Journal of Engineering Science and Technology
Vol. 15, No. 6 (2020) 4032 - 4047

CHEATING ACTIVITY DETECTION ON SECURE ONLINE MOBILE EXAM

ARIEF SETYANTO*, BAYU SETIAJI, MARDHIYA HAYATY, KRISNAWATI

Faculty of Computer Science, Universitas Amikom Yogyakarta
Ring Road Utara, Depok-Sleman, Yogyakarta, Indonesia
*Corresponding Author: arief_s@amikom.ac.id

Abstract

The online exam is one of the important tasks in online learning systems. Online exam proctoring, therefore, is important to ensure the credibility of the exam. The online exam used to ascertain students' knowledge on a given topic, irrespective of their locations. This research proposes an online exam system developed using an android application, with a standard camera and audio recorder installed to capture human activity during the test. Audio-video data were obtained from a total of 20 students, which recorded seven cheating and seven non-cheating activities. The method used in this research, is a system used captured students' faces using the front camera, while the audio recorded the sound. A mid-level representation of the audio-video was conducted before the classification task with data normalization performed into uniform units for each parameter between zeros to ten. Finally, the multilayer perceptron (MLP) carried out the classification of a midlevel signal into cheating and non-cheating activity. The result showed a classification accuracy, precision, recalls, and F1 score of 91.73%, 91.73%, 92.9 % and 91.68%, respectively.

Keywords: Action detection, Eye detection, Face detection, Mobile exam, Multilayer perceptron.

1. Introduction

Electronic learning is currently one of the game-changing applications in the education sector due to its flexibility and convenience. The commonly used traditional way of acquiring knowledge, in classrooms has gradually changed into synchronous/ asynchronous online learning system. This new technique has been enabled by online video meeting applications, facilitated by power point presentations, animation, and the use of more interactive programming tools. According to a survey conducted by [1], 14.3% of American students take online courses without a traditional degree, while 15 % take a combination of both. However, there was a significant improvement in the 2013 survey [2], with an increase from 411,000 to 7.1 million students, while the online education provider improved from 2.6% to 5% over the past years. Although the survey was carried out in the USA, it shows the general trend all over the world.

Examination plays an important role in ensuring students' have understood the learning material. It is either synchronous or asynchronous. Synchronous examination is defined as the process whereby the exam occurs on set schedules and timeframes through online channels. Conversely, the asynchronous online exam is carried out by an interactive questionnaire filled by the participants. It permits the occurrence of the process at any time and place. However, online real time proctoring is not possible in this mode. Therefore, an automatic online synchronous examination system is needed to ensure the test is properly conducted.

Several studies have reported that online exam has a higher possibility of malpractice [3-5]. According to King et al. [4] and King and Case [5], 74% of students stated that it is easy to carry out exam malpractice in online examination, while 29% stated that they participated in the act. According to Corrigan-Gibbs et al. [6], the provision of an honor code tends to provide an insignificant effect of reduction in online malpractice, while warning aids in its reduction.

Therefore, online proctoring is important due to the growth of massive open online courses (MOOC). Some of the online/offline proctoring products available in the market are Kryterion, Web assessor and ProctorU [7]. These applications usually put a surveillance camera on the user site, however, they still rely on humans watching over through the video streaming. Kaiiali et al. [8] proposed a mobile exam system to be embedded in the Moodle system to identify various vulnerabilities capable of violating exam security in m-learning environments and to design the appropriate services and counter measures to ensure exam security.

Bawarith et al. [9] utilised a front face camera to track the participant eye movement. The exploitation of eye movement leads is associated with users' focus, and when it leaves the testing sheet, it is assumed that malpractice has occurred. The multimedia analysis helps online proctoring automation with the Gaze tracker, webcam and EEG installed to assist in the process [10]. Atoum et al. [11] installed a webcam and camera on specially designed eyeglasses in order to record the user's face while capturing the participant's view. However, although the design functions properly, a specially installed camera is needed to reduce the practicality.

This, therefore, led to the development of the audio visual tool for activity/event recognition, social interaction, object tracking and sports analysis. This tool was also designed for social cohesion among group members extracted from video and classified by support vector machine [12]. It is specially designed to identify the good behaviour of job seekers with interview videos as proposed by Nguyen et al. [13]. The process of identifying unwanted behaviour in audio visual data has also been conducted by Zajdel et al. [14] and Lefter et al. [15]. In addition, sport activities such as attack, defence, and other numerous reports associated with matches has been proposed by Wang et al. [16], while goals, saves and kick-of activities are recognized using Bayesian network [17].

Various research has also been conducted to detect cheating behaviour on online exams, however, human supervision still plays a vital role in this process, to achieve detailed surveillance. Moreover, in some research such as in [10, 11], special devices such as additional camera, electroencephalogram (EEG) and gaze tracker need to be installed. This research aims to simplify the online exam proctoring task with standard devices only for easier implementation. This research investigates minimum set of data acquired by the standard camera and audio recorder of low-end smartphone to recognize pre-defined cheating activities using artificial neural network (ANN). The developed online test system was designed on a smart phone with a front camera, and voice recorder. These specifications are standard available currently available on most smart phones to provide multimodal data captured during the exam. The camera and the voice recorder are used for test surveillance.

2. Related Work

There are three attributes associated with online proctored exam model. Firstly, it is conducted without human controls [18-20]. Secondly, humans are involved either online through a webcam or physically in a certain location [21]. Thirdly participant are allowed to take the test without supervision, with the exam process recorded by the multimedia analysis system [11, 22].

Online examinations need to ascertain user identity, and there are many efforts used to ensure computer user identity such as keystroke behaviour or biometrics. Kumar and Rathi [23] used keystroke dynamics for online exam authentication, while in [24], iris recognition was used for biometrics authentication. Furthermore, face recognition was periodically used during the exam to ensure the authentication of the participant [25].

Besides participant identity, user action during an exam is an important clue for identifying cheating behaviour. There are many researches dealing with action recognition such as [26-28]. Head movement is one important clue for activity recognition [26]. Detection of human activity is conducted by analysing the socio-communicative and affective behavioural characteristics of interacting partners using their head motion. There are many techniques for head pose estimation as reported in [27-30]. Liu et al. [27] used the multilevel structured hybrid forest to detect the head and estimate the pose. Dornaika and Raducanu [28] proposed a 3D face pose estimation using 2D Eigen faces and DE algorithm. Qing et al. [29] reported the use of Mutual information (MI) exploited to deal with pose estimation in an uncontrollable environment. Another technique proposed by Drouard et al. [30] utilized a mixture of linear regression with partially-latent output.

Eye-tracking plays an important role in understanding human attention. It is used to point out a user's point of view in accordance with its direction. Currently, most eye movement technology are developed using near-infrared light, which lights up the human pupil. This technology is however expensive and lacks practicability due to the need for a special device. In many medical purposes, this technology is viable due to the accuracy requirement and facility availability.

However, this research was not designed to provide a special camera with near-infrared sources, therefore, the simple technique of the web camera was used to detect and track the pupil. Cheung and Peng [31] and Krafka et al. [32] proposed the installation of a webcam on the desktop to recognize the pupil, eye corner and movements. Krafka et al. [32] proposed a normal camera for eye movement recognition. Subsequently, Meng and Zhao proposed CNN based eye movement detection which was tested in a half-million frames and successfully recognized activities such as reading, browsing and watching video [33].

Sound has been identified as a strong clue for activity recognition [11] and cheating behaviour detection in the presence of a human voice. Therefore, speech extraction is performed in modelling the cheating action and has attracted as the attention of [34, 35].

This research made the following contribution:

- It developed a mobile exam video dataset consisting of cheating and non-cheating activities from twenty subjects, 280 actions and a total of 36.120 frames.
- It defined a midlevel data modeling for audio-video cheating dataset which consisted of head tracking, eye tracking and audio.
- Developed an artificial neural network to perform cheating activity detection based on the mid-level dataset.

The proposed method of the paper is explained in section 3, while section 4 described the results and discussed the achievement with conclusion drawn in the last section.

3. Multimodal Action Detection

3.1. Research framework

The research is divided into four main tasks which are, defining the cheating actions, developing a dataset for pre-defined actions, mid-level feature extraction, and action recognition task. Figure 1 depicts the entire research framework.

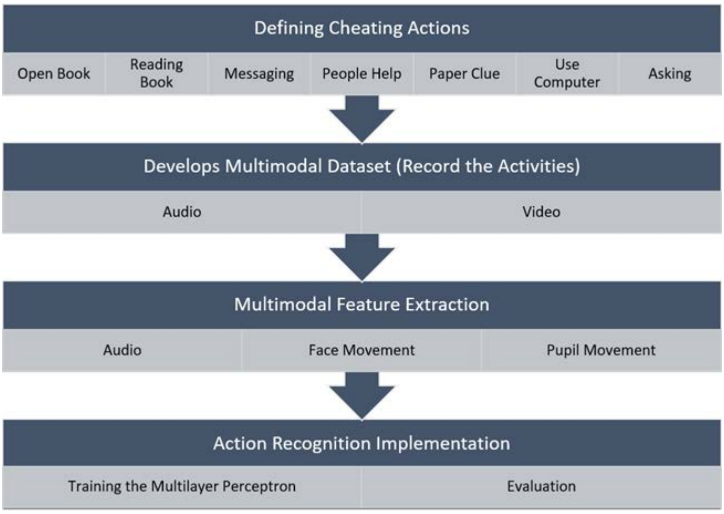


Fig. 1. Research framework.

Online exam cheating activities is defined into seven types. They consist of opening/reading a book, sending/receiving messages through a smart phone, accepting help from people, receiving clues through paper, using the computer to browse testing material, and asking for answers from someone. In order to build a dataset, these activities were recorded in audio and video through standard smart phones. An additional video of non-cheating participants was also recorded with a standardized recording duration of 21.5 seconds in 12 frame per second (fps). A feature extraction was carried out of all the positive and negative cheating videos. The visual and motion data were extracted into face and pupil movement, while the audio was sampled into 12 data per second. The three modalities were fused before entering the network in artificial neural network (ANN), thereby, leading to a total of 1290 data length for each recorded activity. Multi perceptron (MLP), training and evaluation were carried out by using the available data.

3.2. Multimodal input

Multimodal cheating behaviour in online exams involves audio and video data, which is pre-processed from the front camera of the smart phone. Before further processing, a standardization of the visual signal was carried out to reduce the volume of the data while maintaining the quality for feature extraction purposes. The sizes of the frames were under-sampled into 180 x 320 pixels and a reduction of the frame rate was made into 12 per second. The data was considered enough because the video is in slow motion and the distance between the object under surveillance was close. The audio was continually recorded during the data collection for further processing. Figure 2 shows the normal position of a smart phone, with a front camera and the participant's face.

Fig. 2. Normal position of a mobile test examinee.

The video data was extracted into two time series features namely face and pupil movements. Two modalities were obtained in the data collection tasks. The video was explored into the movement using motion vector calculation and Lucas-Kanade methods. Before the tracking started, the face and eye recognition were carried out. This produced a three time series data which are face tracking, eye tracking and audio data. In order to enable the modality fusion for all the data, normalization was carried out in a range of -10 to 10. The periodicity of the data sampled was in 12 data point per second. During the data collection, the original duration was not precisely uniform, while pre-processing and was standardized into 21.5 seconds for each captured audio video. The face tracking produced two variables x and y, which was considered as 1 unit for every 25 pixels movement. Furthermore, the pupil tracking yielded two x and y variables in pixel. Sound on the other hand produced one variable and was sampled in exactly the same frequency at 12 data per second. During the feature extraction, a zero padding is implemented to ensure standard length of each feature for each audio video data. Figure 3 shows the multimodal data pre-processing before a multi-layer perceptron was applied.

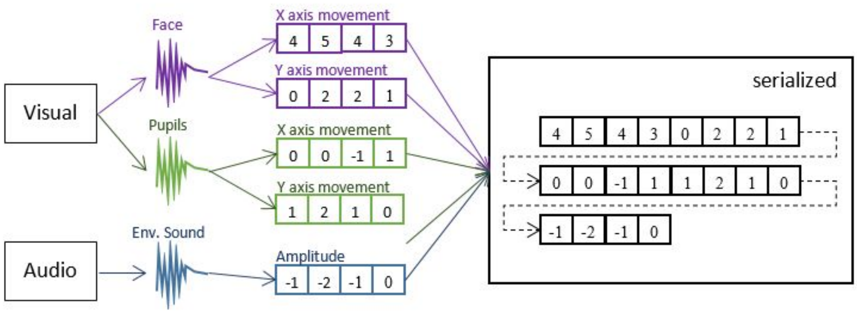


Fig. 3. Multi modal feature extraction and fusion.

3.2.1. Face and eye recognition

The region of interest in the data transformation to the mid-level data representation is the face. Therefore, a number of tasks were needed to analyse the face and the eyes. The first task was to resize the frames into 180 x 320 pixels per frame with 12 frames per second (fps). Secondly, the face location was identified, and its landmark calculated in order to identify its key points such as the nose tip and the eyes.

The face was extracted from the whole image in each frame, using the histogram of gradients HOG [36]. The image was divided into a 16x16 patch and the orientation of the gradient was calculated in 9 bins of unsigned orientation between 0 to 360 degrees [36]. The vector was then normalized to acquire L2-Norm, L1-Norm and L1-Sqrt in Eqs. (1), (2), and (3).

$$L2 - Norm : f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (1)$$

$$1 - Norm : f = \frac{v}{\|v\|_1 + e} \quad (2)$$

$$L1 - Sqrt : f = \sqrt{\frac{v}{(\|v\|_1 + e)}} \quad (3)$$

where v is the non-normalized vector containing all histograms in a given $\|v\|_k$block, $\|v\|_k$ is the k-norm (k = 1,2,..) and e is a constant with a small value.

According to Dalal and Triggs [36], the histogram orientation with L2-Norm and L1-Sqrt provided a good performance. The result of the procedure was a localized face area.

After the face location was determined, its landmark was identified and implemented as proposed by Kazemi and Sullivan [37]. The face orientation in each frame was detected and the relative position of its front direction recorded.

The basic idea was to determine 68 points within the face. Kazemi and Sullivan [37] trained the detector and provided the right model to be used in detecting the task. A cascading ensemble regression tree (ERT) was used to estimate the position of the points in the face landmark as shown in the following formula with an S shape, where the points is a member of the S vector.

