that is easy to learn because of its clear syntax, which can be combined with the use of ready-made modules use, and efficient high-level data structures (Kadir, 2005).[15]

H. Tensorflow

Tensorflow is *libraries* made by Google *open-source*. Previously, tensorflow was found in its image recognition feature *Google Photo's* or *voice recognition* on *Google Now*.[16] *Tensorflow* combines computational algebra with compilation optimization techniques and makes it easier to calculate many mathematical expressions. So far, the time required to carry out calculations has been a problem in carrying out many mathematical computations. As is *tensorflow,* computing can be optimized. On implementation, Tensorflow runs on *backend* system.

Figure 2 Tensorflow logo

I. Convolutional Neural Network (CNN)

Convolutional Neural Network is a machine learning method which is a development of *Multi Layer Perceptron (MLP)* which is designed to process two-dimensional data. CNN is one of a kind *Deep Neural Networks* because of the depth of the network level and is widely implemented in digital image processing. *CNN* has two methods; namely classification using *feedforward* and the learning to use stage *back propagation*. The way CNN works is similar to MLP, but in CNN each neuron is presented in two dimensions, unlike MLP where each neuron is only one dimension in size.[17]

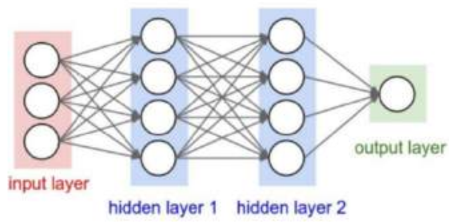


Figure 3 Simple MLP Architecture

Convolutional neural networks is a special type of *neural networks* to process data with a mesh topology or *grid-like topology*. Linear operations on CNN use convolution operations with weights that are not just one dimension, but four dimensions which are a collection of convolution kernels.

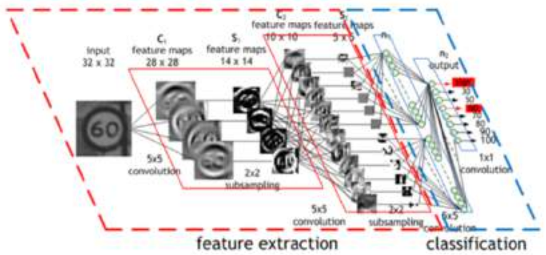


Figure 4 Example of Convolutional Neural Network Architecture

In general, CNN utilizes the convolution process by moving a convolution kernel (filter) of a certain size to an image, the computer obtains new representative information from the results of multiplying parts of the image with the filter used.

J. FER-2013

The Facial Expression Recognition 2013(FER-2013) is a dataset introduced at the 2013 International Conference on Machine Learning (ICML) (*Goodfellow et al.,*2013).[19]FER-2013 contains 35,887 images *grayscale*a 48x48 face consisting of 7 different types of emotions. FER-2013 data has been labeled and classified into 7 classes with an index between 0 and 6 as in Table 1.

Label	Types of Emotions	Amount
0	Angry	4593
1	Disgust (Disgust)	547
2	Scared (Scared)	5121
3	Happy (Happy)	8989
4	Neutral (Neutral)	6077
5	Sad (Sad)	4002
6	Surprised	6198

Table 1 Classification of Emotions in FER-2013

II. PSYSTEM DESIGN

In general, the detection system*engagement* Students will process input video in the form of recordings *video conferencing*, then the video will be processed using *libraries*OpenCV and face detection algorithm. The system will read student attendance levels using two parameters; the length of time the student's eyes stare at the screen and the length of time the student's face faces the screen.

Figure 5 Student Engagement Detection System Workflow

Simultaneously, the system also analyzes reactions students' emotions during learning using the FER-2013 model. Then, the system will process the data and analyze it*value.*The system will display and save the output results in the form of videos that have been annotated and text documents containing values in the form of eye value percentages*on-screen*And*off-screen*, neutral, positive and negative emotional responses and percentage of students' emotional states: neutral, happy, sad, angry, afraid and surprised.

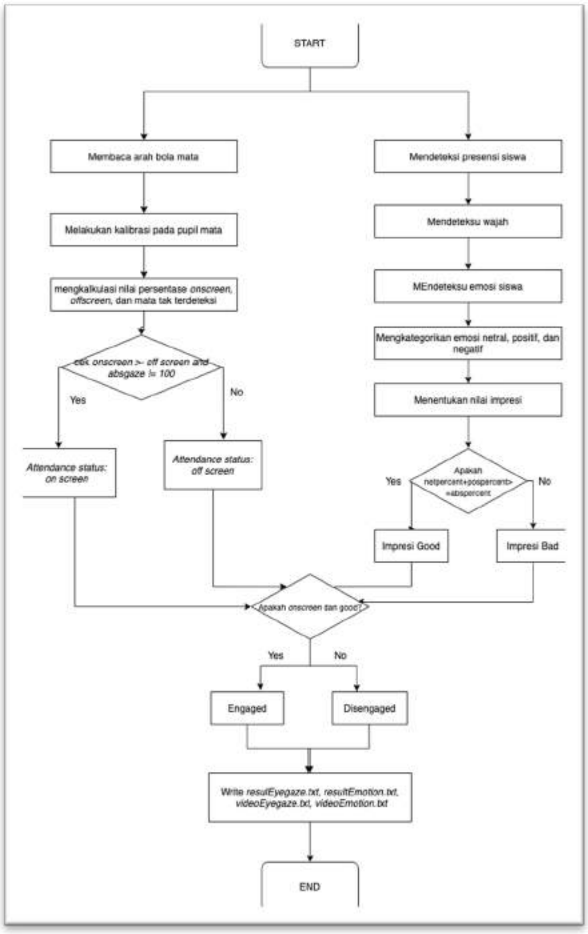


Figure 6 Flowchart of the Student Engagement Detection System

Figure 6 shows*flow chart*detection system *engagement*student. The first process carried out in this system is*terminal*will run the system*Eye Gaze on Screen – Off Screen Detection*and system*Participant Detection and Emotion Recognition*This simultaneously uses GNU Parallel.

