Engineering and Technology Journal
Journal homepage: engtechjournal.org

Detection of confusion behavior using a facial expression based on different classification algorithms

Fatima I. Yasser [a]*, Bassam H. Abd [b], Saad M. Abbas [c]

[a] Department of Electrical Engineering University of Technology-Iraq, Baghdad, Iraq, 319310@student.uotechnology.edu.iq
[b] Department of Electrical Engineering University of Technology-Iraq, Baghdad, Iraq, 30022@uotechnology.edu.iq
[c] Department of Electrical Engineering University of Technology-Iraq, Baghdad, Iraq, saad.m.abbas@uotechnology.edu.iq
*Corresponding author.

Submitted: 25/06/2020 Accepted: 28/07/2020 Published: 25/02/2021

K E Y W O R D S

Confusion detection, FACS, VG-RAM, SVM, Logistic Regression, Quadratic Discriminant.

A B S T R A C T

Confusion detection systems (CDSs) that need Noninvasive, mobile, and cost-effective methods use facial expressions as a technique to detect confusion. In previous works, the technology that the system used represents a major gap between this proposed CDS and other systems. This CDS depends on the Facial Action Coding System (FACS) that is used to extract facial features. The FACS shows the motion of the facial muscles represented by Action Units (AUs); the movement is represented with one facial muscle or more. Seven AUs are used as possible markers for detecting confusion that has been implemented in the form of a single vector of facial action; the AUs that have been used in this work are AUs 4, 5, 6, 7, 10, 12, and 23. The database used to calculate the performance of the proposed CDS is gathered from 120 participants (91males, 29 females), between the ages of 18-45. Four types of classification algorithms are used as individuals; these classifiers are (VG-RAM), (SVM), Logistic Regression and Quadratic Discriminant classifiers. The best success rate was found when using Logistic Regression and Quadratic Discriminant. This work introduces different classification techniques to detect confusion by collecting an actual database that can be used to evaluate the performance for every CDS employing facial expressions and selecting appropriate facial features.

How to cite this article: F. I. Yasser, B. H. Abd, and S. M. Abbas, "Detection of confusion behavior using a facial expression based on different classification algorithms," Engineering and Technology Journal, Vol. 39, Part A, No. 02, pp. 316-325, 2021.
DOI: https://doi.org/10.30684/etj.v39i2A.1750

1. INTRODUCTION

Confusion detection is essential in current days; many studies reveal the vital role that confusion can play in learning, as found in Massive Open Online Course (MOOC). Additionally, in any type of

everyday assignment that required reasoning, for example, feeling confused while driving or in any Situation Awareness (SA) environment also in detection confusion in healthcare applications.

Electromyography (EMG) and Electroencephalogram (EEG) systems have been utilized in confusion detection techniques. These are considered invasive techniques to detect confusion since they require connecting sensors to the participants' bodies during the experiments. EMG works on sensing the electrical activity of the detectable facial musculature related to confusion [1]. Also, employing an EEG system to detect confusion where the assumption considered that confused people EEG signals would diverge from normal state signals [2]. The accuracy for both systems considered to be low compared to the non-invasive systems. Another reason which makes the non-invasive techniques more reliable to detect confusion.

There are other indications of confusion discussed in the previous work, which are eye movement, EEG data, EMG data, facial expression, eye gaze and the way expressing the Language and Discourse Analysis for a learner. The latter one was proposed by Atapattu [3] by adapting the confusion classification technique in MOOC to identify which aspects impact the overall learning process. The dataset contains approximately 30,000 anonymized forum posts. While detecting user's confusion during visualization processing was proposed by Conati [4]. Another work represented by Postma [5] where confusion detection in healthcare delivery applications on the Internet was introduced to help elderly patients using automatic facial recognition systems. Two other works to detect confusion based on facial expression proposed by Grafsgaard [6, 7]. The earlier studies that were represented to clarify confusion back to the date 1987 introduced by De Wit [8]. The hypotheses suggested by D'Mello [9] that confusion appears when their contradictions and conflicts with the informational stream, so the confusion that has been triggered can be effective in learning if it is properly controlled by the learner, or whether there are enough scaffolds which allow learners to deal with their confusion. These hypotheses are partially supported by the achieved results. Recently, other studies give the efficiency of employing facial expression extraction technique for confusion detection according to the number of participants. For instance, a recent study by Borges [10] and Shi [11] both yielded effective results to recognize confusion.