In the training session, n face image (I) and the shape S were provided. The training data consists of $(I_i, S_i) i \in \{1 \dots n\}$ (I_i, S_i) where $i \in \{1 \dots n\}$, n is the number of training images. The cascaded regression trained using Eqs. (4), (5) and (6).

$$\pi_i \in \{1, \dots, n\} \tag{4}$$

$$\hat{S}_i^{(0)} \in \{S_1, \dots, S_n\} \mid S_{\pi_i} \tag{5}$$

$$\Delta S_i^{(0)} = S_{\pi_i} - \hat{S}_i^{(0)} \tag{6}$$

This is repeated for $I = 1 \dots N$ where $N = nR$ where R is the initialization for each image I_i. Based on this data, regression 0 ($r0$) yield $S_i^{(1)}$. The regression is cascaded from $t = 0$ to T. The formula for cascade regressing and estimating the shape is shown in Eqs. (7) and (8).

$$\hat{S}_i^{(t+1)} = \hat{S}_i^t + r_t(S_{\pi_i}, \hat{S}_i^t) \tag{7}$$

$$\Delta S_i^{(t+1)} = S_{\pi_i} - \hat{S}_i^{(t+1)} \tag{8}$$

This process is repeated in a cascade of T regressors $r_0, r_1, \dots, r_{T-1}$ in order to learn the best combination with a sufficient level of accuracy. Once the T regressor trained by n face images was obtained, a regression model for face landmark was defined. The pre-defined model was used to identify the face landmark.

Based on the face landmark, the eye location was identified and denoted by 37 - 42, and 43 - 48 on the left and right eye, respectively. The pupil is located in the middle of the eyes as the darkest object.

3.2.2. Face tracking

Face movement was used to detect the pose of the participant relative to the camera. It is three-dimensional, however, only horizontal and vertical movement were considered.

The video consisted of n frames with i = 1 to n - 1. The x and y positions were calculated by considering the displacement of the face position relative to the previous frame. The nose tips were considered as the centre of the face, while the displacement of the position of the nose tips was considered as face movement data. The value was set between -10 … 10, were recorded as 25 pixels for 1 unit.

$$f(x_i) = \begin{cases} round\left(\frac{x_{i-1} - x_i}{25}\right), & 0 < |x_{i-1} - x_i| < 250 \\ 10, & x_{i-1} - x_i \geq 250 \\ -10, & x_{i-1} - x_i \leq -250 \end{cases} \tag{9}$$

where x_i is the current frame, x_{i-1} is the previous frame face coordinates.

Face tracking produced two output variables which were horizontal (x) and vertical (y) movements. The horizontal and vertical movements, normalized in the range -10 to 10, and were calculated with respect to Eq. (9).

3.2.3. Eye-tracking

Eye-tracking shows user focus to the view. It provides a clue when the participant looks to another side of the test user interface focus. For example, when help is provided from outside the camera view, the user's eyes tends to leave the test sheet focus.

The normal eye condition is when the participant focuses on the testing area on the smart phone. The pupil position becomes in two-dimension space (x and y) with +/- relative to the center in the number of pixel displacement. For a single face, the left and right eye movements are assumed to be in the same direction.

The first task was to identify the face and eyes in the image. Followed by estimating the center of the eye by calculating x_2 - x_1 and y_2 - y_1, where x_2/ x_1 are it's left and right corner and y_1/y_2 are the top and bottom of the eye. Each frame sequence estimates the new position of the black node in the eye relative to the pupil in the previous frame. Figure 4 shows the face identification and pupil tracking during a particular frame in a cheating video. Once the position of the pupil is detected, the relative position is calculated against the normal with respect to Eq. (10).

$$f(x_i) = \begin{cases} x_{i-1} - x_i\,, & 0 < |x_{i-1} - x_i| < 10 \\ 10, & x_{i-1} - x_i \geq 10 \\ -10, & x_{i-1} - x_i \leq -10 \end{cases} \tag{10}$$

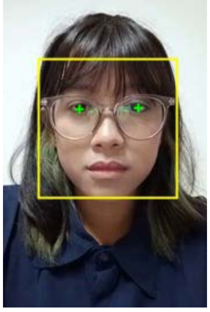


Fig. 4. Face and pupil identification in a particular frame.

3.2.4. Sound

During the simulation, the smart phone recorded the sound of the test environment, which is transformed in certain periods synchronous to the frame sampling. Furthermore, the sounds were normalized into 0 to 10 and put on the time-series data. The audio input amplitude represents the power of the sound and originally the peak of the loudest audio signal(peak). The peak signal usually achieve in certain point. The effective audible signal listened by human usually called the root mean square (RMS) amplitude. The original audio signal is distributed over positive and negative value; therefore, a positive and negative peak are existed. We can consider an audio signal as a composite of many sinusoidal signal in varied frequencies. Let's assume in certain small windows for example 1/12 second we sampled n=10 data {S_1, S_2, …. S_{10}}, RMS can be calculated using Eq. (11).

$$RMS = \sqrt{\frac{({S_1}^2+{S_2}^2+{S_3}^2+\ldots\ldots{S_n}^2\ )}{n}} \tag{11}$$

The RMS therefore all positive number, between 0 and 1. In our experiment this scale is normalized to 0 to 10.

3.3. Multilayer perceptron for activity recognition

Multilayer perceptron (MLP) is a class of artificial neural network (ANN) commonly referred to as "vanilla", especially when it consists of a single hidden layer [19]. The minimum architecture of MLP consists of three layers which are input, hidden, and output. Each node in the hidden and output layer is a neuron with a nonlinear activation function. The network is trained by a supervised learning technique called back-propagation to adjust the weight for each node [20, 21]. MLP is capable of handling non-linear separation of classes due to tits activation function [22]. This capability has become the main difference between MLP and linear perceptron.

It employs two common activation functions sigmoid, in Eq. (12) and rectifier linear unit (RELU) as described in Eq. (13):

$$S(x) = \frac{e^x}{1+e^{-x}} \quad (12)$$

$$R(x) = max(0, x) \quad (13)$$

where S(x) is the sigmoid activation function, x is the input variable and R(x) is the RELU function. Figure 5 shows the architecture of MLP in this research. It consists of input layer with 1290 nodes, varied number of hidden layer and single node in output layer.

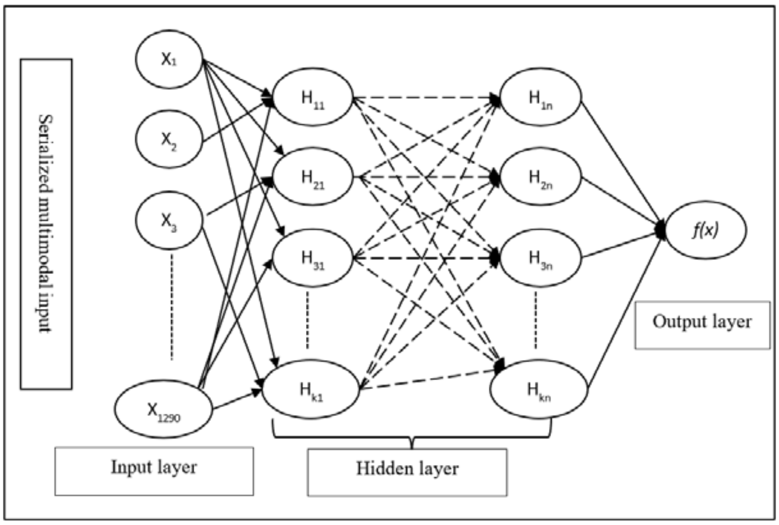


Fig. 5. Multi-layer perceptron architecture.

In the experiments conducted, the network was set with 10, 20, 30 and 40 hidden layers using the ‘RELU’ activation function as described in Eq. (13) and trained with ‘Adam’ optimizer. The input layer was set at 1290 nodes to fit with the number of input data which comes from the multi modal feature extraction of pupil tracking, face tracking and environment sound. The final output is a single node with binary class cheating and non-cheating.

3.4. Experimental design

This experimental research was carried out in a quiet testing environment away from the crowd to benefit the exam room. The research location was a 3x3 meter room which had bright lighting, tables, and chairs. Each participant sitting and working on exam questions all had a smart phone. Their profiles were quite diverse, with male and female, wearing/not wearing glasses, or veil. The data collection was carried out simultaneously among four participants in meeting rooms.

A simulation of suspicious activity was defined in seven possible cheating scenarios as follows:

- Opening a book activity.
- Reading a book next to the smart phone.
- Using a laptop to obtain information.
- Reading someone else's clue outside the camera.
- Sending and receiving messages from the second smart phone.
- Creating someone else's presence outside of the back camera.
- Asking questions from someone else through a voice simulation.

A total of twenty subjects were used to test all the scenarios. It involved recording seven activities for each subject in the cheating scenarios detailed above, as well as seven in the non-cheating categories. This led to a total of 280 activities.

Three low end smart phones with the specification are explained in Table 1.

Table 1. Smartphone specification.

Specs	Samsung A2 Core	Samsung A20	Vivio Y15
CPU	1.6 GHz Octacore	2x 1.6 GHz Octacore	2.0 GHz Octacore
RAM	1 GB	3 GB	4 GB
Display	5" 540 x 960 px	6.4" 720 x 1560 px	6.35"720 x 1544 px
Front Cam	5 MP	8 MP	16 MP
Back Cam	5 MP	13 Mp	13 MP
OS	Android 8.0	Android 9.0	Android 9.0

3.5. Evaluation

The data was divided into 5 folds and a cross-validation was carried out. The proportion of training and testing data were 80% and 20%, respectively. The quality metrics of the classification measured were accuracy, recall, precision, and f1 score.

4. Results and Discussion

The activity classification was carried out by a multilayer perceptron (MLP) on a standardized input parameter. In order to standardize the data, the duration of each activity in the listed scenarios were set to fifteen seconds. The shorter activity was padded with zero value in each variable.

A total of 12 frames per second were considered to be enough since the mobile exam scenario was controlled by ensuring the participants head movement while holding the smart phone was limited by the camera position.

The original audio-video signal was transformed into face, pupil and sampled audio tracking. The length of the output of the transformation was varied due to the duration of the video between 14 to 22 seconds. In order to ensure the equal length of the input vector to the network, zero padding was put at the end of the signal. The expected length of the signal was 258 to fit the number of input nodes in the MLP architecture. Figure 6 shows the snippet of sound signal representation of a particular activity. Meanwhile, Figs. 7 and 8 show the face and eye tracking diagrams, respectively. X and y-axis represent horizontal and vertical movement.

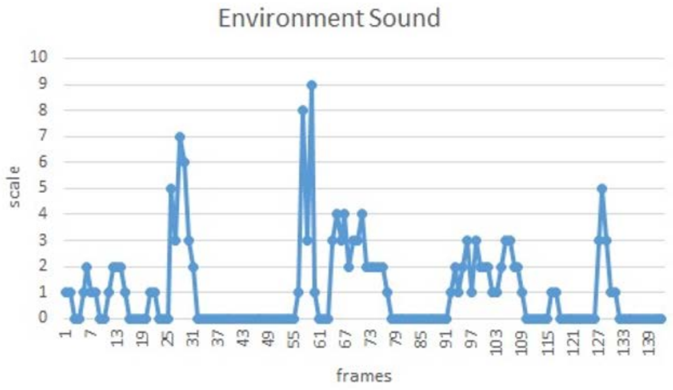


Fig. 6. Sound signal feature.

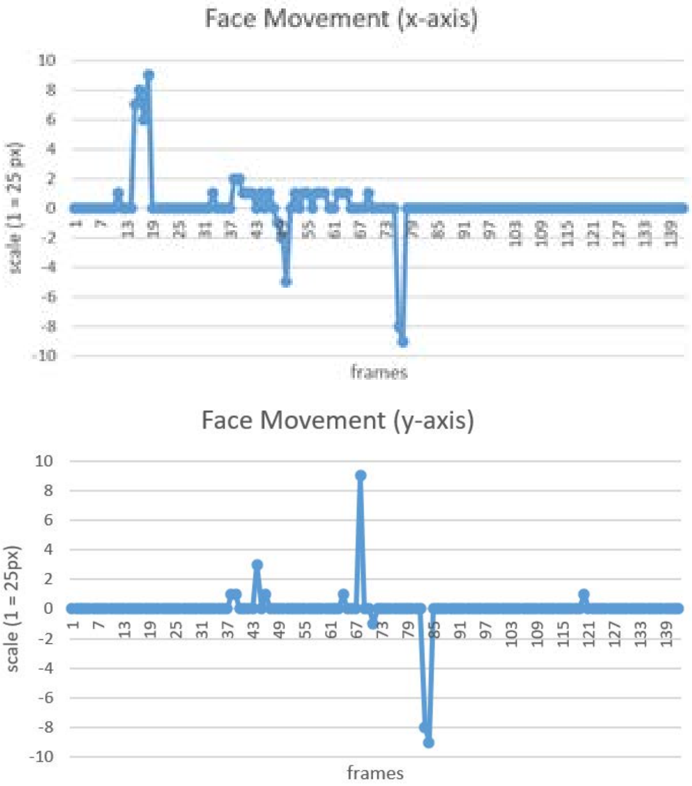


Fig. 7. Face tracking movement.

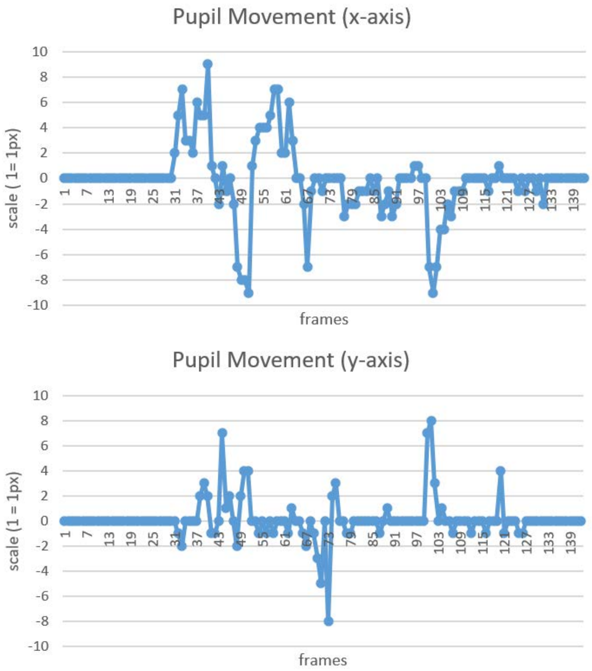


Fig. 8. Eye tracking signal.

The signal was concentrated into a single vector for each activity while the mid-level signal 5×258 was transformed into a 1 x 1290 size.

The entire mid-level signal for all the dataset consisted of 280 rows, with the data randomly divided into five folds and cross-validated using the classifier. Table 2 shows the average, accuracy, precision and f1score of the classification result using MLP with varied hidden layer. Table 2 showed that 20 hidden layers achieved the best classification metrics. Table 3 shows the classification metrics of 5 folds experiments.

Table 2. Average accuracy, recall, precision and F1 score for different hidden layer.

Hidden Layer	Accuracy	Recall	Precision	F1 Score
10	0.910325	0.910325	0.911421	0.910325
20	0.917338	0.917338	0.929832	0.916798
30	0.903312	0.903312	0.909436	0.902822
40	0.903182	0.903182	0.912396	0.902383
50	0.913896	0.913896	0.918950	0.913387

Table 3. Classification result for each fold in 20 hidden layers.