Then, the program will run simultaneously The same,*eyegaze.py*And*emotion.py*will read the program's video input. Python programs*eyegaze.py*is a system*Eye Gaze on Screen – Off Screen Detection.*Python programs *Emotion.py*is a system*Participant Detection and Emotion Recognition.*After that the program will analyze and provide a text output file in the form of *resultEyegaze.txt* And*resultEmotion.txt*contains the value of the python analysis results *scripts*as well as video files*output_eyegaze.mp4* And *output_emotion.mp4*contains videos that have been annotated by the detection system*engagement*student.

A. Design of a Student Engagement Detection System Measurement

In the process of validating and testing the detection system *engagement* students, the author did comparative analysis of 7 videos by detection system *engagement*t students compared with analysis of annotation results by 5 respondents.

Output from the Detection System *Engagement*t Students will provide *value* Conditions in 2 different categories: *"Engaged" and "Disengaged"* which is analyzed based on dimensions *behavioral* and dimensions *emotional*, referring to research *Aslan et al* And *The Facial Expression Recognition-2013 dataset* from *International Conference on Machine Learning*.[20] [21]

Classifying condition categories in videos is processing data resulting from annotation in the form of a questionnaire for 5 respondents using dimensions *behavioral* and dimensions *emotional*. Before annotation is carried out, the annotator has read the information regarding the dimension definition *behavioral* And *emotional* dimensions of measurement *engagement.*

The following is an explanation of the dimensions *behavioral* and dimensions *emotional* used in detection systems *engagement* student:

- Dimensions *Behavioral:*
  - *On-Screen:* The student's eyes are fixed on the screen or slightly looking down, looking *keyboards* on the screen.
  - *Off-Screen:* The student's eyes are not looking at the screen, or the eyes are closed, or the eyes are not detected (the student is not present).
- Dimensions *Emotional:*
  - Emotional Reaction
    - Ø Neutral: Student show neutral emotional state
    - Ø Positive: Students show positive emotional condition: in the form of happiness or surprise.
    - Ø Negative: Students show negative emotional conditions: in the form of anger/disappointment, disgust, fear or sadness.
  - Impression
    - Ø *Good:* Emotion Percentage Value Neutral + Positive >= Negative
    - Ø *Bad:* Emotion Percentage Value Neutral + Positive < Negative

The following is an explanation of the classification used by the system as a reference for determining *engagement* based on dimensions *behavioral* and dimensions *emotional* based on the conditions of the following parameters

- *Engaged,* if the following conditions are met:
  - Condition *onscreen - good*
    - Ø Mark *Overall Attendance* student *= On Screen*
    - Ø Impression Value: *Good*
- *Disengaged,* if one of the following is condition met:
  - Condition *onscreen-bad*
    - Ø Mark *Overall Attendance* student *= on screen*
    - Ø Impression Value: *bad*
  - Condition *offscreen-good*
    - Ø Mark *Overall Attendance* student *= off screen*
    - Ø Impression Value: *good*
  - Condition *offscreen-bad*
    - Ø Mark *Overall Attendance* student *= off screen*
    - Ø Impression Value: *bad*
  - Condition *null-good*
    - Ø Mark *Overall Attendance* student *= null (no attendance)*
    - Ø Impression Value: *good*
  - Condition *null-bad*
    - Ø Mark *Overall Attendance* student *= null (no attendance)*
    - Ø Impression Value: *bad*
- *null,* if all the conditions below are met:
  - Condition *null-null*
    - Ø Mark *Overall Attendance* student *= null (no attendance)*
    - Ø Impression Value: *null (no attendance)*

B. Design and Build a Student Engagement Detection System Program

Detection System Program *Engagement* Student consists of two detection systems running simultaneously using GNU functions *parallel*. These detection systems include: *Eye Gaze on Screen – Off Screen Detection* And *Participant Detection and Emotion Recognition*. *Eye Gaze on Screen – Off Screen Detection* can detect dimensions *behavioral engagement* students based on the position of the student's eyeballs towards the screen using *libraries shape_predictor_68_face_landmarks.dat* from *dlib.*

Participant Detection and Emotion Recognition can detect dimensions *emotional engagement* student based on student facial expressions using haarcascade_frontalface_default.xml – OpenCV's default pretrained mode and FER-2013 dataset in *convolutional neural networks*. This system can also perform dimensional analysis *behavioral engagement* on students by detecting the presence of students' faces to determine student participation in learning.

C. Engagement Analysis Parameters by Annotator

To validate the accuracy of the detection system *engagement* students, the author compares with analysis *engagement* annotation results. The annotator will analyze the video and fill in a checklist every 10 seconds of the video according to the parameters in the detection system*engagement*student.

VIDEO 1	Kondisi / 10 detik	Pilih salah satu pada tiap kategori												Persentase (%)
		10	20	30	40	50	60	70	80	90	100	110	120	
Mata	On Screen													0
	Off Screen													0
	Tidak Terdeteksi													0
Ekspresi Wajah	Netral													0
	Positif													0
	Negatif													0
Q1 : Presensi	Apakah siswa hadir? (Ya/ Tidak) Jawaban:													
Q2: Emosi Overall	Kategori respon emosi : netral, positif, negatif. Pilih satu kondisi yang menggambarkan kondisi keseluruhan emosi siswa! Jawaban:													
Q3: Notes	Catatan Tambahan: Jawaban:													

Notes : Apabila Wajah Tidak terdeteksi, maka tidak perlu analisa emosi pada rentang waktu tersebut.

Figure 7 Engagement Annotation Table for Annotators

The output from the annotation table that has been filled in by the annotator will be processed to validate the level of accuracy of the detection system*engagement*student.