Most CDSs used invasive techniques mentioned earlier may give inaccurate information because it requires connecting sensors to the participant's body, which makes the participant confused/anxious during the experiment. While this potentially can give, false misreading sensor signals leading to confusion detection miss predication. In addition to the contained area in which the experiment was carried. The major purpose of this work is to build an independent CDS without physical contact. Therefore, to design efficient CDS, suitable and reliable visual features should be selected ~~from~~ the collected database. This proposed CDS is an autonomous system and can operate in unconstrained environments.

2. MACHINE-BASED CONFUSION DETECTION SYSTEM

Confusion recognition is an important step in designing effective teaching strategies and interventions. This automatic recognition method, which is based on facial expressions, has been applied and achieved high accuracy to detect confusion [11]. This system consists of three stages, the first one video capturing and pre-processing, which arranges the captured video for the next stage and its responsible for facial identification and finding any important points (landmarks) on the participant's face. The second stage is used to extract features from the captured videos and use them as a potential marker for confusion. The final stage is Decision maker, in which one of, the suggested classifier that reveals the participant state. The general block diagram of confusion detection systems CDS based on facial expressions is shown in Figure 1. The following section discusses these stages in detail.

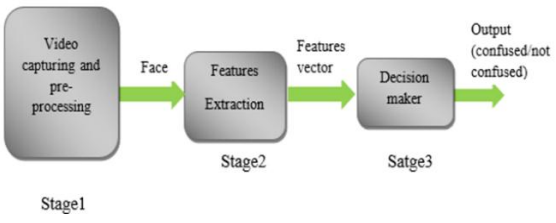

Figure 1: The general block diagram of confusion detection systems CDS.

I. Stage 1

Capturing the video and Pre-Processing it represents the first stage of the proposed system, which consists of three steps (Video Recording and Editing, Face Detection, and Facial Landmark Detection and Tracking). The videos are recorded using a digital camera, then editing the video is performed to recognize and obtain only the needed parts of these videos and to remove undesirable parts. After this step, the viola jones technique is used [12] in such a way that the process of image sequence for facial detection is carried out for the dataset collected. The VJ technique used grayscales of the image to extract feature pixels by using the feature blocks technique, which is a Haar-like feature block set based on the AdaBoost classifier, where five types of Haar-like features are used. The face detection step gives a bounding box around the face. This bounding box leads to initial landmark locations that given initial shape parameters. The CLNF fitting [13] can start based on initial shape parameters. The fitting algorithm finds the optimal shape parameters. If the CLNF fitting was succeeded in locating feature points or otherwise, it would notify the tracker to reinitialize depend on a face detector.

II. Stage 2

Feature Extraction represents the second stage; in this method, detection of AUs depends on 2 kinds of features, which are (appearance and geometry). The AUs movement represented one facial muscle or more, that used as possible markers for detecting confusion. Many techniques based on facial expressions extraction have been utilized for confusion detection, which used a variety of types of AU as an indicator for confusion as found in [6, 7, 10, and 11].

III. Stage 3

In the decision-maker stage, four types of different classifiers are proposed: Virtual Generalizing Random Access Memory (VG-RAM), Support Vector Machine (SVM), Logistic Regression and Quadratic Discriminant classifiers, that used to differentiate between a confused person who is not confused. Each classifier's recognition rate was different from others; therefore, two types of classifier that achieved the higher accuracy in this proposed work of CDS will be discussed in this section, which are Logistic Regression and Quadratic Discriminant classifiers. Logistic Regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique, which can not only be used to predict the probability but also used for classification. Logistic regression can be used to classify individuals in the target categories through the logistic function. It is related to the probability of the chosen outcome event [14, 15]. The Logistic regression equation can be obtained from the Linear Regression equation. The mathematical steps to get Logistic Regression equations are given below, the equation of the straight line can be written as [16]:

$$y = b_0 + b_1 x_1 + b_2 x_2\, b_3 x_3 + \cdots + b_n x_n \quad (1)$$

Where y is a target or dependent variable, b_0 is intercept, x_1, x_2, x_3, x_n are predictors or independent variables $b_1, b_2, b_3\ and\ b_n$ is coefficients or respective predictors. In Logistic Regression y can be between 0 and 1 only, so dividing the above equation by (1-y) result:

$$\frac{y}{1-y}; 0\ for\ y = 0, and\ infinty\ for\ y = 1 \quad (2)$$

After that, a range between (–[infinity] to + [infinity]) is needed, then taking the logarithm of the equation, it will become the final equation for Logistic Regression:

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2\ b_3x_3 + \cdots + b_nx_n \quad (3)$$