Fold	Accuracy	Recall	Precision	F1 Score
1	0.928571	0.928571	0.928571	0.928571
2	0.892857	0.892857	0.909774	0.889796
3	0.964286	0.964286	0.96659	0.964194
4	0.946429	0.946429	0.950893	0.945724
5	0.854545	0.854545	0.893333	0.855703
Average	0.917338	0.917338	0.929832	0.916798
Std Dev	0.039291	0.039291	0.026573	0.039201

According to the experiments, the best model was achieved on the third fold, with a successful 96% classification accuracy based on the available data. The worst experiment achieves 85% accuracy, while the balance weight between the recall and precision in calculating the F1 score was considered due to the balance portion of cheating and non-cheating classes in the dataset.

The conversion of the video data into face and pupil tracking were carried out successfully. The down sampling was reduced into 12 frames per second to produce a good result where the achieved accuracy was above 91.7% on average and 96% in the best model as shown Table 3. The simple mechanism to transform video and audio data into a simple representation still has enough discriminating power for a binary classifier.

An experiment of using multimodal input from webcam, gaze tracked and EEG devices [10]. Three scenarios were performed which are automatics cheating detector (ACD) which is comparable to our research. They also proposed a combination of the ACD with peer cheating detector (PCD) and Final review committee (FRC). PCD involved peer exam taker to review piece of suspicious exam video raised by automatic detection (ACD). Once PCD conclude that the piece of video is suspicious cheating, a final review committee work for final decision.

Table 4 shows the result of Le et al. [10]. Compared to single webcam ACD and Multimodal ACD, our result outperforms in recall, precision and accuracy. The fair comparison with single webcam only (86.1%) However, compare to peer cheating detection (PCD) in Table 4, our achievement slightly under their achievement. This is reasonable due to human involvement in their PCD experiment.

Table 4. The performance of previous research [10].

Cheating Classifier	Recall	Precision	Accuracy
Webcam-only ACD	76.60%	84.60%	83.30%
Multi-modal ACD	82.10%	60.00%	81.10%
PCD	96.40%	94.10%	95.60%
Webcam-only MOOP	78.60%	90.40%	86.10%
Multi-modal MOOP	90.50%	93.80%	92.70%

Atoum et al. [11] proposed a multimodal proctoring using double camera and sound recorder. They need a special camera (wearcam) installed on the eyeglasses to observe the view area of the participant. They achieve 87% of cheating recognition using SVM classifier. Our result was over 4% better at 91.73% compare to theirs’.

This research is limited by the pre-defined cheating action under specified requirements in the experimental setting. However, despite the unavailability of empirical data, and the specified resolution of the video, the result might be degraded. Secondly, other types of cheating are likely to be beyond the simulated activity pre-set in this research, with undefined outcomes. The lighting in this study was set in stable and visible condition, therefore, a darker situation is likely to lead to a significant decrease on recognition quality. The noisy sound is also a challenge as the dataset was taken in a controlled condition where the noise is at a minimum level. Another potential problem of this research is privacy issue, due to the need of recording the video and audio during the exam. In the future, a signal transformation on the recording device would be an alternative to overcome the privacy issue. Due to the limited possibility of cheating/non-cheating actions detected, the exact activity was not precisely recognized such as reading a book, using mobile phone etc. These limitations are a guide for future research.

5. Conclusion

This study presented an audio-video analysis for cheating action recognition to support a mobile exam. Therefore, in conclusion, the audio and video signals are necessary to recognize the cheating and non-cheating activity. The average accuracy, precision, recall, and f1 score were found to be at 91.73±3.9%, 91.73±3.9%, 92.99±2.6% and 91.67±2.3%, respectively. A high recognition rate was achieved, compared to the state of the art in online cheating detection. Compare to multimodal automatic detection and double camera plus audio from the literature, this research achieved better accuracy with almost a 4% margin. In addition, this research was designed for a smart phone exam environment with an implementation of standard accessories without any additional device. Further research needs to be conducted for multiclass classification in order to identify the action conducted. The possibility of using more sensors in the smart phone, such as accelerometer and the gyro tends to enrich the dataset and is expected to improve the accuracy.

Acknowledgment

The research activity leading to the publication has been funded by the Ministry of Research, Technology/National Research and Innovation Agency (Indonesia) Grant No. 227/SP2H/AMD/LT/DRPM/2020. Also, the authors would like to thank the research Directorate of Universitas AMIKOM Yogyakarta for their support.

References

1. Seaman, J.E.; and Seaman, J. (2017). *Distance education state almanac 2017*. Retieved September 5, 2019, from https://onlinelearningsurvey.com/reports/almanac/national_almanac2017.pdf
2. Allen, I.E.; and Seaman, J. (2014). *Grade change tracking online education in the United States*. Retrieved September 5, 2019, from https://www.onlinelearningsurvey.com/reports/gradechange.pdf
3. Fask, A.; Englander, F.; and Wang, Z. (2014). Do online exams facilitate cheating? An experiment designed to separate possible cheating from the effect of the online test taking environment. *Journal of Academic Ethics*, 12(2), 101-112.
4. King, C.G.; Guyette, R.W.; and Piotrowski, C. (2009) Online exams and cheating: An empirical analysis of business students’ views. *Journal of Educators Online*, 6(1), 1-11.
5. King, D.L.; and Case, C.J. (2009). E-cheating: Incidence and trends among college students. *Issues in Information Systems*, 15(1), 20-27.
6. Corrigan-Gibbs, H.; Gupta, N.; Northcutt, C.; Cutrell, E.; and Thies. W. (2015). Deterring cheating in online environments. *ACM Transactions on Computer-Human Interaction*, 22(6).
7. Foster, D.; and Layman, H. (2013). Online proctoring system compared. *U.S. Pat. Appl. No. 12/723,666.*, 1-12.
8. Kaiiali, M.; Ozkaya, A.; Altun, H.; Haddad, H.; and Alier, M. (2016). Designing a secure exam management system (SEMS) for M-learning environments. *IEEE Transactions on Learning Technologies*, 9(3), 258-271.

9. Bawarith, R.; Basuhail, A.; Fattouh, A.; and Gamalel-Din, S. (2017). E-exam cheating detection system. *International Journal Advanced Computer Science and Application*, 8(4), 176-181.
10. Li, X.; Chang, K.; Yuan, Y.; and Hauptmann, A. (2015). Massive open online proctor: Protecting the credibility of MOOCs Certificates. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*. Vancouver BC, Canada, 1129-1137.
11. Atoum, Y.; Chen, L.; Liu, A.X.; Hsu, S.D.H.; and Liu, X. (2017). Automated online exam proctoring. *IEEE Transactions on Multimedia*, 19(7), 1609-1624.
12. Hung, H.; and Gatica-Perez, D. (2010). Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia*, 12(6), 563-575.
13. Nguyen, L.S.; Frauendorfer, D.; Mast, M.S.; and Gatica-Perez, D. (2014). Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Transactions on Multimedia*, 16(4), 1018-1031.
14. Zajdel, W.; Krijnders, J.D.; Andringa, T.; and Gavrila, D.M. (2007). CASSANDRA: Audio-video sensor fusion for aggression detection. Proceedings of the *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*. London, United Kingdom, 200-205.
15. Lefter, I.; Burghouts, G.J.; and Rothkrantz, L.J.M. (2012). Automatic audio-visual fusion for aggression detection using meta-information. *Proceedings of the IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance*. Beijing, China, 19-24.
16. Wang, Z.; Yu, J.; and He, Y. (2017). Soccer video event annotation by synchronization of attack-defense clips and match reports with coarse-grained time information. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(5), 1104-1117.
17. Kolekar, M.H.; and Sengupta, S. (2015). Bayesian network-based customized highlight generation for broadcast soccer videos. *IEEE Transactions on Broadcasting*, 61(2), 195-209.
18. Cluskey, G.R.; Ehlen, C.R.; and Raiborn, M.H. (2011). Thwarting online exam cheating without proctor supervision. *Journal of Academic and Business Ethics*, 4, 1-7.
19. Wahid, A.; Sengoku, Y.; and Mambo, M. (2015). Toward constructing a secure online examination system. *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*. Bali, Indonesia, 1-8.
20. Clarke, N.L.; Dowland, P.; and Furnell, S.M. (2013). e-Invigilator: A biometric-based supervision system for e-assessments. *Proceedings of the International Conference on Information Society*. Toronto, Canada, 238-242.
21. Asha, S.; and Chellappan, C. (2008). Authentication of e-learners using multimodal biometric technology. *Proceedings of the International Symposium on Biometrics and Security Technologies*. Islamabad, Pakistan, 1-6.
22. Jalali, K.; and Noorbehbahani, F. (2017). An automatic method for cheating detection in online exams by processing the student's webcam images. *Proceedings of the 3rd Conference on Electrical and Computer Engineering Technologies*. Tehran, Iran, 1-6.

23. Kumar, A.V.S.; and Rathi, M. (2019). *Biometric Authentication in Online Learning Environments*. Chapter 8: Keystroke dynamics: A behavioral biometric model for user authentication in online exams. IGI Global, 183-207.
24. Bal, A.; and Acharya, A. (2011). Biometric authentication and tracking system for online examination system. *Proceedings of the International Conference on Recent Trends in Information Systems*. Kolkata, India, 209-213.
25. Fayyoumi, A.; and Zarrad, A. (2014). Novel solution based on face recognition to address identity theft and cheating in online examination systems. *Advances in Internet of Things*, 4(2), 5-12.
26. Xiao, B.; Georgiou, P.; Baucom, B.; and Narayanan, S.S. (2015). Head motion modeling for human behavior analysis in dyadic interaction. *IEEE Transactions on Multimedia*, 17(7), 1107-1119.
27. Liu, Y.; Xie, Z.; Yuan, X.; Chen, J.; and Song, W. (2017). Multi-level structured hybrid forest for joint head detection and pose estimation. *Neurocomputing*, 266, 206-215.
28. Dornaika, F.; and Raducanu, B. (2009). Three-dimensional face pose detection and tracking using monocular videos: tool and application. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(4), 935-944.
29. Qing, C.; Jiang, J.; and Yang, Z. (2010). Normalized co-occurrence mutual information for facial pose detection inside videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(12), 1898-1902.
30. Drouard, V.; Horaud, R.; Deleforge, A.; Ba, S.; and Evangelidis, G. (2017). Robust head-pose estimation based on partially-latent mixture of linear regressions. *IEEE Transactions on Image Processing*, 26(3), 1428-1440.
31. Cheung, Y.; and Peng, Q. (2015). Eye gaze tracking with a web camera in a desktop environment. *IEEE Transactions on Human-Machine Systems*, 45(4), 419-430.
32. Krafka, K.; Khosla, A.; Kellnhofer, P.; Kannan, H.; Bhandarkar, S.; Matusik, W.; and Torralba, A. (2016). Eye tracking for everyone. *Proceedings of the IEEE Computer Vision and Pattern Recognition*. Las Vegas, Nevada, 2176-2184.
33. Meng, C.; and Zhao, X. (2017). Webcam-based eye movement analysis using CNN. *IEEE Access*, 5, 19581-19587.
34. Tran, H.D.; and Li, H. (2011). Sound event recognition with probabilistic distance SVMs. *IEEE Transactions on Audio, Speech and Language Processing*, 19(6), 1556-1568.
35. Wang, C.; Wang, J.; Santoso, A.; Chiang, C.; and Wu, C. (2018). Sound event recognition using auditory-receptive-field binary pattern and hierarchical-diving deep belief network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(8), 1336-1351.
36. Dalal, N.; and Triggs, B. (2004). Histograms of oriented gradients for human detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, USA, 1, 886-893.
37. Kazemi, V.; and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, Ohio, 1867-1874.

Judul : Cheating Activity Detection on Secure Online Mobile Exam

Penulis : Arief Setyanto, Bayu Setiaju, Mardhiya Hayaty, Krisnawati

Jurnal : Journal od Engineering Science and Technology

Penerbit : School of eningeering, Taylor's University

Web Jurnal : https://jestec.taylors.edu.my/

Web Paper : https://jestec.taylors.edu.my/Vol%2015%20issue%206%20December%202020/15_6_34.pdf

DOI :

Index : Q3, Engineering

SJR : 0,224

Search Journals | Match Manuscript | Downloads | Help Center

Welcome, Arief Setyanto

Settings | Log Out

Check out our new metric to help you evaluate journals! Dismiss Learn More

JOURNAL OF ENGINEERING SCIENCE AND TECHNOLOGY

Share This Journal

ISSN / eISSN **1823-4690**

Publisher **TAYLORS UNIV SDN BHD, 1 JALAN SS15-8, SUBANG JAYA, SELANGOR, MALAYSIA, 47500**

About

JESTEC (Journal of Engineering Science and Technology) is a peer-reviewed journal that aims at the publication and dissemination of original research articles on the latest developments in all fields of engineering science and technology. The journal publishes original papers in English which contribute to the understanding of engineering science and improvement of the engineering technology and education. Papers may be theoretical (including computational), experimental or both. The contribution should be unpublished before and not under consideration for publication elsewhere. Journal of Engineering Science and Technology (JESTEC) maintains a standard double-blind peer review process. The double-blind process means that the identity of the author and the reviewer are not known to each other. JESTEC is an Open Access Journal, and does not charge readers or their institutions for access to the journal articles. The open access supports the rights of users to read, download, copy, distribute, print, search, or link to the full texts of these articles provided they are properly acknowledged and cited. JESTEC publishes 12 issues per year not including special issues. Publication fees in JESTEC include submission, reviewing, editing, publishing, and uploading the accepted article in JESTEC website. For all these services JESTEC charges USD200 per article. The number of pages per article should be minimum 8 pages and maximum 20 pages.