To calculate evaluation metrics for detection system performance*engagement*students, the author uses 3 *performance metrics*namely accuracy, precision,*recalls,*and F1- *measure.*In the evaluation of these performance metrics, accuracy counts*true positive*And*true negative*of the test video input samples divided by the entire number of samples *testing. (Formula 1)*

$$\boxed{Accuracy = \frac{TP + TN}{TP + FP + TN + FN}}$$

with TP, TN, FP, and FN respectively being*true positive, true negative, false positive,*And*false negative.*

$$\boxed{Precision = \frac{TP}{TP + FP}}$$

Precision is the ratio of the number of positive predictions from a test sample to the total number of samples that have a positive value.

$$\boxed{Recall = \frac{TP}{TP + FN}}$$

Recall(Sensitivity) is the ratio of the number of positive predictions from the experimental sample to the total number of valuable samples*true positive*And*false negative.F1 scores* calculated using Formula 5 below, which is a calculation of the value*precision* And*recalls.*

$$\boxed{F1 = 2 \times \frac{p \times r}{p + r}}$$

With de$\boxed{p}$ined as as $\boxed{\frac{TP}{TP + FP}}$, And $\boxed{r}$ defined $\boxed{\frac{TP}{TP + FN}}$

III. IMPLEMENTATION AND ANALYSIS

The system requires a computer to carry out its work process. The system runs using Terminal. Before carrying out the installation*libraries*or*dependency*, *use*meed to download some files*requirements.*In order for the program to run, the author needs to install it*homebrew*And*python*on the device*macOS.*

```
requirements.txt
attrs==19.3.0
beautifulsoup4==4.9.1
dlb==0.5.0
dlib==19.20.0
falcon==1.0.0
fer==20.0.0
Flask==1.1.2
gast==0.3.3
h5py==2.10.0
idna==2.8
imageio==2.9.0
imutils==0.5.3
ipython-genutils==0.2.0
Keras==2.3.1
Keras-Applications==1.0.8
Keras-Preprocessing==1.1.0
matplotlib==3.2.2
mtcnn==0.1.0
```

Figure 8 Supporting Dependency Library in requirements.txt

A. Dataset Preparation

On*Eye Gaze on Screen – Off Screen Detection,*The system detects the shape of the face using a model*dlib shape predicto*namely dlib shape_predictor_68_face_landmarks.dat as a reference, then projecting the shape of the eyeball pupil using*code*on*eye.py*And*pupil.py.*

Meanwhile, on*Participant Detection and Emotion Recognition,* system use *haarcascade_frontalface_default.xml–pretrained mode* OpenCV built-in to detect user faces and FER- *2013* deep dataset*convolutional neural networks*to detect emotions*user.*

The FER-2013 dataset has 32,000 images*low resolution*which shows emotion*in the wild*(real world as the original, not a pose), so there may be confusion. However, the FER-2103 dataset is a large dataset that has diversity, so it is very good because it can provide a diverse model*robust.*

To be able to use*datasets*FER-2013, author download*datasets*FER-2013 in*Kaggle*which is available in CSV format. Then, the author converts*FER-2013 dataset csv file*from*Kaggle*into PNG format for *training* And*testing*use*script dataset_prepare.py. FER 2013 dataset*contains 7 emotions such as anger, disgust (*disgusted*), Afraid (*scared),*like*(happiness),*sad*(sadness),* and surprised*(surprise).*

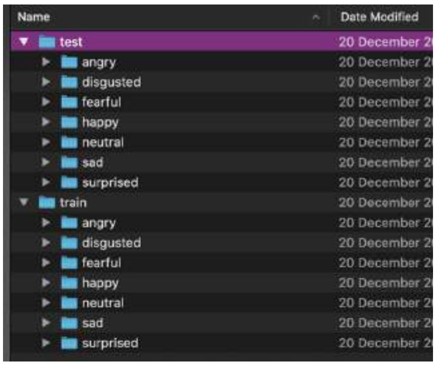


Figure 9 Testing and Training Dataset Folder

After operate *command python dataset_prepare*.py, it will produce output in the form of 2 folders for *testing* And *training* which contains approximately 32,000 images *low resolution* with 7 kinds of human expressions.

Figure 10 Happy Expression Folder for Training Dataset

To be able to-*trains* program, users can also use *command python emotiontrain.py* which will save *weighted model* on *model.h5 file* which can then be used to detect students' emotions.

In this research, the author used *pre-trained model* from *github/atulapra* to detect facial emotions that have been carried out *training* And *test* use *datasets* FER-2013

B. Video Input

The following is an explanation of each video input used:

- input1: Students always look at the screen, students use laptops in a half-lying position, static body position, showing a neutral expression.
- Input2: Students always look at the screen, sitting and sometimes moving, showing interested expressions, laughing and smiling.
- Input3: Students always look at the screen, students use laptops in a prone position, static body position, showing a neutral expression.
- Input4: Students are present, but sometimes they don't look at the screen, their body position is sitting and static, they look a lot, students look confused and frown several times.
- Input 5: Students are present, but almost don't look at the screen at all, more or less only look at the screen for 3 seconds, students focus on playing on their cellphones, students change position: sitting and lying down, neutral expression.
- input6: Students are present, but students are not looking at the screen at all, always looking down, body position sitting, static, neutral expression.
- input7: No students, just room.

Figure 11 Video Input

C. System Testing

Testing carries out tests on a system with 7 video inputs with different conditions for analysis. Testing and data collection were carried out using a Macbook Pro Mid-2017 laptop with the macOS Catallina operating system which has a Python virtual environment.

The author carried out 2 tests in the form of analysis *engagement* 7 videos by detection system *Engagement* t students and analysis *engagement* by 3 annotators. Analysis *engagement* using a detection system *engagement* students are carried out on the author's device with the specifications in the previous section. Meanwhile, analysis *engagement* by 3 annotators carried out by filling out a questionnaire in the form of a parameter checklist that is the same as the detection system *engagement* student. Annotators analyze the video and fill in the parameter checklist *engagement* every 10 seconds of video.

Detection System *Engagement* t Students and analysis *engagement* by the annotator will produce *value* which will be analyzed to determine categories *engagement* the student *"Engaged"* or *"Disengaged".*

Determination of classification *engaged, disengaged,* And *null* based on the conditions of the following parameters:

Condition	***Overall Attendance***	**Impression**	***Engagement***
1	*onscreen*	*good*	*Engaged*
2	*onscreen*	*bad*	*Disengaged*
3	*offscreen*	*good*	*Disengaged*
4	*offscreen*	*bad*	*Disengaged*
5	*null*	*good*	*Disengaged*
6	*null*	*bad*	*Disengaged*
7	*null*	*null*	*null*

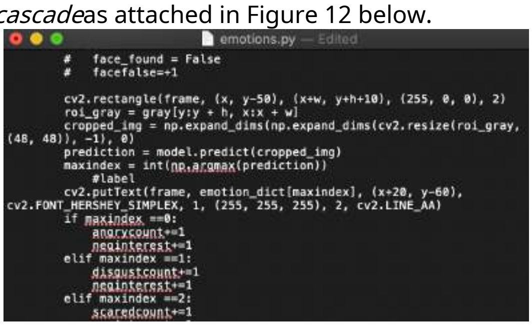
Table 2 Engagement Condition Parameters

D. Convolutional Neural Network

Participant Detection and Emotion Recognition use *convolutional neural networks*as the main architecture. The input video file will be in*capture*every *the frame*using functions*ret, frame = cap.read()* on the program*emotion.py*.

