IAES International Journal of Robotics and Automation (IJRA)
Vol. 11, No. 1, March 2022, pp. 21~32
ISSN: 2722-2586, DOI: 10.11591/ijra.v11i1.pp21-32

Recognizing facial emotions for educational learning settings

Oryina K. Akputu[1], Udoinyang G. Inyang[2], Msugh Ortil[3], Fiza T. Mughal[4], Abel Usoro[5]
[1]Department of Computing Sciences, Admiralty University of Nigeria, Delta State, Nigeria.
[2]Department of Computer Science, University of Uyo, Uyo, Nigeria.
[3]Department of Computer Science, FCT College of Education Zuba, Abuja, Nigeria.
[4]Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia.
[5]School of Engineering and Computing, University of the West of Scotland Paisley, Renfrewshire, United Kingdom.

Article Info

Article history:

Received Jun 20, 2020
Revised Dec 25, 2021
Accepted Dec 30, 2021

Keywords:

Contextual dataset
Educational learning
Emotions recognition
Learning emotions
Personalization

ABSTRACT

Educational learning settings exploit cognitive factors as ultimate feedback to enhance personalization in teaching and learning. But besides cognition, the emotions of the learner which reflect the affective learning dimension also play an important role in the learning process. The emotions can be recognized by tracking explicit behaviors of the learner like facial or vocal expressions. Despite reasonable efforts to recognize emotions, the research community is currently constraints by two issues, namely: i) the lack of efficient feature descriptors to accurately represent and prospectively recognize (detecting) the emotions of the learner; ii) lack of contextual datasets to benchmark performances of emotion recognizers in the learning-specific scenarios, resulting in poor generalizations. This paper presents a facial emotion recognition technique (FERT). The FERT is realized through results of preliminary analysis across various facial feature descriptors. Emotions are classified using the multiple kernel learning (MKL) method which reportedly possesses good merits. A contextually relevant simulated learning emotion (SLE) dataset is introduced to validate the FERT scheme. Recognition performance of the FERT scheme generalizes to 90.3% on the SLE dataset. On more popular but noncontextually datasets, the scheme achieved 90.0% and 82.8% respectively extended Cohn Kanade (CK+) and acted facial expressions in the wild (AFEW) datasets. A test for the null hypothesis that there is no significant difference in the performances accuracies of the descriptors rather proved otherwise ($\chi^2 = 14.619, df = 5$, $p = 0.01212$) for a model considered at a 95% confidence interval.

Corresponding Author:

Oryina K. Akputu
Department of Computing Sciences, Admiralty University of Nigeria
Ibusa Ugwashi-Uku Express way, Delta State, Nigeria
Email: oryinak@gmail.com

1. INTRODUCTION

Previous studies have shown how learners who received personalized, one-on-one educational instructions learn better and faster than those who received traditional one-side-fit-all instructions. However, providing such personalized learning settings might likely go beyond the educational training resources and budgets of most institutions. The e-learnings setting [1] is one instance of a promising alternative to the traditional one-side-fit-all approaches and offers the advantage of being cost-effective [2]–[4]. The personalization of the educational instructions to a learner is achieved through the intelligent tutoring system (ITS) [5]–[7]. A learner model, which is the main component of the ITS, consists of the motivational, cognitive, and affective states that have important effects on the learning performance of the learner [8].

Moreover, since affective states such as emotions are dominant in the teaching and learning process, recognizing them can largely enable the ITS to undertake actions that significantly influence tutoring quality. Besides some researchers have opined that affective state and the learner's emotional state, in particular, should be important factors to consider in designing instructional materials [9]–[11]. Other studies have also emphasized the need to induce and conduct the learner's emotions to the suitable state in learning settings [3], [11], [12]. However, first of all, the learner's emotions have to be recognized by the system. In this regard, there are different methods in the context of human emotions recognition. For instance, by tracking implicit parameters, including, speech recognition [13], [14] facial expression recognition [15]–[17], physiological means [18], [19] body gesture recognition [20] or multimodal or fusion means [12], [21], [22]. While some sensory cues such as speech, body language, and physiological measures, may not yet be realistically and decipherable by a computer as effortless as the human does, the facial expression could be applicable. Notably, multimodal approaches for emotion recognition are reportedly limited due to underlying feature correlation that compromises system performance [23]. Alternatively, facial expression recognition could be recommended; the face cues contribute as much as 55% of the effects in most human communication compared to other cues.