General Information

Journal Website	Visit Site	**Publisher Website**	Visit Site
1st Year Published	2006	**Frequency**	Monthly
Issues Per Year	12	**Country / Region**	MALAYSIA
Primary Language	English	**Aims and Scope**	Visit Site
Editorial Board	Visit Site	**Instructions for Authors**	Visit Site
Avg. Number of Weeks from Submission to Publication	12	**Host Platform**	Taylor's University
Full-Text Formats	PDF		

Journal Citation Report™ (JCR)

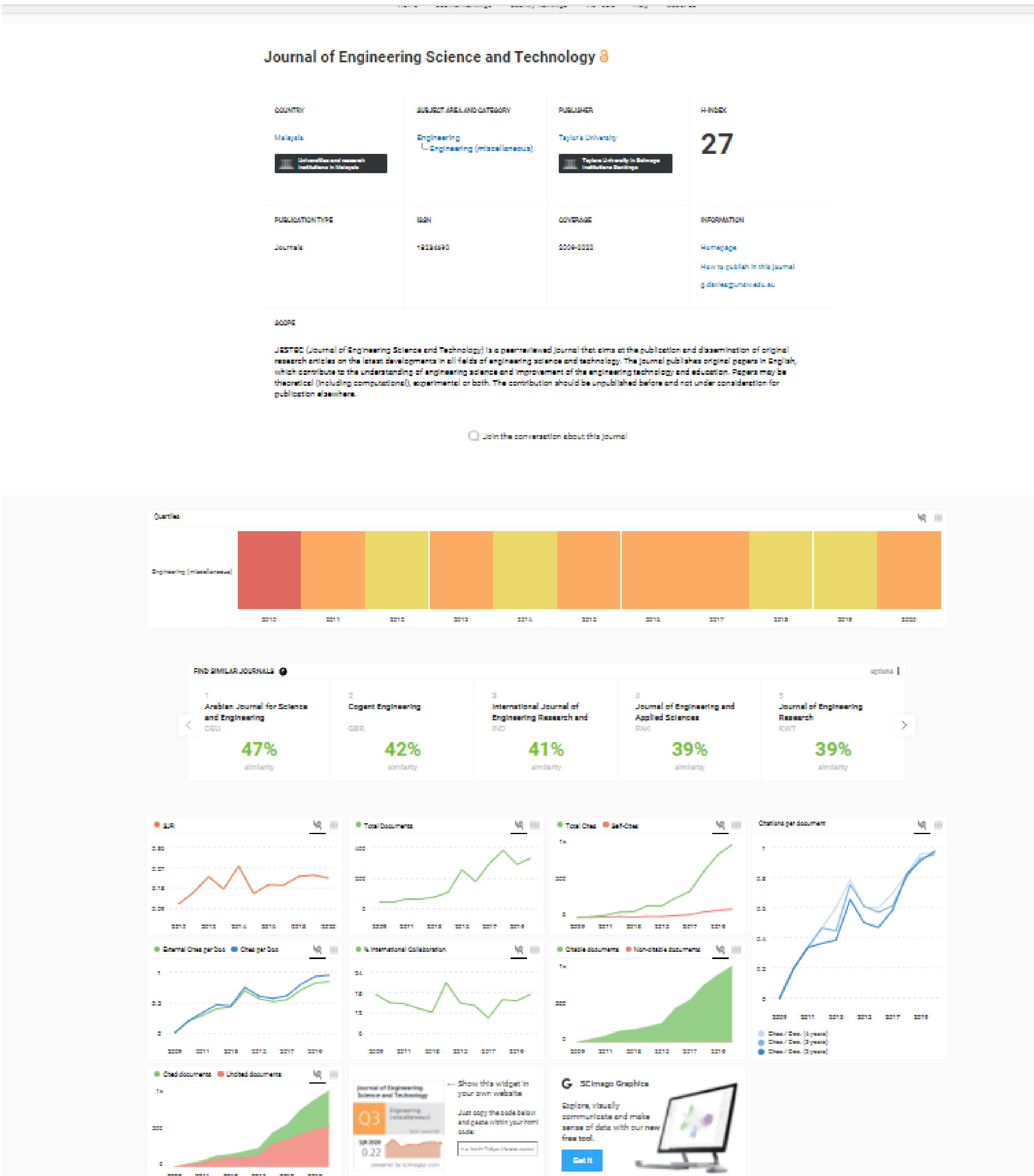
Journal Citation Reports™ 2021

Journal Citation Indicator (JCI) NEW METRIC

The Journal Citation Indicator is a measure of the average Category Normalized Citation Impact (CNCI) of citable items (articles & reviews) published by a journal over a recent three year period. It is used to help you evaluate journals based on other metrics besides the Journal Impact Factor (JIF).

2020	2019
0.24	0.25
Category: Engineering, Multidisciplinary	**Category:** Engineering, Multidisciplinary

Learn About Journal Citation Indicator



Sumber : https://www.scimagojr.com/journalsearch.php?q=18200156709&tip=sid&clean=0

pre-submission

by Arief Setyanto

Submission date: 12-Dec-2019 01:02PM (UTC+0700)
Submission ID: 1232942406
File name: Mobile_Online_Exam_subjectAuth_JESTEC.docx (982.09K)
Word count: 4949
Character count: 25807

7
Journal of Engineering Science and Technology
Vol. XX, No. Y (Year) PPP - QQQ

CHEATING ACTIVITY DETECTION ON MOBILE SECURE ONLINE EXAM

22
Abstract

The online exam is one of the important tasks in online learning systems. Online exam proctoring, therefore, is important to ensure the credibility of the exam. The participants who take apart in the online exam are distributed in many locations. This work proposes an online exam system with an android application. We utilize the camera and audio recorder to capture human activity during the test. We collect the audio-video from 20 participants in seven cheating and one non-cheating scenario. To get the balance data between cheating -non-cheating activity, the participant, take seven cheating and seven non-cheating scenarios. The system record the face of a participant using the front camera and recording the sound. We propose a mid-level representation of the audio-video before the classification task. We perform data normalization into uniform units for each parameter between zero to ten. Finally, the MLP carries out the classification of a midlevel signal into cheating and no- cheating activity. We found that the accuracy of activity classification achieves 90.7%, while precision at 90.7% recalls 91 % and F1 score at 90.6%.

Keywords : Action Detection, Mobile Exam, Eye Detection, Face Detection, Multilayer Perceptron

1. Introduction

Electronics learning currently has become one of the games changing applications in the education sector. The traditional education in classical mode has gradually changed into online mode either in synchronous and asynchronous learning. The online meeting has been enabled by online video meeting applications, while an asynchronous model facilitated by online documents ranging from a static powerpoint, animation software, sharing a video to more complex interactive learning tools such as interactive programming tools. According to a survey in [1], 14.3% of American students taking online courses only without a traditional degree, while 15 % of students take a combination online and traditional degree. A significant improvement happened to compare to the same survey in 2013 [2]. The online education participants improved from 411.000 to 7.1 million, while the online education provider improved from 2.6% to 5% over the past year. Although the survey has been done in the USA, it indicates the general trend all over the world.

Examination plays an important role in ensuring the participants have understood the learning material. The examination can be either synchronous and asynchronous. Synchronous exam where the examiner and examinee work at the same time, it can be done by video call exam. The asynchronous online exam can be done by an interactive questionnaire done by the participants. Asynchronous mode permit examinee examines at any time and any place. Online proctoring would not be possible in this mode; therefore, if we need to ensure the test is properly done, an automatic proctoring is needed.

To date, several researchers have reported that online exam opens the higher possibility of cheating [3]–[5]. According to [4] and [5], 74% of students considered that cheating in the online exam is easy. While 29% of the respondents in their research committed cheating during an online exam. According to [6], honor code only gives the insignificant effect of reduction in online cheating, while warning double the reduction of cheating behavior.

Online proctoring is desirable due to the growth of massive open online courses (MOOC). Some online proctoring products available in the market such as, e.g., Kryterion, Web assessor and ProctorU, offer online proctoring services [7]. They usually put a surveillance camera on the user site; however, they still rely on human watching over through video streaming. In [8], they propose a mobile exam system embedded in the Moodle system and identify various vulnerabilities that may violate exam security in m-learning environments and to design the appropriate security services and countermeasures that can be put in place to ensure exam security.

Automatics proctoring task has received attention from the researcher community. In [9], front face camera to track the participant eye movement. The exploitation of eye movement leads to an understanding of user focus. When the user focus leaves the testing sheet, it can be assumed that something happens. The multimedia analysis helps online proctoring automation [10], [11]. Gaze tracker, webcam and EEG installed to help online exam proctoring [10]. In [11], they set two cameras, which are one webcam, and a camera installed on special designed eyeglass (webcam). They record the user's face, capturing the participant's view area. Their design is working properly; however, they need a special installed camera to reduce the practicality.

Audiovisual has been exploited for activity/event recognition for other purposes such as social interaction, object tracking and sports analysis. Social cohesion among group members extracted from video and classified by support vector machine [12]. An out of the box way of identifying good behavior of job seeker utilized interviews video proposed by [13]. Identifying unwanted behavior in audiovisual data has also done by [14], [15]. Sport activity recognition such as attack defense, match report proposed by [16], while goals, saves, kick of recognized using Bayesian network [17].

This research makes an effort to avoid installing the special device to perform the online proctoring. We, therefore, design the exam on a smartphone with a requirement of double camera front and back, GPS, gyro sensors and voice recorder. Currently, those accessories are available on most of the smartphones. We will use those standard accessories to provide multimodal data captured during the exam. The camera, the voice recorder is responsible for user verification and test surveillance.

2. Related work

There are three possibilities regarding online exam model proctoring. Firstly, the exam leaves without human controlling [18], [19] [20]. Secondly, the human involved either online through a webcam or physically in a certain location [21]. The third option leaves the participant take the test without proctoring but records the exam process and let the multimedia analysis system work[11], [22].

In the Exam, we need to ensure user identity. There are many efforts in ensuring computer user identity, for example, utilizes keystroke behavior or biometrics. Keystroke is a major activity in using a computer; therefore, it attracts researchers' attention to identifying user activity. In [23], researchers use keystroke dynamics for online exam authentication. In [24], iris recognition for biometrics authentication for the same purpose. In [25], face recognition periodically has done during the exam in order to ensure the authentication of the examinee.

Beside participant identity, user action during an exam is an important clue of cheating behavior. There is much research dealing with action recognition such as [26]–[28]. Head movement is one important clue for activity recognition. [26] detect human activity by extracting head motion to analyze the socio-communicative and affective behavioral characteristics of interacting partners. There is much technique for head pose estimation. [27] using the multilevel structured hybrid forest to detect head and estimating the pose. Dornaika et. proposed 3D face pose estimation using 2D eigenfaces and DE algorithm [28]. Mutual information (MI) exploited to deal with pose estimation in a uncontrollable environment [29]. Another technique proposed by [30] exploited a mixture of linear regression with partially-latent output.

Eye-tracking plays an important role in understanding human attention. In the examination proctoring, exploitation of eye direction will be able to point out the user's point of view. Currently, most eye movement technology developed using near-infrared light. The human pupil will light up when they receive near-infrared light sources, and therefore, the currently available technology benefit from that pupil behavior. It is, however, costly and lacks practicability due to a need for the special device and even location for eye movement detection. In many medical purposes, this technology is viable due to the accuracy requirement and facilities availability.

However, this research would not be able to provide a special camera with near-infrared sources; therefore, we seek simple techniques to detect and track the pupil with a simple web camera. Some researchers made an effort to enable tracking the eye without near-infrared sources such as [31], [32] Cheung et. All proposed a webcam in desktop to recognize the pupil, eye corner and movements [31]. In 2016, [32] propose a normal camera for eye movement recognition. Meng et. All proposed CNN based eye movement detection and tested in a half-million frames and successfully recognize activities such as reading, browsing and watching video [33].

Sound has been identified as a strong clue for activity recognition in [11] sound considered to be a clue of cheating if a human voice exists. Therefore, speech extraction performed in modeling the cheating action. Sound event recognition has been attracted researcher attention, such as [34][35].

The contribution of this paper is:

Developing a mobile exam video cheating dataset consists of cheating and one non-cheating activity from twenty subjects, 280 actions with 36.120 frames in total.

Defining a midlevel data modeling for audio-video cheating dataset consists of head pose, eye tracking and audio.

Develop an artificial neural network to perform cheating activity detection based on the mid-level dataset.

The outline of the remaining paper. We explain our proposed methods in section 3, and part 4 is described result and discuss the achievement, in the last part we draw our conclusion

3. Multimodal Action Detection

3.1 Research Framework

This research is divided into four main tasks which are, defining the cheating actions, develops a dataset for pre-defined actions, mid-level feature extraction, and action recognition task. Figure 1 depicts the entire research framework.

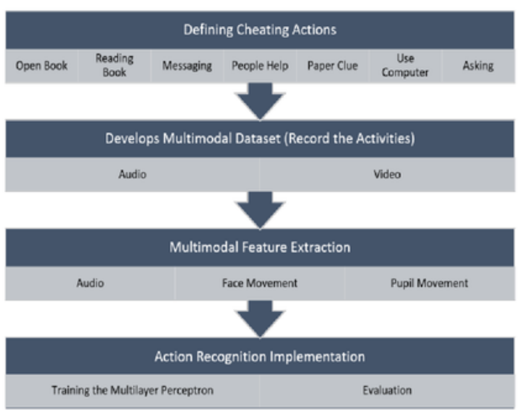


Figure 1 Research Framework

3.2 Multimodal Input

Multimodal cheating behavior in online exams involves audio and video data. We pre-process the video from the front camera of the smartphone. Before further processing, a standardization of the visual signal is carried out to reduce the volume of the data while maintaining the quality for feature extraction purposes. The size of the frames is under-sampled into 180 x 320 pixels size and reduces the frame rate into 12 per second. We consider that this data is enough considering that the video is quite low motion and the distance between the object under surveillance is closed enough. The audio is taken continuedly during the data collection for further processing. Figure 2 shows the condition of the normal position of a smartphone, a front camera and the face of the test taker.

Figure 2 Normal position of a mobile exam test taker

3.2.1. Face and Eye Recognition

The region of interest in the data transformation to the mid-level data representation is the face. Therefore, we need a number of task to localized the face and the eyes. Firstly, resize the frames into 180 x 320 pixels per frame with 12 frames per second (fps). Secondly, the face location has to be identified and the face landmark is calculated in order to identify key point of the face such as nose tip and eyes.

We extract the face from the whole image in each frame. use the histogram of gradients HOG [36]. We follow the implementation of [36] by divide the image into a 16x16 patch and calculate the orientation of the gradient in 9 bins of unsigned orientation between 0 to 360 degrees. The vector is then normalized to get L2-Norm, L1-Norm and L1-Sqrt in equations 1,2, and 3.

$$L2-Norm : f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad \textbf{(1)}$$

$$L1-Norm : f = \frac{v}{\|v\|_1 + e} \quad \textbf{(2)}$$

$$L1-Sqrt : f = \sqrt{\frac{v}{(\|v\|_1 + e)}} \quad \textbf{(3)}$$

Where v is the non-normalized vector containing all histograms in a given block and $\|v\|_k$ is the k-norm (k = 1,2 ..) and e is constant with a small value?

According to the Dalal et.all in [36], the histogram orientation with L2-Norm and L1-Sqrt gives a good performance. The result of this procedure is a localized face area, as figured out below.

After the face location is determined, the face landmark has to be identified. We implement the face landmark recognition proposed by [37]. Detect the face orientation in each frame and record the relative position to front face direction. The basic idea is finding 68 points within the face. Kazemi et.all has trained the detector and provide the model to be used in the detection task. They use a cascading ensemble regression tree (ERT) to estimate the position of the points in the face landmark. They use the formula below to define the face landmark as a shape S, where the points in the face is a member of the S vector.

In the training session, n face image (I) provided and the shape S. The training data consists of (I_i, S_i) where $i \in \{1 \dots n\}$, n is the number of training images. The cascaded regression trained using equations 4,5 and 6.

$$\pi_i \in \{1, \dots, n\} \quad (4)$$

$$\hat{S}_i^{(0)} \in \{S_1, \dots, S_n\} \mid S_{\pi_i} \quad (5)$$

$$\Delta S_i^{(0)} = S_{\pi_i} - \hat{S}_i^{(0)} \qquad (6)$$

This is repeated for I = 1… N where N =nR where R is the initialization for each image I_i. Based on this data, regression 0 (r0) learn to yield the $S_i^{(1)}$. The regression is cascaded till from t = 0 to T. The formula for cascade regressing an estimating the shape can be seen in equations 7 and 8.

$$\hat{S}_i^{(t+1)} = \hat{S}_i^t + r_t(S_{\pi_i}, \hat{S}_i^t) \qquad \mathbf{(7)}$$

$$\Delta S_i^{(t+1)} = S_{\pi_i} - \hat{S}_i^{(t+1)} \qquad (8)$$

This process is repeated in a cascade of T regressors $r_0, r_1, \ldots, r_{T-1}$ to learn the best combination give a sufficient level of accuracy. Once the T regressor trained by n face images, a regression model for face landmark defined. This paper uses the pre-defined model to identify the face landmark.