Every*frames*which has been in*capture*will be processed using the method*haar cascade*which is used to detect faces in each*frames*from input*files*videos. The facial area will be affected*crop*and done*resizing*to a size of 48x48 pixels and then processed and used as input for the model*convolutional neural networks.*

Process on*convolutional neural networks*this happens after the face position is detected using*frontal facial haarcascade*as attached in Figure 12 below.



```
emotions.py — Edited
        #    face_found = False
        #    facefalse=+1

        cv2.rectangle(frame, (x, y-50), (x+w, y+h+10), (255, 0, 0), 2)
        roi_gray = gray[y:y + h, x:x + w]
        cropped_img = np.expand_dims(np.expand_dims(cv2.resize(roi_gray,
(48, 48)), -1), 0)
        prediction = model.predict(cropped_img)
        maxindex = int(np.argmax(prediction))
            #label
        cv2.putText(frame, emotion_dict[maxindex], (x+20, y-60),
cv2.FONT_HERSHEY_SIMPLEX, 1, (255, 255, 255), 2, cv2.LINE_AA)
        if maxindex ==0:
            angrycount+=1
            neginterest+=1
        elif maxindex ==1:
            disgustcount+=1
            neginterest+=1
        elif maxindex ==2:
            scaredcount+=1
```

Figure 12 CNN architecture in the emotions.py script

Figure 12 shows*neural network processes*occurs in *command prediction=model.predict(cropped_img).*

Output from*networks*This will give an output in the form of:*lists*containing*softmax score*of 7 emotion classes based on the FER-2013 dataset. The system will store and display emotions with*maximum score*highest. Function *int(np.argmax(prediction))*will provide output in the form of emotions with the highest value.*Convolutional Neural Networks* used in this research was implemented based on research results*Correa et al.*[22]

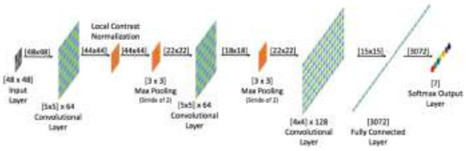


Figure 23 Convolutional Neural Network Architecture

The following is an implementation of*convolutional neural networks*on. The implementation of this emotion testing system can detect the emotions of all faces in the video. However, for testing the detection system*engagement* students, the researcher decided to conduct a test on 1 (one) face in order to measure the students' emotional responses more precisely.

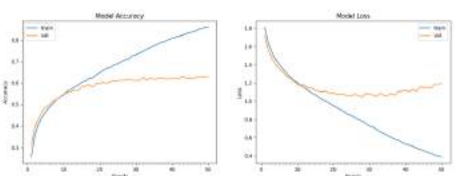


Figure 14 Accuracy Model and Convolutional Loss Model neural networks

Detection system*engagement*students use *simple-4 layer CNN*, with test accuracy reaching a value of 63.2% in 50*epochs.*

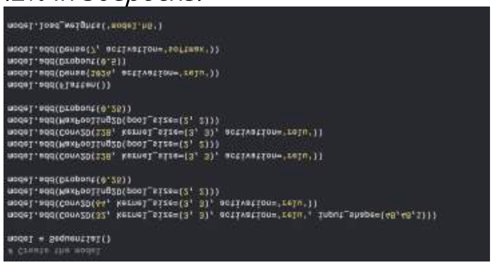


Figure 15 Implementation of Convolutional Neural Networks on emotion.py

*Network*This is divided into 4 layers, where a 48x48 pixel image is used as input, then the next layer is *convolutional layer, local contrast normalization layer,* And*max pooling layer.*

Usage*max pooling layer*applied to reduce the number of parameters in emotion classification.*Max pooling layer*secondly on this architecture can reduce *computational intensity*from*networks*This

After*max pooling layer,*there are 2*convolutional layers*and 1*fully connected layer*which is connected on *softmax output layer*. All layers in this CNN architecture use*ReLu. Fully connected layer*on this architecture uses*dropout*.

E. Test Results

Based on the results*running*on 7 videos using a detection system*engagement*students, the results obtained are:

1. Video Output 1

The following is the output of the detection system test results*engagement*students using video input 1:

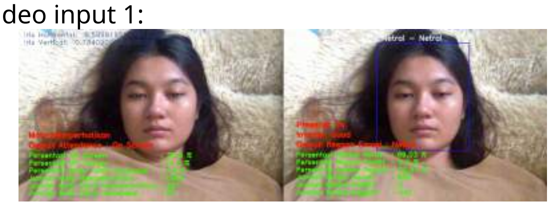


Figure 16 Video Annotation Snippet 1

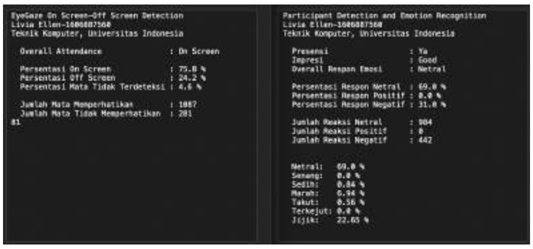
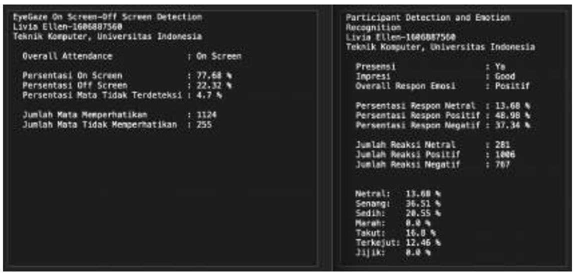


Figure 17 Video Engagement Output 1

2. Video Output 2

The following is the output of the detection system test results *engagement* students with use *input* video 2:

Figure 18 Engagement Video Annotation Snippet 2



```
EyeGaze On Screen-Off Screen Detection
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Overall Attendance                 : On Screen

  Persentasi On Screen               : 77.68 %
  Persentasi Off Screen              : 22.32 %
  Persentasi Mata Tidak Terdeteksi   : 4.7 %

  Jumlah Mata Memperhatikan          : 1124
  Jumlah Mata Tidak Memperhatikan    : 255
```

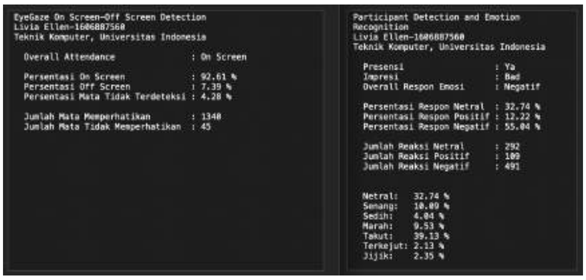
```
Participant Detection and Emotion
Recognition
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Presensi                    : Ya
  Impresi                     : Good
  Overall Respon Emosi        : Positif

  Persentasi Respon Netral    : 13.68 %
  Persentasi Respon Positif   : 48.98 %
  Persentasi Respon Negatif   : 37.34 %

  Jumlah Reaksi Netral        : 281
  Jumlah Reaksi Positif       : 1006
  Jumlah Reaksi Negatif       : 767

  Netral:    13.68 %
  Senang:    36.51 %
  Sedih:     20.55 %
  Marah:     0.0 %
  Takut:     16.0 %
  Terkejut:  12.46 %
  Jijik:     0.0 %
```

Figure 19 Output Engagement Video 2

3. Video Output 3

The following is the output of the detection system test results *engagement* students with use *input* video 3:

Figure 20 Video Annotation Snippet 3



```
EyeGaze On Screen-Off Screen Detection
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Overall Attendance                 : On Screen

  Persentasi On Screen               : 92.61 %
  Persentasi Off Screen              : 7.39 %
  Persentasi Mata Tidak Terdeteksi   : 4.28 %

  Jumlah Mata Memperhatikan          : 1348
  Jumlah Mata Tidak Memperhatikan    : 45
```

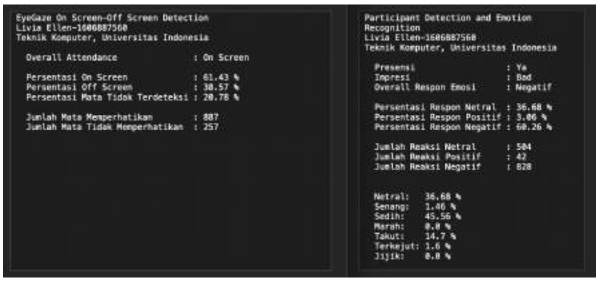
```
Participant Detection and Emotion
Recognition
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Presensi                    : Ya
  Impresi                     : Bad
  Overall Respon Emosi        : Negatif

  Persentasi Respon Netral    : 32.74 %
  Persentasi Respon Positif   : 12.22 %
  Persentasi Respon Negatif   : 55.04 %

  Jumlah Reaksi Netral        : 292
  Jumlah Reaksi Positif       : 109
  Jumlah Reaksi Negatif       : 491

  Netral:    32.74 %
  Senang:    10.09 %
  Sedih:     4.04 %
  Marah:     9.53 %
  Takut:     39.13 %
  Terkejut:  2.13 %
  Jijik:     2.35 %
```

Figure 21 Output Engagement Video 3

4. Video Output 4

The following is the output of the detection system test results *engagement* students using video input 4:

Figure 22 Engagement Video Annotation Snippet 4



```
EyeGaze On Screen-Off Screen Detection
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Overall Attendance                 : On Screen

  Persentasi On Screen               : 61.43 %
  Persentasi Off Screen              : 38.57 %
  Persentasi Mata Tidak Terdeteksi   : 20.78 %

  Jumlah Mata Memperhatikan          : 887
  Jumlah Mata Tidak Memperhatikan    : 257
```

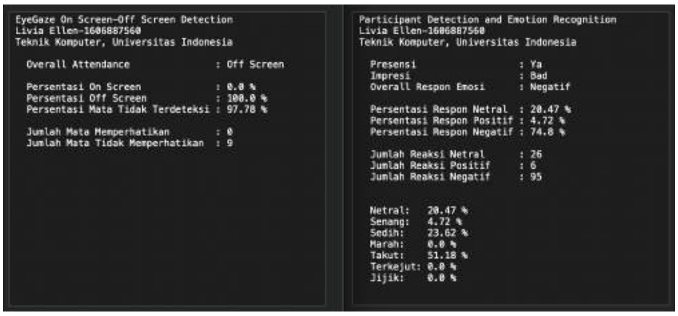
```
Participant Detection and Emotion
Recognition
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Presensi                    : Ya
  Impresi                     : Bad
  Overall Respon Emosi        : Negatif

  Persentasi Respon Netral    : 36.68 %
  Persentasi Respon Positif   : 3.06 %
  Persentasi Respon Negatif   : 60.26 %

  Jumlah Reaksi Netral        : 504
  Jumlah Reaksi Positif       : 42
  Jumlah Reaksi Negatif       : 828

  Netral:    36.68 %
  Senang:    1.46 %
  Sedih:     45.56 %
  Marah:     0.0 %
  Takut:     14.7 %
  Terkejut:  1.6 %
  Jijik:     0.0 %
```

Figure 23 Output Engagement Video 4

5. Video Output 5

The following is the output of the detection system test results *engagement* students using video input 5

Figure 24 Engagement Video Annotation Snippet 5



```
EyeGaze On Screen-Off Screen Detection
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Overall Attendance                 : Off Screen

  Persentasi On Screen               : 0.0 %
  Persentasi Off Screen              : 100.0 %
  Persentasi Mata Tidak Terdeteksi   : 97.78 %

  Jumlah Mata Memperhatikan          : 0
  Jumlah Mata Tidak Memperhatikan    : 9
```

```
Participant Detection and Emotion Recognition
Livia Ellen-1606887560
Teknik Komputer, Universitas Indonesia

  Presensi                    : Ya
  Impresi                     : Bad
  Overall Respon Emosi        : Negatif

  Persentasi Respon Netral    : 20.47 %
  Persentasi Respon Positif   : 4.72 %
  Persentasi Respon Negatif   : 74.8 %

  Jumlah Reaksi Netral        : 26
  Jumlah Reaksi Positif       : 6
  Jumlah Reaksi Negatif       : 95

  Netral:    20.47 %
  Senang:    4.72 %
  Sedih:     23.62 %
  Marah:     0.0 %
  Takut:     51.18 %
  Terkejut:  0.0 %
  Jijik:     0.0 %
```