Consequently, several efforts have focused on studying facial emotion recognizers. But relatively few studies emphasize the need to experiment with the emotion recognizers on realistic learner’s dataset. Such concentration should provide a valuable reference for designing educational materials since personalized learning settings have emerged to some extent in response to the need for diversity in educational resources. One problem often encountered is perhaps how to accurately recognize or detect the emotional states of the learners. In this regard, a variety of feature descriptors [3], [24]–[26] are utilized with traditional classifiers [15], [27], [28] for enhanced performance accuracy of the emotion systems. However, a serious challenge of most systems, which this paper tries to solve is that previously reported performances have rarely been done on contextual datasets that reflect real learning settings. Findings using settings other than the prospective learning environment, cannot be generalized upon as the contextual ones. As a contribution, this paper tries to synthesize and harmonize underlying interdisciplinary linkages between learning settings and emotion recognition research into a coherent whole. A comparative study has been conducted across selected feature descriptors resulting in a suggested facial expression recognition (FER) application, herein referred to as FER technique (FERT). A contextually relevant dataset-simulated learning emotion (SLE) is introduced for experimental analysis to study the influence of the contextuality on performances of the scheme for the prospective learning settings. Additionally, two more popular benchmark datasets namely extended Cohn Kanade (CK+) and active facial expression in the wild (AFEW) datasets have also been used to validate performance results.

The remaining part of this paper is organized as follows; Section 2 briefly describes the research methods. Section 3, presents results and discussions. Section 4 concludes the paper.

2. RESEARCH METHOD

This section discusses previous research approaches on the use or efficacy of human emotion in learning settings. The processing steps leading to the design of the emotion recognition method, FERT are also discussed.

2.1. Educational learning settings

The educational learning settings (ELS) include all education-centered learning that can exploit cognitive characteristics of the learner and any predefined variable to adapt educational content to learning needs. The ELS provides the framework to express functionalities of learning adaptation and how this could be arrived at. Two aspects (problems) of the adaptation have been identified in the literature [29]. One aspect pertains to ‘what can be adapted to?’-various adaptive characteristics of the learner, such as cognitive traits (working memory capacity, inductive reasoning ability, and meta-cognitive skills) [1], interests, experiences, learning styles, context, and environment. Another aspect of the adaptation pertains to ‘what can be adapted?’-various strategies for the learning content presentation (adapt the actual content, or media used) as well as navigation (link destination and overview for orientation support). In this context, the focus is on the former adaptation scenario (i.e., what can be adapted to?). The purpose here is to consider several learner-level adaptation parameters that contribute to describing their characteristics and contexts. While some studies have consistently shown how humans have some cognitive traits that can be adapted to [30], [31]. Other researches have also shown how the affective processes of the learner also influence their learning process [29], [32], [33]. The effective processes involve emotions and how they are regulated to impact learning.

2.2. Emotion in learning to set

The neurology of emotion suggests learning, attention, memory, and human social functioning, are all connected with emotional processes [4], [34]. Generally, the emotions of the learner, especially the positive ones (e.g., happiness, engagement, satisfaction, and hopefulness) can have a positive effect on learning [11], [33], [35]. The impact the emotions have on learning performances underscores the need for their perception and consideration in learners' modeling [36]. For emotion modeling in learning settings, educationists and cognitive psychologists often refer to well-established Russell's dimensional circumflex model of affect [37]. Studies that utilize Russell's model could also be found [38], [39]. In particular, Craig *et al*. [38], report the occurrence of six emotion states of frustration, boredom, flow, confusion, eureka, and neutral during learning interaction with the intelligent learning system. However, some of the emotional states (e.g., eureka and neutral) may not be relevant to learning settings. Moreover, it is believed that expert human teachers react to offer remedial support to learners based on just a few sets of emotions as opposed to a large set. Elsewhere, Akputu *et al*., [40], seven frequently occurring learning emotions are considered, viz., engagement, confusion, frustration, boredom, hopefulness, satisfaction, and sadness. One relevance of that study is the creation of the SLE-dataset which conceptualizes the seven classes of learning emotions in a realistic learning scenario. Such conceptualization becomes imperative if we must ensure the emotional well-being of learners besides cognitive factors. Perhaps among the first efforts in that direction is designing emotions recognition techniques for potential inclusion in future personalized learning systems.