Based of the face landmark, Identify the eye location Eye location is denoted by 37 – 42 of the left eye in the face landmark, and 43 – 48 for the right eye. The pupil is located in the middle of the eyes as the darkest object.

3.2.2. Face Tracking

Face movement responsible for detecting the pose of the participant relative to the camera. The face able to move in three-dimensional; however, we consider horizontal and vertical movement only.

Let the video is consist of n frames. The first frame would be frame 0, and the face tracking data start with i = 1 to n - 1. The x and y position is calculated by considering the displacement of the face position relative to the previous frame. The nose tips are considered as the center of the face. We consider the displacement of the position of the nose tips as face movement data. The value is set between 0 – 10 and whenever the value > 10, then it is recorded as 10.

$$f(x_i) = \begin{cases} x_{i-1} - x_i, & 0 < |x_{i-1} - x_i| < 10 \\ 10, & |x_{i-1} - x_i| \geq 10 \end{cases} \qquad \mathbf{(9)}$$

Where x_i is the current frame, x_{i-1} is the previous frame face coordinates.

Face tracking produces two output variable which is horizontal movement (x) and vertical movement y. The horizontal and vertical movement, then normalized in the range 0 to 10, is calculated respect to equation 9.

3.2.3. Eye-tracking

Eye-tracking indicates user focus to the user view. It will give us a clue when the user looks to another side of the test user interface focus. For example, when someone helps him with the paper outside of the camera view, the user's eyes will leave the test sheet focus.

The normal eye condition when the participant focus on the testing area (smartphone). Pupil position would be in two-dimension space x and y with +/- relative to the center of the eyes. For a single face, we assume that the left and right eye movements in the same direction.

The first task is to identify the face and the eyes in the face image. Secondly, estimate the center of the eye by identifying calculating $x_2 - x_1$ and $y_2 - y_1$, where x_2 and x_1 are the left and the right corner of the eye and y_1 and y_2 as the top and bottom of an eye. The job in every frame sequence is estimating the new position of the black node in the eye relative to the pupil in the previous frame. Figure 4 shows the face identification and pupil tracking during a particular frame in a cheating video. Once the position of the pupil detected, the relative position is calculated against the normal position of respect to equation 9.

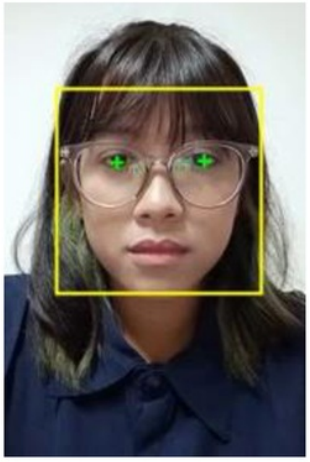


Figure 3 Face and pupil identification in a particular frame

3.2.4. Sound

During the simulation, the smartphone records the sound of the test environment. We transform the recorded sound and sample the amplitude of the sound in certain periods synchronous to the frame sampling. We normalized the sound into 0 to 10 and put on the time-series data.

3.3 Multilayer perceptron for activity recognition

Multilayer perceptron (MLP) is a class of artificial neural network (ANN). Multilayer perceptrons commonly referred as "vanilla" neural networks, especially it have a single hidden layer only [19]. The minimum architecture of MLP consists three layers which are an input, a hidden, and an output layer. Each node in the hidden and output layer is a neuron with a nonlinear activation function. Training the network to adjust the weight for each nodes is carried out by a supervised learning technique called backpropagation [20][21]. MLP capable to deal with non linear separation of classes due to the function of nonlinear activation function [22]. This capability is become the main differences between MLP and a linear perceptron.

MLP employs two common activation functions sigmoid in equation, 6 and rectifier linear unit (RELU) as described in equation 7:

$$S(x) = \frac{e^{x}}{1+e^{-x}} \tag{6}$$

$$R(x) = \max(0, x) \tag{7}$$

Where S(x) is the sigmoid activation function, x is the input variable, R(x) is the RELU function.

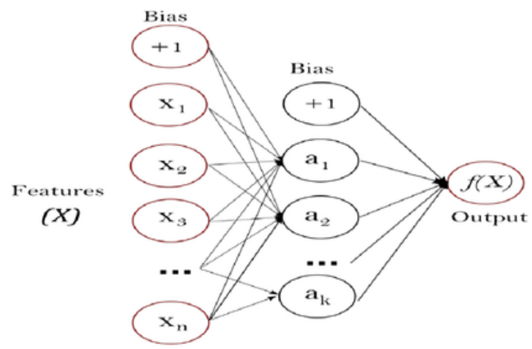


Figure 4 Multi-Layer Perceptron

In our experiments, we set the network with 20 hidden layers with 'RELU' activation function as described in equation 7 and trained using 'adam' optimizer. The input layer is set at 1290 nodes and the final output is a single node with binary class cheating and non-cheating.

3.4 Experiment Design

This research is designed as experimental research. A simulation of suspicious activity will be defined in seven possible cheating scenarios as follow

- Opening a book activity
- Reading a book on next to the smartphone
- Using a laptop to browse information
- Read some else clue outside the camera

- Send and receive a message from the second smartphone
- Someone else presence outside of the back camera
- Asking for someone else trough voice

Those activities are considered as cheating during an online exam. In this research, we make a real experiment using 20 subjects to test all the scenarios. We recorded 7 activities for each subject, and in total 140 activities recorded and another 7 times of each subject in normal (non-cheating) activity. Therefore, we have 280 activities in total, 140 cheating and 140 non-cheating.

3.5 Evaluation

We divide the data into 5 folds and carry out cross-validation. The proportion of training and testing data 80%, 20%, respectively. The evaluation is done for those five classifiers in each round. The quality metrics of the classification measured are accuracy, recall and precision and f measure.

4. Result and Discussion

The activity classification is carried out by a multilayer perceptron (MLP) on a standardized input parameter. In order to standardize the data, the duration of each activity in the listed scenarios are set in a fifteen-second duration. The shorter activity is padded with zero value in each variable.

We consider 12 frames per second is enough since the mobile exam scenario is controlled. In fact, when the examinee uses the smartphone the movement of the head will be limited by the camera position hold by hand would not be possible to move too far.

The original audio-video signal is transformed into face tracking, pupil tracking and sampled audio. The length of the output of the transformation is varied due to the duration of the video between 14 to 16 seconds. In order to ensure the equal length of the input vector to the network, we put zero paddings at the end of the signal. The expected length of the signal is 258 to fit the number of input nodes in the MLP architecture. Figure 5 shows the snippet of mid-level sound signal representation of a particular activity. Figure 6 and 7 shows the face tracking and eye tracking respectively. X and y-axis represent horizontal and vertical movement.

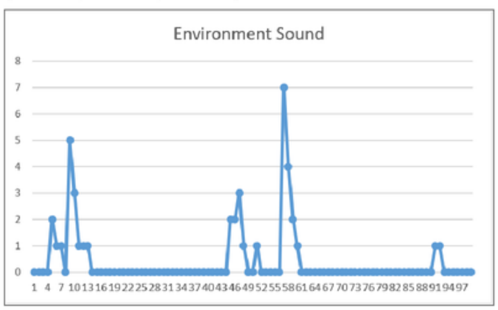
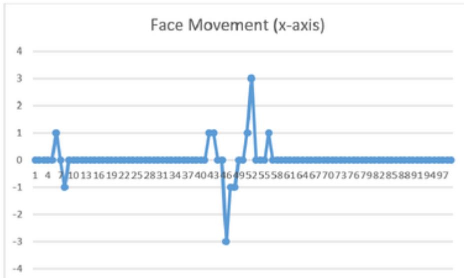


Figure 5 Sound mid-level representation



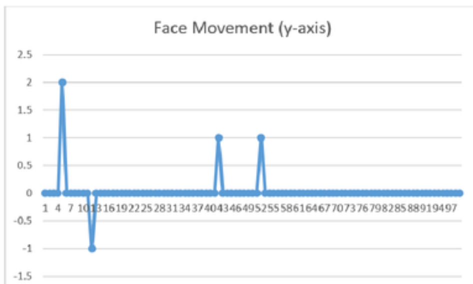
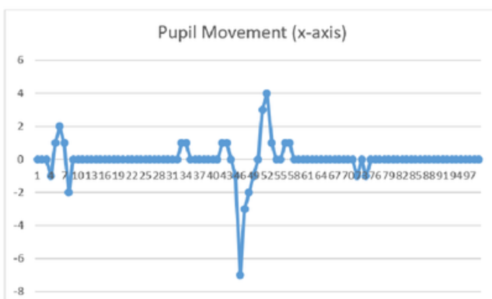


Figure 6 **Face Tracking movement mid-level signal**



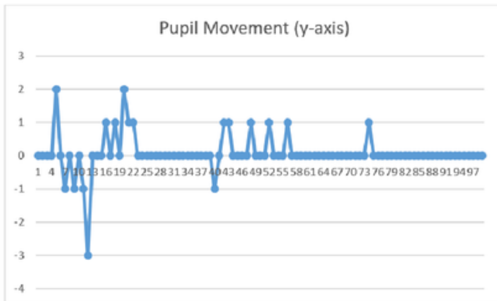


Figure 7 Output Signal

We concatenate the signal into a single vector for each activity. We transform the mid-level signal 5 x 258 into a vector with 1 x 1290 size.

The entire mid-level signal for all the dataset consists of 280 rows. We divide the data into fivefold randomly and validate the classifier using five-fold cross-validation. Table 1 presents the classification metrics of the experiments.

Table 1 Classification result for each fold.

Fold	Accuracy	Recall	Precision	F1 Score
1	0.892857	0.892857	0.902527	0.893131
2	0.892857	0.892857	0.912338	0.892033
3	**0.964286**	**0.964286**	**0.964286**	**0.964286**
4	0.892857	0.892857	0.894081	0.892299
5	0.890909	0.890909	0.899609	0.890475
Average	0.906753	0.906753	0.914568	0.906445
Std Dev	0.023013	0.023013	0.019887	0.023136

According to the experiments, we got the best model on the fold three, where the classifier successfully classifies 96% of the activity based on the available data. The rest of the experiments achieves 89% accuracy. We consider the balance weight between the recall and precision in calculating the F1 score. This is because of the balance portion of cheating and non-cheating class in the dataset.

The conversion of the video data into face tracking and pupil tracking has been done successfully. As shown in figure 5,6 and 7 The downsampling into 12 frames per second still produce a good result where the accuracy achieves well over 90% on average and in the best model (fold 3) 96% of accuracy is achieved. The simple mechanism to transform video and audio data into a simple representation still has enough discriminating power for a binary classifier.

Table 2 The performance of previous research [10]

Cheating Classifier	Recall	Precision	Accuracy
Webcam-only ACD	76.60%	84.60%	83.30%
Multi-modal ACD	82.10%	60.00%	81.10%
PCD	96.40%	94.10%	95.60%
Webcam-only MOOP	78.60%	90.40%	86.10%
Multi-modal MOOP	90.50%	93.80%	92.70%

Comparing to the previous research in [10] shows in table 2 — their research using webcam, EEG and gaze detector in their research. The accuracy of our result is slightly under their PCD and multimodal MOOP. It is reasonable because they use much more complicated tools. If we fairly compare to their results with a single webcam, our proposal overperforms compare [10]. We achieve 90.7% of accuracy while their web-cam only achieved 86.1% — another online proctoring with a double camera and sound recording proposed by [11]. Our result was 3.7% better than their accuracy at 87%.

5. Conclusion

This research presents an audio-video analysis for cheating action recognition to support a mobile exam proctoring. We conclude that the audio and video signal feasible to recognize the cheating and non-cheating activity. The average accuracy, precision, recall, and score at 90.7 ± 2.3%, 91.4 ± 2.3 %, 90.7 ± 1.99 % and 90.6 ± 2.3% respectively. Compared to the state of the art in online cheating detection, we achieve a high recognition rate. Compare to webcam only proctoring in [10] and double camera plus audio in [11], we achieve better accuracy with almost a 4% margin. In addition, our research is designed for a smartphone exam environment with an implementation of standard smartphone accessories without any additional device. This research still needs more data for multiclass classification to identify which action happened. The possibility of using more sensors in the smartphone, such as accelerometer, the gyro will enrich the dataset and expected to improve the accuracy.

References

1. J. E. Seaman, D. Ph, J. Seaman, and D. Ph, "Distance Education State Almanac 2017," 2017.
2. I. E. Allen and J. Seaman, "Grade Change Tracking Online Education in the United States," 2014.
3. A. Fask, F. Englander, and Z. Wang, "Do Online Exams Facilitate Cheating? An Experiment Designed to Separate Possible Cheating from the Effect of the Online Test Taking Environment," *J. Acad. Ethics*, vol. 12, no. 2, pp. 101–112, 2014.
4. C. G. King, R. W. Guyette, and C. Piotrowski, "Online exams and cheating: An empirical analysis of business students' views," *J. Educ. Online*, vol. 6, no. 1, pp. 1–11, 2009.