Figure 25 Video Engagement Output 5

6. Video Output 6

The following is the output of the detection system test results *engagement* students by using *input* video 6:

Figure 26 Engagement Video Annotation Snippet 6

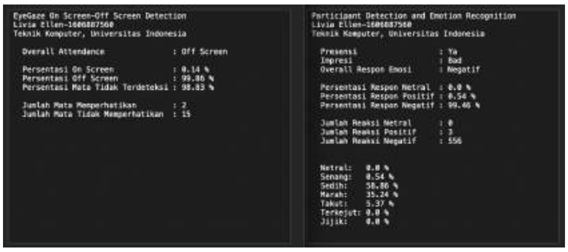


Figure 27 Video Engagement Output 6

7. Video Output 7

The following are:*output*detection system test results*engagement*students with use*input*video 7.

Figure 28 Video Engagement Annotation Snippet 7

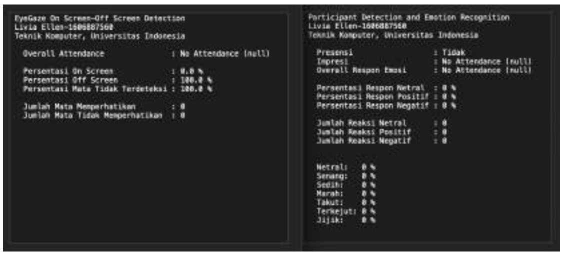


Figure 29 Video Engagement Output 7

Table 3 shows the detection system*engagement*students provide final condition output*null-null*. Condition*nullnull* means video 3 is included in the category*null*, because no one was detected.

Based on the results*running*on 7 videos using detection system*engagement*students, here are the results, namely:

Inputs	On Screen (%)	Off Screen (%)	Tak's eyes Detected (%)	Response Neutral (%)	Response Positive (%)	Response Negative (%)
1	75.8	24.2	4.60	69.00	0.00	31.00
2	77.68	22.32	4.70	13.68	48.98	37.34
3	92.61	7.39	4.28	32.74	12.22	55.04
4	61.43	38.57	20.78	36.68	3.06	60.26
5	0.00	100.0	97.7	20.47	4.72	74.8
6	0.14	99.86	98.83	0.00	0.54	99.46
7	0.00	0.00	100	0.00	0.00	0.00
MIN	**0.00**	**0.00**	**4.60**	**0.00**	**0.00**	**0.00**
MAX	**92.61**	**100**	**100**	**69.00**	**48.98**	**99.46**
AVG (%)	**43.95**	**41.76**	**42.27**	**24.65**	**9.93**	**51.12**

Table 3 Engagement Output Data by Detection System Student Engagement

Input	Presensi		Overall Respon	Attendance	Impresi	Status Engagement	
1	Ya	1	Netral	*onscreen*	*Good*	***Engaged***	**1**
2	Ya	1	Positif	*onscreen*	*Good*	***Engaged***	**1**
3	Ya	1	Netral	*onscreen*	*Good*	***Engaged***	**1**
4	Ya	1	Negatif	*onscreen*	*Bad*	***Disengaged***	**0**
5	Ya	1	Netral	*offscreen*	*Good*	***Disengaged***	**0**
6	Ya	1	Netral	*null*	*Good*	***Disengaged***	**0**
7	Tidak	0	*null*	*null*	*null*	***null***	**-**

Table 4 Analysis of Engagement by Detection System Student Engagement

Based on the average annotation results from 3 annotators for validation of the detection system*engagement*students, the output appears in the form of

Inputs	On Screen (%)	Off Screen (%)	Tak's eyes Detected (%)	Response Neutral (%)	Response Positive (%)	Response Negative (%)
1	80.56	19.44	5.56	75.75	6.06	21.21
2	86.11	13.89	0.00	36.11	50.00	11,11
3	94.44	5.56	0.00	86.11	0.00	13.89
4	50.00	44.44	5.56	17,17	3.03	76.76
5	2.78	97.22	88.89	100	0.00	0.00
6	0.00	100	100	0.00	0.00	0.00
7	0.00	100	100	0.00	0.00	0.00
MIN	**0.00**	**5.56**	**0.00**	**0.00**	**0.00**	**0.00**
MAX	**94.44**	**100**	**100**	**86.11**	**50.00**	**76.76**
AVG (%)	**44.84**	**54.28**	**43.25**	**40.25**	**8.44**	**17.56**

Table 5 Data on Average Engagement Output by Annotator

Input	Presensi		Overall Respon	Attendance	Impresi	Status Engagement	
1	Ya	1	Netral	*onscreen*	*Good*	***Engaged***	**1**
2	Ya	1	Positif	*onscreen*	*Good*	***Engaged***	**1**
3	Ya	1	Netral	*onscreen*	*Good*	***Engaged***	**1**
4	Ya	1	Negatif	*onscreen*	*Bad*	***Disengaged***	**0**
5	Ya	1	Netral	*offscreen*	*Good*	***Disengaged***	**0**
6	Ya	1	Netral	*null*	*Good*	***Disengaged***	**0**
7	Tidak	0	*null*	*null*	*null*	***null***	**-**

Table 2 Analysis of Engagement by Annotator

F. Test Results

In this final assignment testing there are several tests to be implemented directly from the system. This test will be carried out by applying the features contained in this system. In this research, accuracy, precision,*recalls,*And*f1 score*based on parameters, presence, *attendance,*impression, and status *engagement.*

This analysis and testing will be carried out according to the scenario parameters above the detection system results *engagement* students are compared with the results of the analysis *annotator*, the results of this test will be displayed in table form and the results of the formula calculation. In general, this sub-chapter will be divided into 6 parts, namely testing performance metrics for accuracy, precision,*recalls,*And*f1 score*based on presence parameters,*attendance,*emotions, impressions, and status*engagement,*as well as evaluation of detection system performance metrics*engagement*student.