2.3. Emotion recognition methods

The main objective of an emotion recognition method is perhaps to determine the emotional category or class to which a given sample may represent. This problem is challenging in the sense that, emotion samples within a class would usually exhibit some feature diversity, whereas those of other classes may correlate significantly. One well-known approach to addressing the problem is the use of suitable feature descriptive techniques to effectively represent the facial feature cue. Generally, a good descriptor can enhance the subsequent descriptive power of the classifiers, thereby improving recognition accuracy. A variety of descriptor techniques are available for facial feature representation. These includes, Gabor filters [41], principal component analysis (PCA) [24], [27], and Fisher linear discriminant (FLD) (e.g., linear discriminant analysis (LDA)) [27], [42]. Notably, each of these feature extraction methods has individual merits over their counterpart's approaches [23]. However, it also remains to be seen how performances of these features can be generalized in contextual settings. The settings considered here are the educational learning scenarios or learning datasets.

Emotion recognition using the learning datasets requires reliable classifier methods besides the feature descriptors. In this regard, support vector machines (SVM) could be used [27]. However, a drawback of the SVM is that it encodes feature diversity via a single parametric kernel, which impairs classification accuracy. Therefore instead of the classical SVM, the multiple kernel learning (MKL) pioneered by Bach, Lanckriet, and Jordan [43] is used. Moreover, the MKL has been recommended for emotion recognition tasks [14], [15]. The MKL work by simultaneously learning an optimal kernel combination of distinct kernel and associated parameters. The distinct kernels can encode various feature attributes in a higher dimension, which has merit in enhancing class discrimination over classical SVM classifiers. Notably, a few limitations of the MKL including the inability to discriminate redundant features as well as misclassification behavior have been addressed in previous studies [14], [15]. This study is particularly inspired by the recently introduced MKL decision tree with WFA (MKLDT-WFA). In the study of Researchers, the MKLDT-WFA efficiently encodes face image feature diversity for enhanced emotion classification accuracy. However, it remains to be seen how the MKLDT-WFA can fare across feature descriptors (PCA, LDA, with their possible combinations) for potential usage in the emerging facial emotion recognition technologies.

2.4. Proposed facial emotion recognition technique

The suggested FERT scheme is shown in Figure 1. First, the face is detected from the input image sequence, using a variant of the Viola and Jones method [44]. The detected face, denoted, is resized to using bi-cubic greyscale interpolation. The second step, which is feature extraction is based on the results of a preliminary performance analysis across feature descriptors. The descriptors that were studied include, Gabor wavelet, PCA, and LDA, as well as three possible combinations among these, denoted as $PCA + LDA$, $PCA + Gabor$, and $PCA + LDA + Gabor$. Nevertheless, the third step of the FERT scheme which is emotion classification which utilizes a reliable variant of the MKLDT-WFA pipeline.

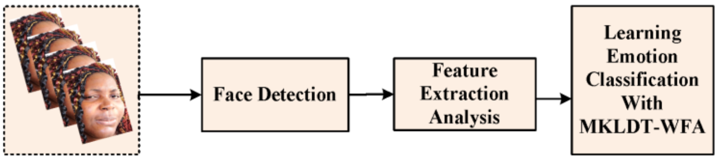


Figure 1. Processing pipeline of FERT

2.4.1. Feature extraction using Gabor wavelet

The 2D Gabor filters [41], are utilized to extract features of every detected facial image. Let, $I(x,x')$ denotes the grey scaled image with (x,x') as the coordinates of image center parts. Exactly 24 Gabor filters (12 real and 12 imaginary filters) are derived for an image frame.

The set, $S_{\psi} = \{\psi_{0,0}(I): \mu \in \{0,1,2,3,4,5\},\ldots,\tau = \{0,1,2,3\}\}$, contains the Gabor wavelet filter representation of the image, $I(x,x')$. The choice of the number of frequencies and magnitude is believed [45], to offer optimal discrimination. Notably filtering the face image frame with the 24 from the Gabor filter bank results in an inflation of the dimensionality 24 times the initial size of 256×256 pixels. That is the 24 Gabor magnitude resides in the $1572864\ (256 \times 256 \times 24)$ dimension, which would be too expensive. Therefore, the resulting Gabor magnitude is normalized to zero mean and unit variance presented in the form (1).