5. D. L. King and C. J. Case, "E-CHEATING: INCIDENCE AND TRENDS AMONG COLLEGE STUDENTS," *Issues Inf. Syst.*, vol. 15, no. 1, pp. 20–27, 2014.
6. H. Corrigan-Gibbs, N. Gupta, C. Northcutt, E. Cutrell, and W. Thies, "Deterring cheating in online environments," *ACM Trans. Comput. Interact.*, vol. 22, no. 6, 2015.
7. D. Foster and H. Layman, "Online proctoring system compared," *U.S. Pat. Appl. No. 12/723,666.*, pp. 1–12, 2013.
8. M. Kaiiali, A. Ozkaya, H. Altun, H. Haddad, and M. Alier, "Designing a Secure Exam Management System (SEMS) for M-Learning Environments," *IEEE Trans. Learn. Technol.*, vol. 9, no. 3, pp. 258–271, 2016.
9. R. Bawarith, D. Abdullah, D. Anas, and P. Dr., "E-exam Cheating Detection System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 4, pp. 176–181, 2017.
10 X. Li, K. M. Chang, Y. Yuan, and A. Hauptmann, "Massive open online proctor: Protecting the credibility of MOOCs Certificates," in *CSCW 2015 - Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 2015, pp. 1129–1137.
11 Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu, and X. Liu, "Automated Online Exam Proctoring," *IEEE Trans. Multimed.*, vol. 19, no. 7, pp. 1609–1624, Jul. 2017.
12 H. Hung and D. Gatica-Perez, "Estimating cohesion in small groups using audio-visual nonverbal behavior," *IEEE Trans. Multimed.*, vol. 12, no. 6, pp. 563–575, 2010.
13 L. S. Nguyen, D. Frauendorfer, M. S. Mast, and D. Gatica-Perez, "Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior," *IEEE Trans. Multimed.*, vol. 16, no. 4, pp. 1018–1031, 2014.
14 W. Zajdel, J. D. Krijnders, T. Andringa, and D. M. Gavrila, "CASSANDRA: Audio-video sensor fusion for aggression detection," *2007 IEEE Conf. Adv. Video Signal Based Surveillance, AVSS 2007 Proc.*, pp. 200–205, 2007.
15 I. Lefter, G. J. Burghouts, and L. J. M. Rothkrantz, "Automatic audio-visual fusion for aggression detection using meta-information," *Proc. - 2012 IEEE 9th Int. Conf. Adv. Video Signal-Based Surveillance, AVSS 2012*, pp. 19–24, 2012.
16 Z. Wang, J. Yu, and Y. He, "Soccer Video Event Annotation by Synchronization of Attack-Defense Clips and Match Reports with Coarse-Grained Time Information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 5, pp. 1104–1117, 2017.
17 M. H. Kolekar and S. Sengupta, "Bayesian Network-Based Customized Highlight Generation for Broadcast Soccer Videos," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 195–209, 2015.
18 G. R. Cluskey, C. R. Ehlen, and M. H. Raiborn, "Thwarting online exam cheating without proctor supervision," *J. Acad. Bus. Ethics*, 2011.
19 A. Wahid, Y. Sengoku, and M. Mambo, "Toward constructing a secure online examination system," in *ACM IMCOM 2015 - Proceedings*, 2015.
20 N. L. Clarke, P. Dowland, and S. M. Furnell, "e-Invigilator: A Biometric-

Based Supervision System for e-Assessments," in *International Conference on nformation Society (i-Society), 2013*, 2013, p. 5.

21 S. Asha and C. Chellappan, "Authentication of e-learners using multimodal bometric technology," in *IEEE- International Symposium on Biometrics and Security Technologies, ISBAST '08*, 2008.

22 jALALI KHATEREH, "An Automatic method for cheating detection."

23 A.V. Senthil Kumar and M. Rathi, "Keystroke Dynamics: A Behavioral Biometric Model for User Authentication in Online Exams," in *Biometric Authentication in Online Learning Environments*, www.igi-global.com, 2019, pp. 183–207.

24 A. Bal and A. Acharya, "Biometric authentication and tracking system for online examination system," in *2011 International Conference on Recent Trends in Information Systems, ReTIS 2011 - Proceedings*, 2011, pp. 209–213.

25 A. Fayyoumi and A. Zarrad, "Novel Solution Based on Face Recognition to Address Identity Theft and Cheating in Online Examination Systems," *Adv. Internet Things*, vol. 04, no. 02, pp. 5–12, 2014.

26 B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE Trans. Multimed.*, vol. 17, no. 7, pp. 1107–1119, Jul. 2015.

27 Y. Liu, Z. Xie, X. Yuan, J. Chen, and W. Song, "Multi-level structured hybrid forest for joint head detection and pose estimation," *Neurocomputing*, vol. 266, pp. 206–215, 2017.

28 F. Dornaika and B. Raducanu, "Three-Dimensional Face Pose Detection and Tracking Using Monocular Videos: Tool and Application," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 39, no. 4, pp. 935–944, 2009.

29 C. Qing, J. Jiang, and Z. Yang, "Normalized Co-Occurrence Mutual Information for Facial Pose Detection Inside Videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1898–1902, 2010.

30 V. Drouard, R. Horaud, A. Deleforge, S. Ba, and G. Evangelidis, "Robust Head-Pose Estimation Based on Partially-Latent Mixture of Linear Regressions," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1428–1440, 2017.

31 Y. M. Cheung and Q. Peng, "Eye Gaze Tracking with a Web Camera in a Desktop Environment," *IEEE Trans. Human-Machine Syst.*, vol. 45, no. 4, pp. 419–430, 2015.

32 K. Krafka *et al.*, "Eye Tracking for Everyone," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2176–2184, 2016.

33 C. Meng and X. Zhao, "Webcam-Based Eye Movement Analysis Using CNN," *IEEE Access*, vol. 5, pp. 19581–19587, 2017.

34 H. D. Tran and H. Li, "Sound event recognition with probabilistic distance SVMs," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 6, pp. 1556–1568, 2011.

35 C. Y. Wang, J. C. Wang, A. Santoso, C. C. Chiang, and C. H. Wu, "Sound Event Recognition Using Auditory-Receptive-Field Binary Pattern and Hierarchical-Diving Deep Belief Network," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 8, pp. 1336–1351, 2018.

36 N. Dalal and W. Triggs, “Histograms of Oriented Gradients for Human Detection,” *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05*, vol. 1, no. 3, pp. 886–893, 2004.

37 V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

pre-submission

ORIGINALITY REPORT

10%	7%	4%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Nazarbayev University Student Paper	1%
2	Submitted to Universiti Teknologi MARA Student Paper	1%
3	Submitted to Taylor’s Education Group Student Paper	1%
4	Xuanchong Li, Kai-min Chang, Yueran Yuan, Alexander Hauptmann. "Massive Open Online Proctor", Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15, 2015 Publication	1%
5	itechprosolutions.in Internet Source	1%
6	arxiv.org Internet Source	1%
7	jestec.taylors.edu.my Internet Source	<1%

8	manualzz.com Internet Source	<1 %
9	Submitted to University of Westminster Student Paper	<1 %
10	Submitted to ITM University Student Paper	<1 %
11	Submitted to University of Stirling Student Paper	<1 %
12	Submitted to De La Salle University - Manila Student Paper	<1 %
13	Submitted to Auckland University of Technology Student Paper	<1 %
14	Yousef Atoum, Liping Chen, Alex X. Liu, Stephen D. H. Hsu, Xiaoming Liu. "Automated Online Exam Proctoring", IEEE Transactions on Multimedia, 2017 Publication	<1 %
15	Submitted to Asian Institute of Technology Student Paper	<1 %
16	Submitted to University of Sheffield Student Paper	<1 %
17	journalofbigdata.springeropen.com Internet Source	<1 %
18	M.-Y. Chow, R.N. Sharpe, J.C. Hung. "On the	

application and design of artificial neural networks for motor fault detection. I", IEEE Transactions on Industrial Electronics, 1993
Publication <1%

19 dspace.lboro.ac.uk
Internet Source <1%

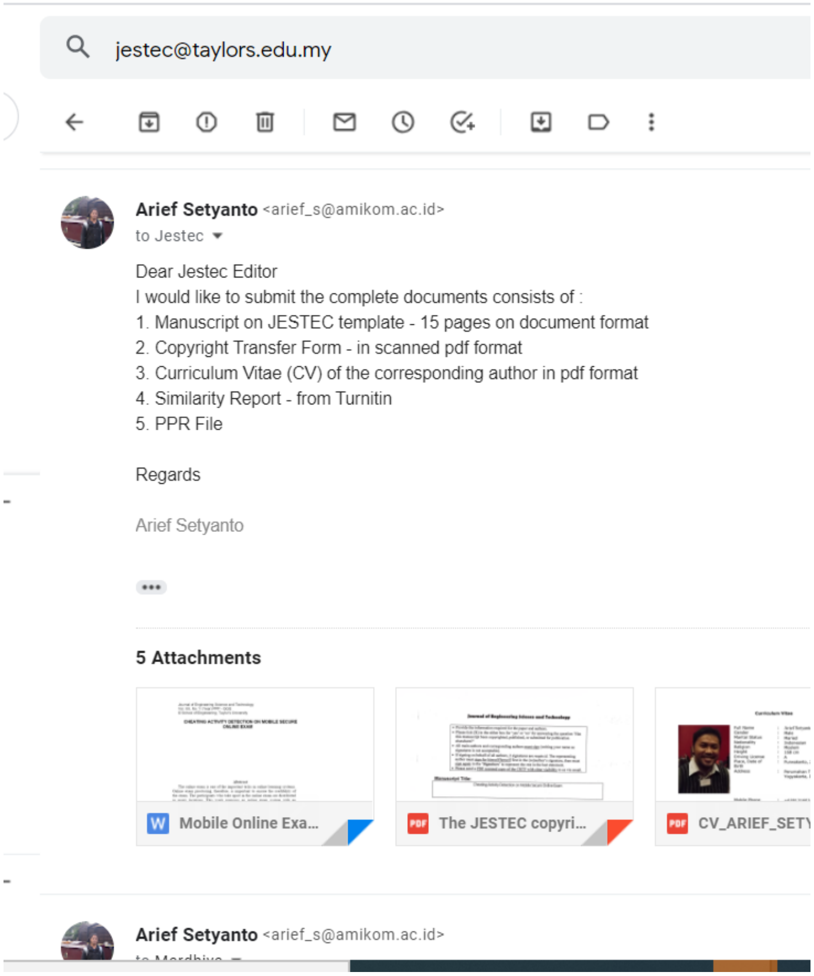
20 Xiaolong Zhou, Haibin Cai, Youfu Li, Honghai Liu. "Two-eye model-based gaze estimation from a Kinect sensor", 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017
Publication <1%

21 pure.qub.ac.uk
Internet Source <1%

22 Submitted to School of Business and Management ITB
Student Paper <1%

23 Abdourahman Houssein Ahmed, Kidiyo Kpalma, Abdoulkader Osman Guedi. "Human Detection Using HOG-SVM, Mixture of Gaussian and Background Contours Subtraction", 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2017
Publication <1%

1. Pengiriman paper pertama pada tanggal 12 Desember 2019


jestec@taylors.edu.my

Arief Setyanto <arief_s@amikom.ac.id>
to Jestec

Dear Jestec Editor
I would like to submit the complete documents consists of :
1. Manuscript on JESTEC template - 15 pages on document format
2. Copyright Transfer Form - in scanned pdf format
3. Curriculum Vitae (CV) of the corresponding author in pdf format
4. Similarity Report - from Turnitin
5. PPR File

Regards

Arief Setyanto

5 Attachments

Mobile Online Exa...

The JESTEC copyri...

CV_ARIEF_SETY

Arief Setyanto <arief_s@amikom.ac.id>

2. Paper lengkap dan dimulai proses Review round 1

Submission of a Manuscript (EE19229) / First Round of the Review Process Inbox ×

Jestec <Jestec@taylors.edu.my>
to me

Wed, Dec 18, 2019, 3:39 PM

Dear Author

Thank you for submitting your research paper to the Journal of Engineering Science and Technology (JESTEC)

Kindly note that we have received the paper entitled

CHEATING ACTIVITY DETECTION ON MOBILE SECURE ONLINE EXAM

Your paper ID is **EE19229** *(Please quote the above manuscript ID in all future correspondence with us.)*

Soon we will initiate the first round of the review process.

Please be reminded that upon the full acceptance of your paper, publication fee in amount of USD300 must be paid before the article is published in the journal website.

Best regards

JESTEC Editor
http://jestec.taylors.edu.my

3. Paper selesai di review pada putaran pertama dinyatakan diterima dengan revisi minor dan revisi mayor – oleh para reviewer pada 9 Februari 2020, komentar reviewer terlampir

Paper ID EE19229 /Review of a paper, First Round Result/ Inbox ×

Jestec <Jestec@taylors.edu.my>
to me

Dear Author

The first round of the review process has been completed.

I am glad to advise that your paper has been conditionally accepted for publication with

☐ No modification ☑ Minor corrections ☑ Major modification.

Attached herewith, please find

☐ 1 ☑ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 reviewers' reports.

Please notice the following:

1. Address all the concerns/recommendations of the reviewers
2. All amendments made are to be highlighted in red color in the revised paper.
3. Send an outlining following the instructions in the attached file on how did you address each reviewers' concern/recommendations.
4. In order to complete the review process on time, we highly appreciate it if we can receive the revised paper within **three weeks** from today.
5. In case that you will need more time to complete the revision, please indicate how much time you need via an email so we can get the approval from the Editorial Board.

Please note that the final acceptance of the paper depends on the final decision of the Review Panel and after the paper successfully passed all the review rounds.

Best Regards

Journal of Engineering Science and Technology (JESTEC)

REVIEW FORM

Title of paper:	Cheating Activity Detection on Mobile Secure Online Exam

For sections A & B, please tick a number from 0 to 5, where 0 = strongly disagree and 5 = strongly agree.

A. Technical aspects

1. The paper is within the scope of the Journal.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5
2. The paper is original.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5
3. The paper is free of technical errors.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5

B. Communications aspects

1. The paper is clearly readable.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5
2. The figures are clear & do clearly convey the intended message.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5
3. The length of the paper is appropriate.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5

C. Comments to the authors (You may use another sheet of paper.)

There are much Error spelling in this paper, such as Section 2., Section 3. … equation (1), (2), (3) …

The author should write more succinctly.

Any requirements to obtain test results are presented in this paper

This paper presents an audio-video analysis for cheating action recognition to support a mobile exam proctoring, only using standard smartphone accessories.

The author did not give the algorithms of the Multimodal Action Detection – It is main core-content of the paper.

D. Recommendation (Tick one)

1. Accepted without modifications. ☐
2. Accepted with minor corrections. ☐
3. Accepted with major modification. ☑
4. Rejected. ☐

E. Comments to the editors (These comments will not be sent to the authors)

Journal of Engineering Science and Technology (JESTEC)

REVIEW FORM

Title of paper: Cheating Activity Detection on Mobile Secure Online Exam

For sections A & B, please tick a number from 0 to 5, where 0 = strongly disagree and 5 = strongly agree.

A. Technical aspects

1. The paper is within the scope of the Journal.	☐ 0	☐ 1	☐ 2	☐ 3	✓ 4	☐
2. The paper is original.	☐ 0	☐ 1	☐ 2	☐ 3	✓ 4	☐
3. The paper is free of technical errors.	☐ 0	☐ 1	☐ 2	✓ 3	☐ 4	☐

B. Communications aspects

1. The paper is clearly readable.	☐ 0	☐ 1	☐ 2	✓ 3	☐ 4	☐
2. The figures are clear & do clearly convey the intended message.	☐ 0	☐ 1	☐ 2	✓ 3	☐ 4	☐
3. The length of the paper is appropriate.	☐ 0	☐ 1	☐ 2	☐ 3	✓ 4	☐

C. Comments to the authors (You may use another sheet of paper.)

1. There are several grammatical mistakes that must be fixed.
2. Every variable or value must be equipped with unit.
3. Problem statement must be declared clearly before research purpose.
4. Research framework as it is shown in Figure 1 should be explained.
5. Imperative sentences should be avoided.
6. System limitations should be declared.
7. Graphics or charts should be equipped with unit.
8. Testing scenario and environment should be described clearly.
9. System diagram should be added so that in which parts the work implemented is clear.
10. Reasoning in perceptron design should be added. Why there are 20 hidden layers and 1290 nodes in input layer.