Evaluasi Metrik Performa Sistem Pendeteksi *Engagement* Siswa				
Parameter	**Akurasi**	**Presisi**	***Recall***	***F1 score***
Presensi	100	100	100	100
Status *Attendance*	100	100	80	88,89
Emosi	66,67	75	75	60
Impresi	50	100	40	57,14
***Engagement* status**	**83,33**	**100**	**66,67**	**80,00**

Table 7 Evaluation of Detection System Performance Metrics Student Engagement

Table 7 shows the values of accuracy, precision, *recalls,* And*f1 score*of each parameter. The table shows the values of accuracy, precision,*recalls,*And*f1 score* more than 50% on presence, status parameters *attendance*, emotions and status*engagement,*which means the presence, status parameter model*attendance*, emotions and status *engagement*on the detection system *engagement*students have accurate models. Meanwhile, the impression parameters have accuracy values and*f1 score*the lowest is respectively 50% accuracy and*recalls* by 40%.

Things that need to be optimized further are the precision reading of face detection, students' emotional responses through facial expressions, and the precision of reading the direction of the eyeballs in the program. The program error in detecting faces is shown in Figure 36 below:

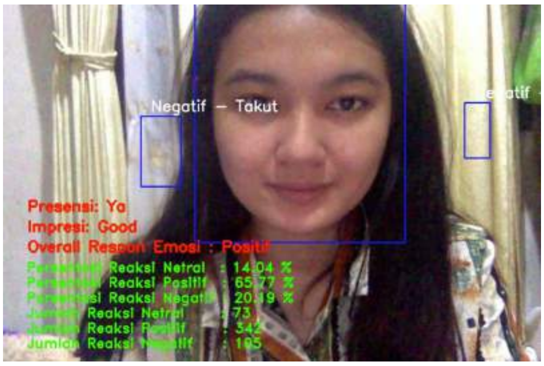


Figure 30 Face Detection Error in Video 2

In figure 30, you can see the program reading another face when*background*own*patterns*which resembles the shape of a face.

Figure 31 Emotion Reading Errors in Video 3, (a) actual emotions, (b) misreading

Figure 32 Misreading of Emotions in Video 6, (a) sad, (b) afraid

Figures 31 and 32 display errors in reading emotional responses. Figure 31.a displays the correct emotion reading, while image 31.b displays the incorrect emotion reading. The program reads students' emotions as negative when students show neutral emotional responses.

Figures 31a and 32.b show errors in reading students' emotions when students look down, *frames*above is categorized as sad and afraid because it resembles the emotions of sadness and fear*datasets* FER-2013 which the author uses in this program, is attached in figure 33.

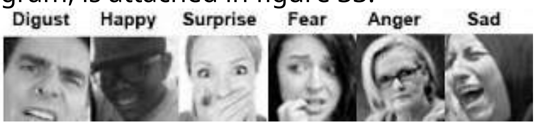


Figure 33 Emotion Classification FER-2013 dataset The lack of precision in reading emotions is also caused by the position of the student's face towards the camera and screen*video conferencing*. When students are sitting, that means*frontal face* students are detected clearly, the number of errors in reading students' emotions is less than when students are asleep, which is *frontal face*appears to be on a certain slope that influences the reading of emotions.

In addition, because the system detects students' emotional responses as long as faces are detected, the emotional responses detected are not just emotional responses to learning materials or learning materials. *elearning*, but rather all emotional responses during the student's face are detected. This includes students' emotions when students interact with other people and other activities: opening their cell phones, interacting with other people during videos*conferencing*and caught on camera.

In this test, the author used the FER-2013 dataset from*International Conference on Machine Learning (ICML) 2013*with a reading accuracy rate of 63.2% in 50*epochs*to detect students' emotions. From the test results, emotional accuracy was 66.67%. For further development, detection system*engagement* students can*trains*dataset again or using a model *machine learning*which is more accurate.

Accuracy in reading emotions affects the reading results of the detection system *engagement* students directly, because the emotional response value influences the impression value, where the impression value is one of the variables to determine status *engagement*

Figure 34 Reading Eye Gaze on Screen-Off Screen Video Detection 4,
(a) Eyes Not Noting, (b) Eyes Not Detected

Figure 34 displays the readings *eye Gaze on screen-off screen detection* in video 4. Figure 40.a shows the condition where the eyes are detected but not paying attention, which is indeed the correct condition: the student is reading a message on a cellphone. The “eyes are not paying attention” condition is valuable *true* then it will increase the percentage value *off screen*. Apart from the “eyes not paying attention” condition, the “eyes not detecting” condition also increases the percentage value *off screen*.Figure 34.b shows the condition when the student closes his eyes because he is starting to feel sleepy, then the condition is on *frames* this also increases the percentage value *off screen.*This means that if students often blink or close their eyes, it will affect the reading of their grades *off screen.*The reason the blinking condition is not disallowed is because the frequency of blinking is related to the student's level of focus, based on research *Nakano et al.*[23]

Based on studies published in journals *Proceedings of the National Academies of Science* this, when the subject blinks, part of the brain *visual cortex* And *somatosensory cortex* those involved in processing visual stimuli and regulating attention are in the down or off position. This indicates that people who blink too often indicate that their minds are not focused. The research also discusses that under 10% of human consciousness is spent blinking. So the reading value of "eyes not detected" in videos 1,2, and 3, which is the condition where students focus on paying attention to the screen without distraction, is worth 4.60%, 4.70%, and 4.28% respectively, which is still within the threshold. normal human blink frequency.

From testing the detection system *engagement* students, it can be concluded that the detection system *engagement* students are effective enough to measure the level *engagement* students on learning methods *e-learning*. Detection system *engagement* students can measure presence, status *attendance* student pay attention to the screen, emotions, impressions and status *engagement* student.