This current prediction function returns a probability score between 0 and 1. In order to map this to a discrete class (true/false, confused/not), selecting a threshold value or tipping point above which will classify values into class 1 and below, which will classify values into class 0.

$$P \geq 0.5, \text{class} = 1$$

$$P < 0.5, \text{class} = 0$$

An example of a logistic regression function is shown below in Figure 2:

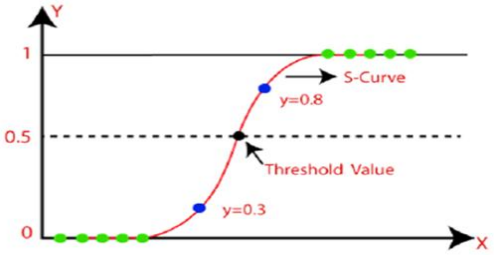

Figure 2: An example of a logistic regression function.

The quadratic discriminant classifier (QDC) is a well-known parametric Bayesian classifier that has been successfully applied to statistical pattern recognition problems. One such application is in automatic face recognition. Experiments indicate that the proposed classifier leads to remarkable improvement in face recognition accuracy as compared to the existing classifiers such as the nearest neighbor, support vector machine, and Naive Bayes [16]. In classification, the objective is to assign a given pattern, such as a face, to one of K classes, namely, $\omega 1$, $\omega 2$,...,ωk on the basis of a feature vector F = [$f1$,$f2$,$f3$,…,fT]′associated with the pattern. Statistical pattern recognition considers that the feature vector F is a T-dimensional observation drawn randomly from the class-conditional PDFp (**F**|$\omega\ell$), where $\omega\ell$, is the class to which the feature vector belongs [17]. If the class-conditional PDFs are multivariate Gaussian, then the optimal Bayesian decision rule based on a 0/1 loss function yields the QDC. This classifier assigns an unknown pattern represented by F to the class $\omega\ell$ that minimizes [18]:

$$\mathrm{d}(F) = ln|S_\ell| + (F + \mathcal{F}_\ell)'S_\ell^{-1}(\mathrm{F}\_\mathcal{F}_\ell)\_2\ln\ \pi_\ell \quad (4)$$

Where $\pi\ell$ is the prior probability associated with, $\omega\ell$,ℓ is the MLE of the true covariance matrix $\boldsymbol{\Sigma}\ell$ given by the equation below Moreover, $\boldsymbol{F}\ell$ is the sample mean vector of the $\boldsymbol{\lambda}tr$ feature vectors in class $\omega\ell$:

$$S_\ell = \frac{1}{\lambda_{tr}-1}\sum_{k=1}^{\lambda_{tr}}(F_k^\ell\_\mathcal{F}_\ell)(\ F_k^\ell\_\ \mathcal{F}_\ell\ )' \quad (5)$$

3. DATASETS COLLECTION

A database is a collection of data that provides a forum for checking both the confusion detection systems and algorithm robustness. To date, a confusion detection study has used a few databases, and none of them has been made publicly accessible. Hence, it is important to collect databases to assess the efficiency of the proposed confusion detection. The datasets are gathered from 120 participants (29 females, 91 males) between the ages of 18-45. Each participant was asked to answer a set of

different questions (personal social questions) in the interview. Questions were asked while the camera (CANON) recorded behavioral facial expressions during the entire session in ordinary circumstances. The data sets are collected with backgrounds that are widely accessible, and the participants' head pose is not constrained. As shown in Figure 3, in other words, subjects are able to shift their heads unrestrictedly. The database gathered is consistent with realistic facial expressions, rather than been produced or created.

Figure 3: The interview session. Face expression of the subject is registered while asking the questions by the examiner.

4. THE PROPOSED CONFUSION DETECTION SYSTEM

The proposed confusion detection system is constructed with three stages. The first stage is responsible for video capturing, facial identification, and finding any important points on the participant's face. This stage is called "Video Capturing and Pre-Processing," which arranges the captured video for the next second stage, "Dynamic Feature Extraction and Action Units (AUs) detection," this stage-manages the participant's face, which is detected in order to extract the appropriate features from it. Representing by AUs that detected and extracted from the face based on two types of features (appearance and geometry) features. "Decision Maker (Classification)" is the final stage where a confused individual is detected and recognized from whom not confused depending on AUs extracted in the earlier stage. Which is VG-RAM, SVM, Logistic Regression, and Quadratic Discriminant. Figure 4, illustrates the step structure of the proposed system.