D. Recommendation (Tick one)

1. Accepted without modifications. ☐

2. Accepted with minor corrections. ✓

3. Accepted with major modification. ☐

4. Rejected. ☐

E. Comments to the editors (These comments will not be sent to the authors)

4 saya meminta perpanjangan masa koreksi 2 minggu karena belum selesai (28 Februari) di aprove pada 29 Februari perpanjangan masa koreksi

15 of 26

Arief Setyanto <arief_s@amikom.ac.id>
to Jestec

Feb 28, 2020, 3:40 PM

Dear executive editor,
Thank you very much for the feedback for the paper that I submitted.
Currently, I am still revising the paper. I still working with some issues to address the reviewer's comments. I would like to ask for an extension for two weeks but if I can make it earlier I will submit it as early as possible.

Regards

Arief Setyanto

Jestec <Jestec@taylors.edu.my>
to me

Feb 29, 2020, 7:04 AM

accepted

Best Regards

JESTEC Editor
http://jestec.taylors.edu.my

5. Tanggal 11 Maret saya mengirim hasil koreksi paper, Tanggal 29 Maret di balas dokumen yang dikirimkan masih terdapat personalisasi sehingga perlu dikirim ulang dengan menghilangkan identitas penulis karena prosesnya merupakan double blind review

Arief Setyanto <arief_s@amikom.ac.id> Mar 11, 2020, 2:46 PM

to Jestec

Dear Chief Editor

I would like to submit the correction of my paper to address to reviewer's comments.

Regards

Arief Setyanto

2 Attachments

Review Report_FM...

Manuscript_Correc...

Jestec <Jestec@taylors.edu.my> Mar 29, 2020, 11:17 AM

to me

Dear Author

Thank you for your email and submitting your revised paper.

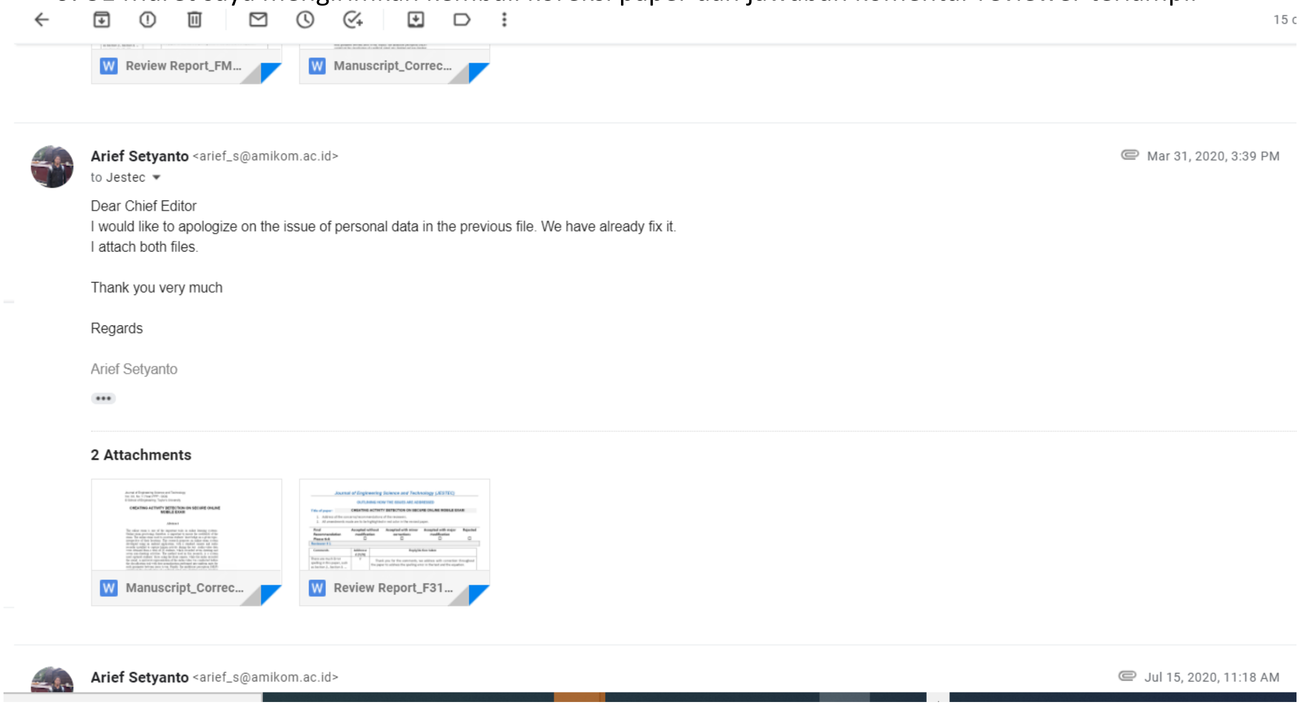
However, there is an issue you must fix before we start the 2nd round of the review process.

When you do the revision of your paper, please remove anything which reveals your identity. This is because our review process is double-blind peer review.

Kindly, check and resend the documents.

Best Regards

6. 31 Maret saya mengirimkan kembali koreksi paper dan jawaban komentar reviewer terlampir



OUTLINING HOW THE ISSUES ARE ADDRESSED

Title of paper:	CHEATING ACTIVITY DETECTION ON SECURE ONLINE MOBILE EXAM

1. Address all the concerns/recommendations of the reviewers.
2. All amendments made are to be highlighted in red color in the revised paper.

Reviewer # 1

Final Recommendation	Accepted without modification	Accepted with minor corrections	Accepted with major modification	Rejected
Please tick	☐	☐	☐	☐

Comments	Addressed (Y/N)	Reply/Action taken
There are much Error spelling in this paper, such as Section 2., Section 3. … equation (1), (2), (3) …	Y	Thank you for the comments, we address with correction throughout the paper to address the spelling error in the text and the equation. The vector was then normalized to acquire L2-Norm, L1-Norm and L1-Sqrt in equations (1), (2), and (3). $$L2 - Norm : f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (1)$$ $$L1 - Norm : f = \frac{v}{\|v\|_1 + e} \quad (2)$$ $$L1 - Sqrt : f = \sqrt{\frac{v}{(\|v\|_1 + e)}} \quad (3)$$
The author should write more succinctly.	Y	Already corrected by professional Proof Reader

<table>
<tr>
<td>Any requirements to obtain test results are presented in this paper</td>
<td>Y</td>
<td>
Thank you for the comments, we add specific requirement during the experiments and also the specification of smartphone used in our research.

Page 10 , Last Paragraph :

This experimental research was carried out in a quiet testing environment away from the crowd to benefit the exam room. The research location was a 3x3 meter room which had bright lighting, tables, and chairs. Each participants sitting and working on exam questions all had a smart phone. Their profiles were quite diverse, with male and female, wearing/not wearing glasses, or veil. The data collection was carried out simultaneously among four participants in meeting rooms.
Page 11 :
Table 1. SmartPhone Spesification
<table>
<tr><th>Specs</th><th>Samsung A2 Core</th><th>Samsung A20</th><th>Vivio Y15</th></tr>
<tr><td>CPU</td><td>1.6 GHz Octacore</td><td>2x 1.6 GHz Octacore</td><td>2.0 GHz Octacore</td></tr>
<tr><td>RAM</td><td>1 GB</td><td>3 GB</td><td>4 GB</td></tr>
<tr><td>Display</td><td>5” 540 x 960 px</td><td>6.4” 720 x 1560 px</td><td>6.35”720 x 1544 px</td></tr>
<tr><td>Front Cam</td><td>5 MP</td><td>8 MP</td><td>16 MP</td></tr>
<tr><td>Back Cam</td><td>5 MP</td><td>13 Mp</td><td>13 MP</td></tr>
<tr><td>OS</td><td>Android 8.0</td><td>Android 9.0</td><td>Android 9.0</td></tr>
</table>
</td>
</tr>
<tr>
<td>This paper presents an audio-video analysis for cheating action recognition to support a mobile exam proctoring, only using standard smartphone accessories.</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>The author did not give the</td>
<td>Y</td>
<td>Thank you for this comment, we address the comments by providing additional explanation of our multimodal data and the figure to draw our approach in.</td>
</tr>
</table>

<table>
<tr>
<td>algorithms of the Multimodal Action Detection – It is main core-content of the paper.</td>
<td></td>
<td>

Page 5-6

The video data was extracted into two time series features namely face and pupil movements. Two modalities were obtained in the data collection tasks. The video was explored into the movement using motion vector calculation and *lucas-kanade* methods. Before the tracking started, the face and eye recognition was carried out. This produced a three time series data which are face tracking, eye tracking and audio data. In order to enable the modality fusion for all the data, normalization was carried out in a range of -10 to 10. The periodicity of the data sampled was in 12 data point per second. During the data collection, the original duration were not precisely uniform, while pre-processing and was standardized into 21.5 seconds for each captured audio video. The face tracking produced two variables *x* and *y, which was* considered as 1 unit for every 25 pixel movement. Furthermore, the pupil tracking yielded two x and y variables in pixel. Sound on the other hand produced one variable and was sampled in exactly the same frequency at 12 data per second. During the feature extraction, a zero padding is implemented to ensure standard length of each feature for each audio video data. Figure 3 shows the multimodal data pre-processing before a multi-layer perceptron was applied.

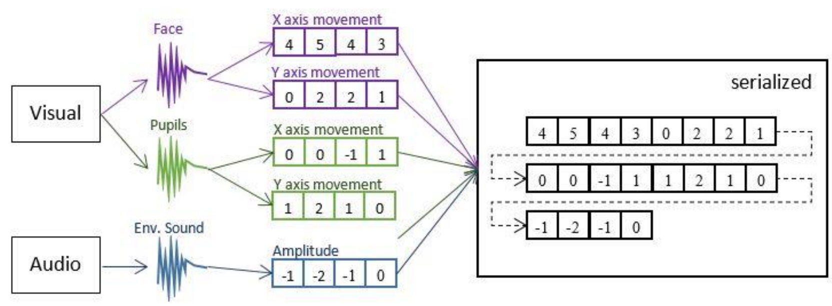


Fig 1. Multi Modal Feature Extraction and Fusion

</td>
</tr>
</table>

(Please add more rows if needed)

Reviewer # 2				
Final Recommendation	**Accepted without modification**	**Accepted with minor corrections**	**Accepted with major modification**	**Rejected**

Please tick ☐ ☐ ☐ ☐

Comments	Addressed (Y/N)	Reply/Action taken
There are several grammatical mistakes that must be fixed.	Y	Thank you for your comments, The grammatical error Already corrected by professional Proof Reader
Every variable or value must be equipped with unit.	Y	Page 8 - 3.2.2 Face Tracking, Paragraph 2 : The value was set between -10 … 10, were recorded as 25 pixel for 1 unit. Page 8 – 3.2.3 Eye-tracking , Paragraph 2 : in number of pixel displacement
Problem statement must be declared clearly before research purpose.	Y	Thankyou for your comments, we underline the aims of the research in the **Pages 2, Paragraph 5 :** ......Moreover, in some research such as in [10], [11], special devices such as additional camera, electroencephalogram(EEG) and gaze tracker need to be installed. This research aims to simplify the online exam proctoring task with standard devices only for easier implementation. This research investigates minimum set of data acquired by the standard camera and audio recorder of low-end smartphone to recognize pre-defined cheating activities using artificial neural network (ANN).
Research framework as it is shown in Figure 1 should be explained.	Y	Thank you for the comments, we add the explanation of the research framework in **Page 4 (paragraph 3) and Page 5 (paragraph 1):** Online exam cheating activities is defined into seven types. They consist of opening/reading a book, sending/receiving messages through a smart phone, accepting help from people, receiving clues through paper, using the computer to browse testing material, and asking for answers from someone. In order to build a dataset, these activities were recorded in audio and video through standard smart phones. An additional video of non-cheating participants was also recorded with a standardized recording duration of 21.5 seconds in 12 frame per second (fps). A feature extraction was carried out of all the positive and negative cheating videos. The visual and motion data were extracted into face and pupil movement, while the audio was sampled into 12 data per

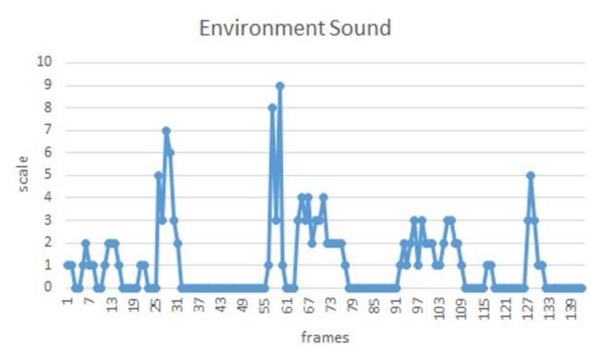
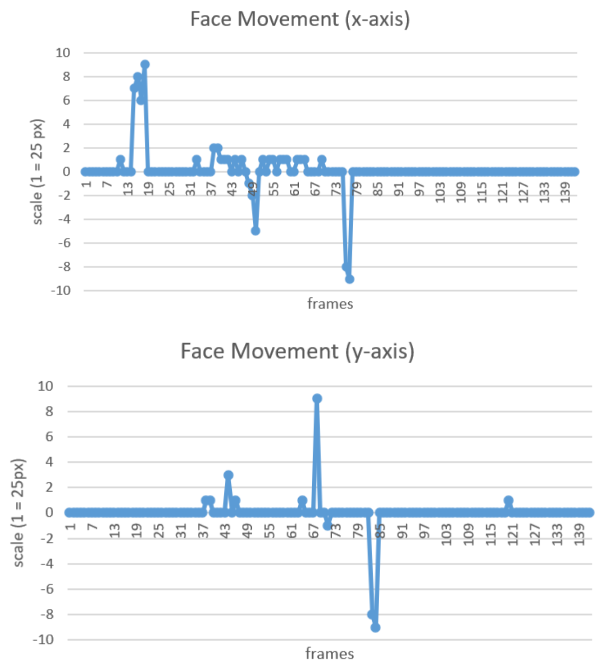
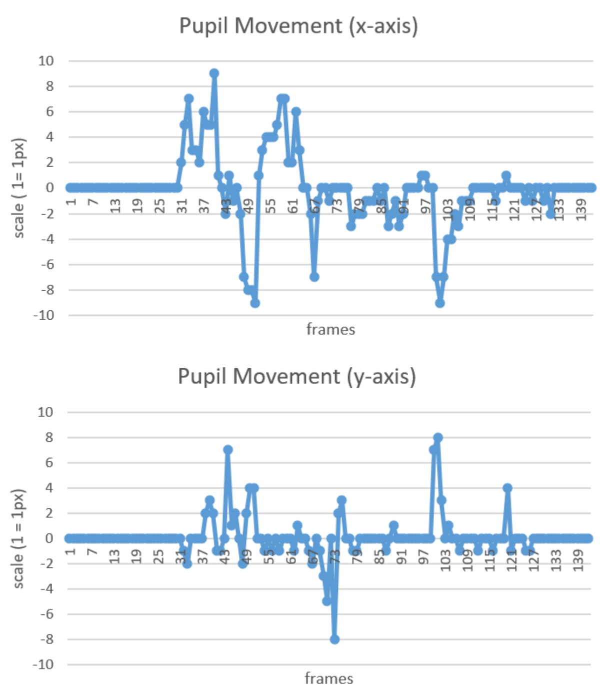
		second. The three modalities were fused before entering the network in artificial neural network (ANN), thereby, leading to a total of 1290 data length for each recorded activity. Multi perceptron (MLP), training and evaluation were carried out by using the available data.
Imperative sentences should be avoided.	Y	Already corrected by professional Proof Reader
System limitations should be declared.	Y	Thank you for your comments, we address the comment by a new paragraph in **page 15, last paragraph :** This research is limited by the pre-defined cheating action under specified requirements in the experimental setting. However, despite the unavailability of empirical data, and the specified resolution of the video, the result might be degraded. Secondly, other types of cheating are likely to be beyond the simulated activity pre-set in this research, with undefined outcomes. The lighting in this study was set in stable and visible condition, therefore, a darker situation is likely to lead to a significant decrease on recognition quality. The noisy sound is also a challenge as the dataset was taken in a controlled condition where the noise is at a minimum level. Another potential problem of this research is privacy issue, due to the need of recording the video and audio during the exam. In the future, a signal transformation on the recording device would be an alternative to overcome the privacy issue. Due to the limited possibility of cheating/non-cheating actions detected, the exact activity was not precisely recognized such as reading a book, using mobile phone etc. These limitations are a guide for future research.
Graphics or charts should be equipped with unit.	Y	We add the unit in each graphics. Especially for sound we do not have the units, the number 0 – 10 is a normalized value relative to maximum amplitude. **Page 12 – 13** 