$$V^{(d)} = \left[\left(\psi_{0,0}(i)^{\mathrm{T}}, \ldots, \psi_{3,5}(i)^{\mathrm{T}}\right)^{\mathrm{T}}\right]_{0,0}^{\psi}(i)^{\mathrm{T}} = \{V_1, \ldots, V_i\}_{i=1}^{g} \tag{1}$$

2.4.2. Feature extraction using Gabor wavelet

The Gabor magnitude dimensionality feature issues offer a problem space where the PCA [46], fits in. A good property of the PCA is that it can extract discriminant features from high-dimensional data and presents them in relatively lower-dimensional space, thus lowering the computation cost. Besides lowering the cost of dimensionality, another good property is the fact that it reduces underlying redundancy while maintaining the diverse information in the original data distribution. We consider a set S of N-dimensional sample images $S = \{x_i\}_{i=1}^{N}$, and assume that each image belongs to one of the P classes, i.e., $\{p_1, p_2, \ldots, p_i\}_{i=1}^{P}$, $p_i = \{x_1, x_2 \ldots, x_i\}_{i=1}^{n}$. The p_i is a set of face samples with the same expression class and N is the total number of sample images. Consider, a linear transformation that maps the original N-dimensional image space onto an M-dimensional feature space, $v_i \in \Re^{N\times M}$, such that, a resultant scatter of transform features spaces $\{v_1, \ldots, v_i\}_{i=1}^{g}$, becomes [42].

$$W^T S_t W \tag{2}$$

Where $W \in \Re^{N\times M}$ and S_t are orthonormal column and the total scatter matrixes respectively. Furthermore, the determinant of the projected samples is maximized by choosing an optimal projection W_o as (3).

$$W_o = arg \ \max_{W} |W^T S_t W| = [w_1, w_2, \ldots, w_m] \tag{3}$$

2.4.3. Feature extraction using LDA

The $\{w_i | i = 1,2,\ldots,m\}$ is the set of m-dimensional Eigenvectors of the scatter S_t. Even though the feature representation with the PCA offer (i.e., reduces redundancy) lower dimensionality, its performances in the latter experiment are not encouraging. This is because the PCA projects unwanted components along the Eigenvectors due to factors including lighting and facial expression [25], [47]. Therefore, this paper followed procedure in [47] to discarding the first three principal components to improve the performance of this descriptor,

$$S_B = \sum_{i=1}^{N} N\,(\theta_i - \theta)(\theta_i - \theta)^T \tag{4}$$

and, the within-class scatter matrix as (5).

$$S_W = \sum_{i=1}^{n} \sum_{x_i \epsilon p_i} (x_i - \theta_i)(x_i - \theta_i)^T \tag{5}$$

The n represents the number of samples in the class p_i.take S_W as a non-singular, then we have the following optimal projection as (6).

$$W_o = arg \max_W \left|\frac{W^T S_B W}{W^T S_W W}\right| = [w_1, w_2, \ldots, w_m]. \quad (6)$$

The matrix $W_o$is orthonormal columns that maximize the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix. The $\{w_i|i = 1,2,\ldots,m\}$ denotes a set of generalized Eigenvectors of S_B and S_W which corresponds to m larger generalized Eigenvalues, $\{\lambda_i|\ i = 1,2,\ldots,m\}$ of the form (7).

$$S_B W_i = \lambda_i S_W W_i, i = 1,2,\ldots,m \quad (7)$$

2.4.4. Emotion classification using MKL

Besides obtaining facial feature representation such as Gabor wavelet, PCA, or LDA, an efficient feature classifier is additionally required to generalize emotion classes to distinct feature diversity for the emotion recognition task. In this regard, the MKLDT-WFA has been utilized [15]. The MKLDT classifies input data by simultaneously learning an optimal combination of distinct kernels along with associated parameters. The distinct kernels learned to transform various features information encoded from data by the descriptors; features from distinct classes are then mapped into a new dimensional space patterning to diverse emotional classes. The objective in context is to designate *p*-classes emotion classes. Classification with MKLDT-WFA begins from the root nod n_1 of the DT routine. The *p*-classes of a father node are divided into binary disjoint clusters or child nodes or non-leaf node n_i.Consider the cluster group a node to be G_1 and G_2 each containing at least one class or possibly multiple classes. The approach realized the grouping of the data using the following distance measure [15].