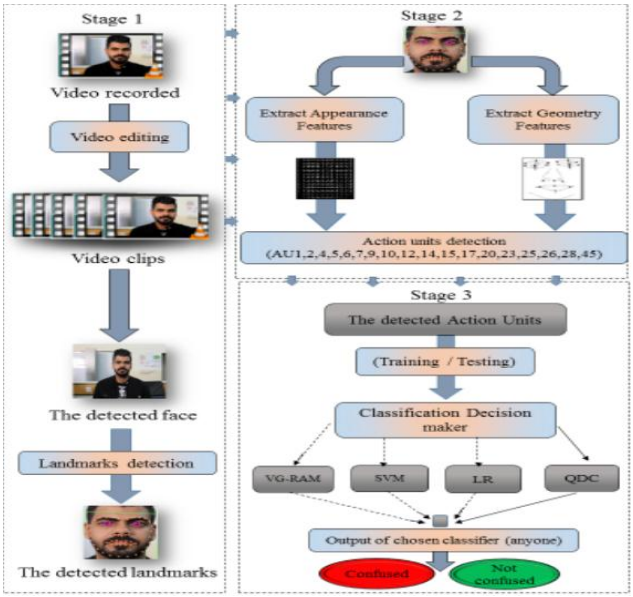

Figure 4: illustrates the step structure of the proposed system.

A. Video recording and editing

The videos are recorded using a digital camera; the camera has a small display LCD screen. The file format type of the recorded video is the MOV extension. In this work, the videos are recorded at a resolution of 1920x1080 pixels per frame for 120 participants to test the robustness of the proposed CDS. After the videos are recorded, these videos must be modified before extracting features from it, splitting between each sentence expressed by the participant. Where every clip has a length of time of

about 1 second. Such editing is essential so that undesirable parts of the original videos can be removed and separating confusion responses from other responses. In total, the number of clips obtained from all original videos is 490 video clips: 245 for the confused response and 245 for the not confused response. The numbers of original videos are 120; each participant has one video that contains confused and not confused responses. After editing, for both responses, each participant has approximately 8-12 video clips (4-6 video clips for each response).

Face detection is used to check the existence of the face region in the video, Where the VJ technique is used, as mentioned in the previous section, to process the image sequence for facial detection that was carried out for the dataset collected. The faces of all participants are passed by all these 31 AdaBoost layers, then implementing skin cheek only to the cascaded AdaBoost. Finally, the algorithm used the decision threshold to pick the best AdaBoost layer. After the face detection step gives a bounding box around the face. The Constrained Local Neural Field (CLNF) fitting can start based on initial shape parameters. The fitting algorithm finds the optimal shape parameters. The system is beginning to applying landmark detection to each image in the input video independently in a sequence, So CLNF gives the location of 68 facial landmarks. The face detection and CLNF landmark detection and tracking are shown in Fig 5, 6 respectively.

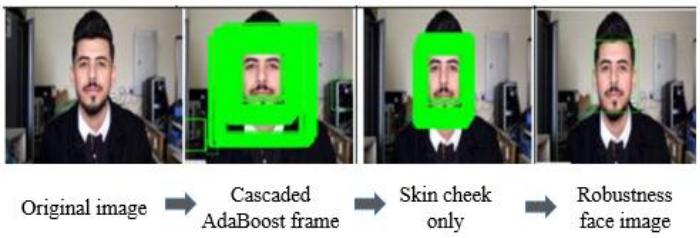

Figure 5: Face detection using VJA.

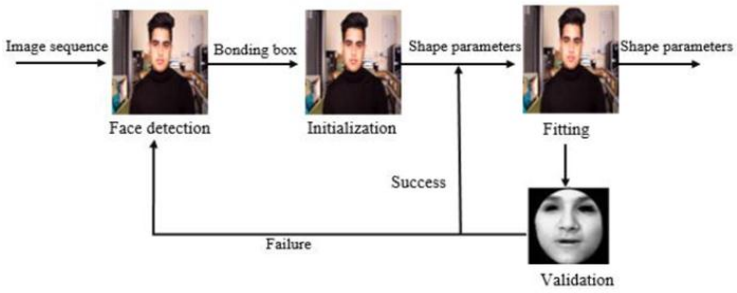

Figure 6: CLNF landmark detection and tracking.