		Fig 2. Sound Signal Feature  **Fig 3. Face tracking movement**  **Fig 4. Eye Tracking Signal**

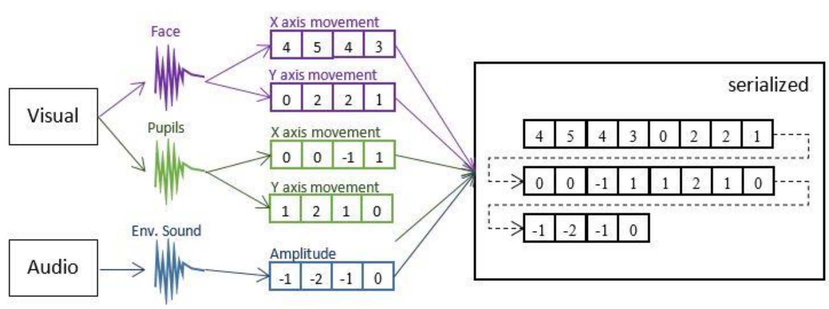
<table>
<tr>
<td>Testing scenario and environment should be described clearly.</td>
<td>Y</td>
<td>
Thankyou for your comments, We address by an additional explanation in
Page 10, 3.4. Experimental Design
This experimental research was carried out in a quiet testing environment away from the crowd to benefit the exam room. The research location was a 3x3 meter room which had bright lighting, tables, and chairs. Each participants sitting and working on exam questions all had a smart phone. Their profiles were quite diverse, with male and female, wearing/not wearing glasses, or veil. The data collection was carried out simultaneously among four participants in meeting rooms.
....
....
Three low end smart phones with the specification below were used.
Table 2. SmartPhone Spesification
<table>
<tr><th>Specs</th><th>Samsung A2 Core</th><th>Samsung A20</th><th>Vivio Y15</th></tr>
<tr><td>CPU</td><td>1.6 GHz Octacore</td><td>2x1.6 GHz Octacore</td><td>2.0 GHz Octacore</td></tr>
<tr><td>RAM</td><td>1 GB</td><td>3 GB</td><td>4 GB</td></tr>
<tr><td>Display</td><td>5” 540 x 960 px</td><td>6.4” 720 x 1560 px</td><td>6.35”720 x 1544 px</td></tr>
<tr><td>Front Cam</td><td>5 MP</td><td>8 MP</td><td>16 MP</td></tr>
<tr><td>Back Cam</td><td>5 MP</td><td>13 Mp</td><td>13 MP</td></tr>
<tr><td>OS</td><td>Android 8.0</td><td>Android 9.0</td><td>Android 9.0</td></tr>
</table>
</td>
</tr>
<tr>
<td>System diagram should be added so that in which parts the work implemented is clear.</td>
<td>Y</td>
<td>
Page 6, Paragraph 1 :
Figure 3 shows the multimodal data pre-processing before a multi-layer perceptron was applied.


Fig 5. Multi Modal Feature Extraction and Fusion
</td>
</tr>
<tr>
<td>Reasoning in perceptron design should be</td>
<td>Y</td>
<td>
Thankyou for you comments, provide an experimental result, a varied hidden layer applied and 20 layer produce the best result, therefore, we choose 20 hidden layers
Page 10, Paragraph 2 :
....the network were set with 10, 20, 30 and 40 hidden layers
</td>
</tr>
</table>

<table>
<tr>
<td>added. Why there are 20 hidden layers and 1290 nodes in input layer.</td>
<td></td>
<td>
Page 14, Paragraph 2 :

.....Table 2 shows the average, accuracy, precision and f1score of the classification result using MLP with varied Hidden Layer. ....

Table 3. Average Accuracy, Recall, Precision and F1 Score for different Hidden Layer
<table>
<tr><th>Hidden Layer</th><th>Accuracy</th><th>Recall</th><th>Precision</th><th>F1 Score</th></tr>
<tr><td>10</td><td>0.910325</td><td>0.910325</td><td>0.911421</td><td>0.910325</td></tr>
<tr><td>20</td><td>0.917338</td><td>0.917338</td><td>0.929832</td><td>0.916798</td></tr>
<tr><td>30</td><td>0.903312</td><td>0.903312</td><td>0.909436</td><td>0.902822</td></tr>
<tr><td>40</td><td>0.903182</td><td>0.903182</td><td>0.912396</td><td>0.902383</td></tr>
<tr><td>50</td><td>0.913896</td><td>0.913896</td><td>0.91895</td><td>0.913387</td></tr>
</table>
</td>
</tr>
</table>

7. Progress review rounde ke 2 dimulai 5 April 2020

jestec@taylors.edu.my

Active

16 of 26

Paper ID EE19229 /A progress of Review Process/ Inbox ×

Jestec <Jestec@taylors.edu.my> Sun, Apr 5, 2020, 2:43 PM

to me

Dear Author

This email is to confirm that your paper is currently undergoing the

☐ 1st ☑ 2nd ☐ 3rd round of the review process.

Thank you for your patience.

Best regards

JESTEC Editor

http://jestec.taylors.edu.my

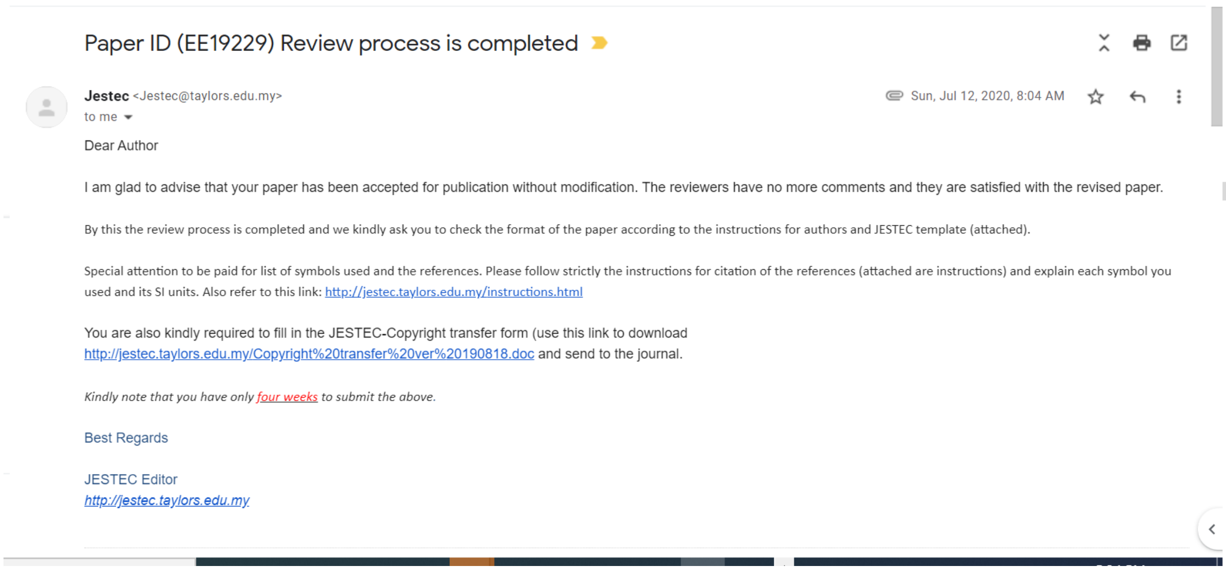
This message (including any attachments) is intended only for the use of the individual or entity to which it is addressed and may contain information that is non-public, proprietary, privileged, confidential, and ex from disclosure under applicable law or may constitute as attorney work product. If you are not the intended recipient, you are hereby notified that any use, dissemination, distribution, or copying of this communi is strictly prohibited. If you have received this communication in error, notify us immediately by telephone and (i) destroy this message if a facsimile or (ii) delete this message immediately if this is an electronic communication.

Arief Setyanto <arief_s@amikom.ac.id> Sun, Apr 5, 2020, 4:37 PM

to Jestec

Thank you for the update.

8. Review Ronde ke 2 selesai dan paper dinyatakan diterima tanpa perlu dilakukan koreksi lagi



Paper ID (EE19229) Review process is completed

Jestec <Jestec@taylors.edu.my> Sun, Jul 12, 2020, 8:04 AM
to me

Dear Author

I am glad to advise that your paper has been accepted for publication without modification. The reviewers have no more comments and they are satisfied with the revised paper.

By this the review process is completed and we kindly ask you to check the format of the paper according to the instructions for authors and JESTEC template (attached).

Special attention to be paid for list of symbols used and the references. Please follow strictly the instructions for citation of the references (attached are instructions) and explain each symbol you used and its SI units. Also refer to this link: http://jestec.taylors.edu.my/instructions.html

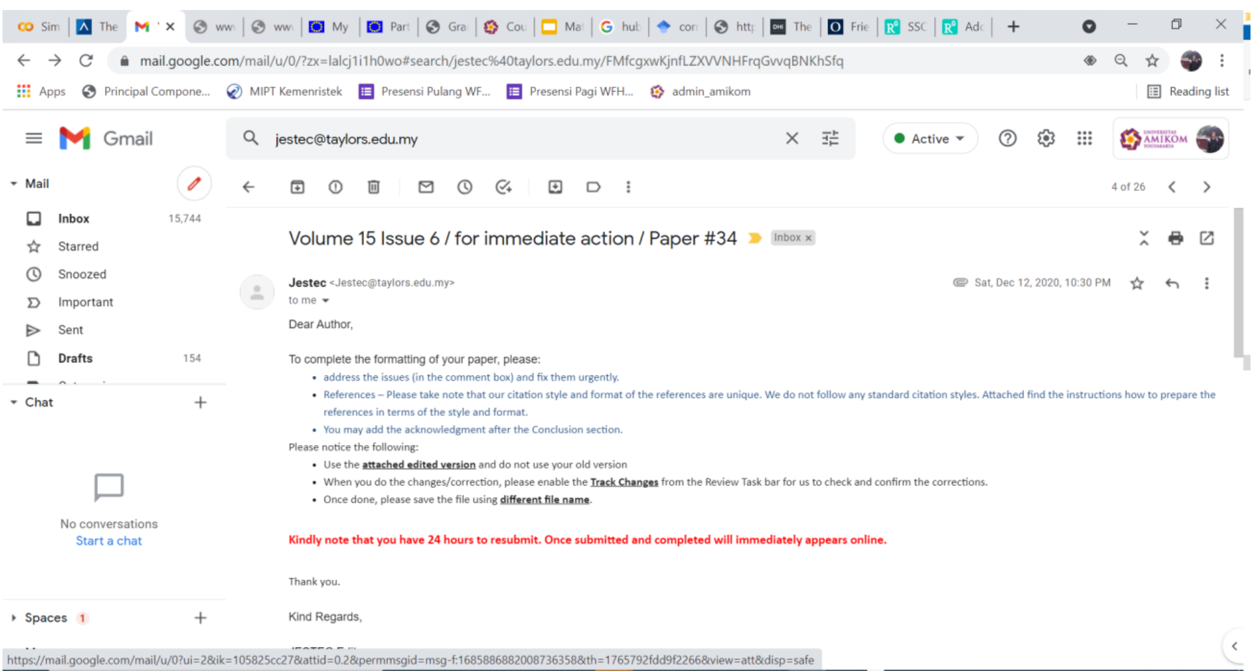
You are also kindly required to fill in the JESTEC-Copyright transfer form (use this link to download http://jestec.taylors.edu.my/Copyright%20transfer%20ver%20190818.doc and send to the journal.

Kindly note that you have only four weeks to submit the above.

Best Regards

JESTEC Editor
http://jestec.taylors.edu.my

9. Paper terbit pada Desember 2020



Volume 15 Issue 6 / for immediate action / Paper #34 Inbox

Jestec <Jestec@taylors.edu.my> Sat, Dec 12, 2020, 10:30 PM
to me

Dear Author,

To complete the formatting of your paper, please:

- address the issues (in the comment box) and fix them urgently.
- References – Please take note that our citation style and format of the references are unique. We do not follow any standard citation styles. Attached find the instructions how to prepare the references in terms of the style and format.
- You may add the acknowledgment after the Conclusion section.

Please notice the following:

- Use the **attached edited version** and do not use your old version
- When you do the changes/correction, please enable the **Track Changes** from the Review Task bar for us to check and confirm the corrections.
- Once done, please save the file using **different file name**.

Kindly note that you have 24 hours to resubmit. Once submitted and completed will immediately appears online.

Thank you.

Kind Regards,

mail.google.com/mail/u/0/?zx=lalcj1i1h0wo#search/jestec%40taylors.edu.my/FMfcgxwKjngSpTtqrnCnPnQLQWsHRKss

Your paper is now available online in Volume 15 Issue 6/Check and Proofread/

Jestec <Jestec@taylors.edu.my> Sun, Dec 13, 2020, 10:30 AM

Dear Author,

I am glad to inform you that your papers are currently available online in Issue 15 Volume 6 (December 2020).
Please access your paper using this link http://jestec.taylors.edu.my/V15Issue6.htm for final proofreading and checking if there are any mistakes/changes.

Please reply before or latest by Thursday, 10/12/2020, 12 p.m. (Malaysia local time).
After this date, no changes will be allowed.

Thank you for publishing with JESTEC.

Regards,

JESTEC Editor,
http://Jestec.taylors.edu.my

Arief Setyanto <arief_s@amikom.ac.id> Tue, Dec 22, 2020, 11:41 AM