IV. KCONCLUSION

With the results of the experiments that have been carried out, it can be concluded that:

- Detection system *engagement* students can measure presence, status *attendance* students pay attention to the screen, emotions, impressions and status *engagement* students, with an accuracy rate of 83.33%, precision of 100%, *recalls* amounted to 66.67% and *f1 score* of 80.00%
- Performance Value in reading each detection system parameter *engagement* students, namely presence, status *attendance* students pay attention to the screen, emotions, impressions and status *engagement* students are as follows:

Evaluasi Metrik Performa Sistem Pendeteksi *Engagement* Siswa				
Parameter	**Akurasi**	**Presisi**	***Recall***	***F1 score***
Presensi	100	100	100	100
Status *Attendance*	100	100	80	88,89
Emosi	66,67	75	75	60
Impresi	50	100	40	57,14
***Engagement* status**	**83,33**	**100**	**66,67**	**80,00**

- Emotion detection with datasets *FER*-2013 and model *convolutional neural network-4* layer has a reading accuracy rate of 63.2% in 50 *epochs*. In the implementation using 7 video samples, the accuracy level was 66.67%.
- Face detection errors occur due to conditions *background* videos that have a lot *patterns,* so that *patterns* that resembles a face will be detected by the system.
- The position of the student's face towards the screen and the position and tilt of the student's body towards the screen influence the reading of the direction of the student's eyeballs and the reading of the student's emotional response.
- The test objectives have been achieved in the analysis *engagement* students in various conditions.

DAFTERPLITERATURE

[1] Law Number 12 of 2012 concerning Higher Education

[2] Handelsman, Mitchell & Briggs, William & Sullivan, Nora & Towler, Annette. (2010).*A Measure of College Student Course Engagement.* The Journal of Educational Research. 98. 184- 192. 10.3200/ JOER.98.3.184-192.

[3] Isen, A. M., & Reeve, J. (2005).*The Influence of Positive Affect on Intrinsic and Extrinsic*

*Motivation: Facilitating Enjoyment of Play Responsible Work Behavior, and Self-Control. Motivation and Emotion,*29, 295-323. http://dx.doi.org/10.1007/s11031-006-9019-8

[4] Slameto. (1988).*Learning and the factors that influence it*. Building Literacy.

[5] Connell, J. P., & Wellborn, J. G. (1991). *Competence, autonomy, and relatedness: A motivational analysis of self-system processes.*In MR Gunnar & LA Sroufe (Eds.), *The Minnesota symposia on child psychology, Vol. 23. Self processes and development*(p. 43–77). Lawrence Erlbaum Associates, Inc.

[6] Matsumoto & Linda Juang. (2008). Culture & Psychology. USA: Thomson Wadsworth

[7] Matsumoto. (2008).*Introduction to Cross-Cultural Psychology*. Yogyakarta: Student Library

[8] Safaria, T. and Saputra, NE (2009).*Emotion management: A smart guide on how to manage positive emotions in your life*. Jakarta: Bumi Literacy.

[9] Dervish Hude. (2006). *Emotion-Exploration Religio-Psychological about Human Emotions in the Qur'an*. Jakarta: Erlangga

[10] Sutoyo. Q, Mulyanto. Edy, Suhartono. Vincent, Dwi Nurhayati Oky, Wijanarto,*"Digital Image Processing Theory"*, Andi Yogyakarta and UDINUS Semarang, 2009.

[11] Kaehler, Adrian and Bradski , Gary .2008. *Learning OpenCV*. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472

[12] Viola, Paul & Jones, Michael. (2001).*Rapid Object Detection using a Boosted Cascade of Simple Features.*IEEE Conf Comput Vis Pattern Recognit. 1. I-511. 10.1109/CVPR.2001.990517.

[13] WS Pambudi and BMN Simorangkir, " *"Facetracker uses the Haar Like Feature and PID methods in the simulation model."*J. Technology and Informatics, vol. 2, no. 2, 2012.

[14] Papageorgiou, CP & Oren, Michael & Poggio, Tomaso. (1998).*General framework for object detection.*Proceedings of the IEEE International Conference on Computer Vision. 6:. 555 - 562. 10.1109/ICCV.1998.710772.

[15] Abdul Kadir, (2005).*Python Programming Basics*. Yoyakarta : ANDI

[16] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. & others (2016*). TensorFlow: A System for Large-Scale Machine Learning..OSDI*(p./pp. 265--283), .

[17] S. Albawi, TA Mohammed and S. Al-Zawi, *"Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET),*Antalya, 2017, pp.1-6, doi: 10.1109/ICEngTechnol.2017.8308186.

[18] van der Walt, Stéfan & Colbert, S. & Varoquaux, Gael. (2011).*The NumPy Array: A Structure for Efficient Numerical Computation.*Computing in Science & Engineering. 13. 22 - 30. 10.1109/ MCSE.2011.37.

[19] Goodfellow, Ian & Erhan, Dumitru & Carrier, Pierre & Courville, Aaron & Mirza, Mehdi & Hamner, Ben & Cukierski, Will & Tang, Yichuan & Thaler, David & Lee, Dong-Hyun & Zhou, Yingbo & Ramaiah , Chetan & Feng, Fangxiang & Li, Ruifan & Wang, Xiaojie & Athanasakis, Dimitris & Shawe-Taylor, John & Milakov, Maxim & Park, John & Bengio, Y.. (2013). Challenges in Representation Learning:*A Report on Three Machine Learning Contests. Neural Networks.*64. 10.1016/ j.neunet.2014.09.005.

[20] Aslan, S., Alyuz, N., Okur, E.*et al.*Effect of emotion-aware interventions on students' behavioral and emotional states.*Education Tech Research Dev*66,1399–1413 (2018). https://doi.org/10.1007/s11423-018-9589-7

[21] Nezami, O.M., Dras, M., Hamey, L., Richards, D., Wan, S., & Paris, C. (2019). Automatic Recognition of Student Engagement Using Deep Learning and Facial Expression.*ECML/PKDD*.

[22] V. Enrique Correa, Arnoud Jonker, Michael Ozo, Rob Stolk. (2016)*"Emotion Recognition using Deep Convolutional Neural Networks"*

[23] Nakano, Tamami & Kato, Makoto & Morito, Yusuke & Itoi, Seishi & Kitazawa, Shigeru. (2012).*Blink-related momentary activation of the default mode network while viewing videos.* Proceedings of the National Academy of Sciences of the United States of America. 110.

10.1073/pnas.1214804110.

[24] Whitehill, Jacob & Serpell, Zewe & Lin, Yi-Ching & Foster, Aysha & Movellan, Javier. (2014). *The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions.* Affective Computing, IEEE Transactions on. 5. 86-98. 10.1109/TAFFC.2014.2316163.