B. Feature extraction

The input video clip consists of 10-35 frames (depend on the length of video clips); each frame has its corresponding AUs vector. This leads to 10-35 AU vectors for each video clip. To train and test the classifier on these visual feature vectors, the video clips should have the same number of AUs vectors. In this work, one AUs vector is chosen for each video clip by select the most repeated AUs on frames sequence. The most repeated AUs of frames sequence (belong to the same video clip) are selected depending on the presence or absence of each AU in the frames sequence of the video clips. The proposed system recognizes 18 facial AUs (AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45) some of these AUs have no effect on confusion detection processes. Based on the result of the proposed AUs detection system using collected datasets (for confused and not confused responses) show that some AUs not affected during the confused and not confused response approximately remain in the same state during all interviews. Only seven AUs have a significant effect during confused and not confused responses, as shown in Table I, which summarizes these seven AUs used in this work as potential indicators for detection confusion.

TABLE I: Name and region of the effective AUs based on FACS.

Action Units	FACS name	Facial Expression
AU4	Brow Lower	
AU5	Upper Lid Raiser	
AU6	Cheek Raiser	
AU7	Lid Tightener	
AU10	Upper Lip Raiser	
AU12	Lip Corner Puller	
AU23	Lip Tightener	

C. Classification

In the decision-maker stage, four types of different classifiers are proposed: VG-RAM, SVM, Logistic Regression, and Quadratic Discriminant. These classifiers are trained and tested from the collected database, the training and evaluation database was selected on a random basis. Four hundred ninety video samples have been collected for both males and females (245 video clips for confused response and 245 video clips for not confusing response). This dataset is divided into training and testing (50% for training and 50% for testing).

The collected data sets are divided into two groups; the first group used to learn the classifiers, and the second group used to evaluate their performance. The accuracy of each classifier is considered a key parameter for evaluating the performance of each classifier. Based on the collected database and identified AUs, the performance of suggested classifiers is compared to finding the best classifier which can distinguish a confused person from someone who is not confused.

5. THE PROPOSED CLASSIFIER PERFORMANCE RESULTS

The classifiers results for each SVM, VG-RAM, Logistic Regression, and Quadratic discriminant are shown in Table II, based on the recognition rate of the result obtained when using all datasets. The Table reveals the accuracy of detection for the four proposed systems by separating confused participants from unconfused participants. Logistic Regression and Quadratic discriminant classifiers have the highest detection accuracy results, which is 96.3415% for both classifiers based on all collected datasets. While detection accuracy results for SVM and VG-RAM classifiers were also promising but less than Logistic Regression and Quadratic discriminant classifiers.

TABLE II: The proposed confusion detection classifiers Performance.

Proposed classifier	Detection accuracy
SVM	93.0894%
VG-RAM	95.5285%
LR	96.3415%
QDA	96.3415%

6. PERFORMANCE COMPARISON WITH PREVIOUS WORKS

The best performance of the suggested systems will be compared to earlier works so that the proposed CDS robustness is evaluated. The proposed system achieved the best performance results by using logistic Regression and Quadratic discriminant classifiers. It is difficult to compare results for the same environment; the Datasets used in previous works are not publicly available and are different from the collected data sets used in the proposed CDS are the reason for this difficulty. Many of the earlier works have gathered the datasets through interviews and online courses, resulting in huge differences in the database that have been collected from the database of previous work. An example of these variations is the participants' psychological state, the lighting condition, and the occlusion effect. There are a variety of technologies on which previous systems rely, for example (muscles and brain activity, facial expressions, and eye movement). The comparisons were made depending on the participant's number and the detection accuracy since the comparison according to methods, which were employed, the algorithms, the classifiers, and features number is difficult to make. Table III, listed comparisons depending on the participant's number, type of features, and the accuracy of detection of the proposed CDS with the confusion systems in the previous works. Unconcerned with the technology used, the Table display different techniques were used in previous work to detect confusion. The two main challenges facing CDS were the number of participants and the accuracy of the detection to evaluate CDS robustness.

According to the Table below, the proposed method is essentially preferable for all previous works since the currently available dataset is greater than the number of datasets that have been used in previous work.

TABLE III: A Comparison based on the number of participants and detection accuracy results with previous works.

Ref. No.	Type of features	No. of participants	Detection Accuracy
[6]	Action unit AU4	14 videos	86%
[1]	EMG data	24	87.5%
[7]	Action units AU4 and AU7	67	85%
[19]	Fifteen different eye movement	20	--
[20]	The urgency and Question Variables	--	74%
[21]	Participants' gaze trajectories and fixations	14	--
[22]	driver's behavior and the traffic conditions	11	--
[2]	EEG data	10	73.3%
[23]	EEG data	17	71.3%
[10]	Action units AU25, 26 and 27.	20	81%
[11]	narrowing of the eyes and lowering of the eyebrows, similar to frowning	82	93.8%
Proposed system	Facial expressions based on seven AUs	120 (490 video clips)	96.3415%

7. CONCLUSIONS

The proposed CDS has proven the theory of using facial movements to detect confusion. Where a new database provided to enhance and evaluate the performance of the proposed confusion detection system. These samples were extracted from 120 participants. Seven Action Units (AU 4, 5, 6, 7, 10, 12, and 23) have a sufficient effect in the detection of confusion depending on the database that has been collected, while the features that have been redacted are (AU 1, 2, 9, 14, 15, 17, 20, 25, 26, 28, and 45) which have no effect on the recognition rate. The suggested CDS is an autonomous system capable of recognizing the input video clip that belongs to anybody in unconstrained environments, even though it is not part of the data training of the system.

References

[1] Durso, T. Francis, K. M. Geldbach, and P. Corballis. "Detecting confusion using facial electromyography." Human factors 54.1 (2012): 60-69

[2] Ni, Zhaoheng, et al. "Confused or not confused? disentangling brain activity from eeg data using bidirectional lstm recurrent neural networks." *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2017.

[3] Atapattu, Thushari, et al. "An Identification of Learners' Confusion through Language and Discourse Analysis." arXiv preprint arXiv:1903.03286 (2019).

[4] Conati, Cristina, et al. "When to Adapt: Detecting User's Confusion During Visualization Processing." UMAP Workshops. 2013.

[5] P. Nilsenová, Marie, E. Postma, and K. Tates. "Automatic detection of confusion in elderly users of a web-based health instruction video." Telemedicine and e-Health 21.6 (2015): 514-519.

[6] Grafsgaard, Joseph F., K. E. Boyer, and J. C. Lester. "Predicting facial indicators of confusion with hidden Markov models." International Conference on Affective computing and intelligent interaction. Springer, Berlin, Heidelberg, 2011.

[7] Grafsgaard, Joseph, et al. "Automatically recognizing facial expression: Predicting engagement and frustration." Educational Data Mining 2013. 2013.

[8] D. Wit, Han F. On the methodology of clarifying confusion. North Holland: Elsevier Science Publishers Bv, 1987.

[9] D'Mello, Sidney, et al. "Confusion can be beneficial for learning." Learning and Instruction 29 (2014): 153-170.

[10] Borges, Niklas, et al. "Classifying Confusion: Autodetection of Communicative Misunderstandings using Facial Action Units." *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2019.

[11] Shi, Zheng, et al. "Automatic Academic Confusion Recognition in Online Learning Based on Facial Expressions." 2019 14th International Conference on Computer Science & Education (ICCSE). IEEE, 2019.

[12] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, vol. 57, no. 2, p. 137–154, 2004.

[13] T. Baltrušaitis, L. P. Morency and P. Robinson, "Constrained local neural fields for robust facial landmark detection in the wild," in IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 2-8 Dec. 2013.

[14] Zhou, Changjun, et al. "Face recognition based on PCA and logistic regression analysis. "Optik 125.20 (2014): 5916-5919.

[15] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (11) (2010) 2106–2112.

[16] Stoltzfus, Jill C. "Logistic regression: a brief primer." *Academic Emergency Medicine* 18.10 (2011): 1099-1104.

[17] A. S. Shahnewaz, T. Howlader, and S. M. Rahman. "Pooled shrinkage estimator for quadratic discriminant classifier: an analysis for small sample sizes in face recognition." International Journal of Machine Learning and Cybernetics 9.3 (2018): 507-522.

[18] Kittler J (1994) Statistical pattern recognition in image analysis. J Appl Stat 21(1–2):61–75.

[19] Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 22(1):4–37.

[20] DeLucia, Patricia R., et al. "Eye movement behavior during confusion: Toward a method." Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 58. No. 1. Sage CA: Los Angeles, CA: SAGE Publications, 2014.

[21] Agrawal, Akshay, et al. "YouEDU: addressing confusion in MOOC discussion forums by recommending instructional video clips." (2015).

[22] Pachman, Mariya, et al. "Eye tracking and early detection of confusion in digital learning environments: Proof of concept." Australasian Journal of Educational Technology 32.6 (2016).

[23] Hori, Chiori, et al. "Driver confusion status detection using recurrent neural networks." 2016 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2016.

[24] Zhou, Yun, et al. "Confusion State Induction and EEG-based Detection in Learning." 201840th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018.