$$d_{i,j}(x_{e,i}, c_j(n_i)) = \sqrt{K_{x_i x_i} - 2K_{x_i c_j}(n_i) + K_{c_j c_j}(n_i)}, i,j = 1,\ldots,N \quad (8)$$

Where,

$$d_{i,j} = \begin{cases} 1, d_{i,j}(x_{e,i}, \bar{c}(n_i)) \\ 0, \text{other.} \end{cases} = \underset{m}{Min}\{d_{i,j} x_{e,i}, \bar{c}_m(n_i)\} \quad (9)$$

The c_j is the cluster center at a node n_i; K_{x_i,c_j} and $K_{c_i,c_j}$are cluster center kernel map with its group's sample and cluster center of another group respectively. With the cluster groups obtained, the MKLDT-WFA classifier decides (10).

$$f(x_{e,i}) = \sum_{q=1}^{Q} \sum_{e,i,j=1}^{N,E} \sum_{m=1}^{M} d_m^{c_i} d_m^{c_j} \alpha_i y_j Km(x_{e,i}, x_{e,j}) + b \quad (10)$$

The terms α^* and b is the Lagrange multiplier and the offset constant respectively. The K_m is kernel combination function. Finally, the decision function in (10) is solved for the values of d_m, a_* and b using the objective function [15].

$$\underset{d}{Min}\, J(d) = \sum_{q=1}^{Q} \sum_{i,j=1}^{N} \sum_{m=1}^{M} (d) \quad (11)$$

Where,

$$J(d) = \left\{ \underset{\alpha}{\text{Max}} \sum_{q=1}^{Q} \alpha_{i,q} - \frac{1}{2} \sum_{j=1}^{N} \alpha_{i,q} \alpha_{j,q} y_i y_j \sum_{m=1}^{M} d_m K_m(x_{e,i}, x_{e,j}) \right. \quad (12)$$

Notably,

$$\begin{aligned} \alpha_{i,q}\alpha_{j,q} &\in [0, C] \\ \sum_{j=1}^{N} \alpha_{i,q} y_i &= \forall i,j = 1,\ldots,N \end{aligned} \quad (13)$$

2.4.5. Experiment

In this section, the experiments are presented to achieve three main objectives. The first is to study possible performance improvement of the FERT across the six descriptors viz. the Gabor wavelet, the PCA, the LDA, the $PCA + LDA$, the $PCA + Gabor$, and $PCA + LDA + Gabor$. The second objective is to address the lack of comprehensive studies on emotion recognizers in contextually relevant learning settings. In this regard, this paper adopts SLE [40], a contextually relevant dataset for studying performances of the FERT for a potential learning environment. Moreover, two more popular datasets have been used viz. AFEW 4.0 [48] and the Cohn-Kanade dataset (or CK+ in short) [49]. A brief description of each dataset is as follows: i) The

SLE dataset: [40], is the only contextual emotion recognition dataset used. Although a majority of other datasets (e.g., camera and interface) exist, most of them are not recorded under ideal learning settings. The SLE dataset presents seven learning emotions-engagement, hopefulness, happiness, boredom, frustration, surprise, and confusion; ii) The AFEW 4.0 [48] dataset: contains samples of subjects expressing one of the following seven emotions: neutral (neu), happy (hap), sadness (sad), disgust (dis), fear (fea), anger and surprise (sup); iii) the CK+ [49] dataset: used contains 593 facial image sequences from 123 subjects. However, only 327 of the sequences met the criteria of the seven universal emotions of anger, disgust, fear, happiness, sadness, surprise, and neutrality. Only 327 portions resulting from frame extraction have been utilized in this paper. Figure 2 reflects sample images from each of the datasets, Figure 2(a) is shown sample images of datasets SLE, Figure 2(b) is shown sample images of datasets AFEW, Figure 2(c) is shown sample images of datasets CK+ datasets, and Table 1 highlights important attributes of each dataset.

(a) (b) (c)

Figure 2. Sample images of datasets: (a) SLE, (b) AFEW, and (c) CK+ datasets

Table 1. Dataset attributes

Dataset	SLE	CK+	AFEW 4.0
No. of emotion classes	7	7	7
Size of dataset	1350 (225 videos x6)	5831 images	1.368
No. of actors	25	123	428
Size of train set	810 (135 videos x6)	3.498	3468 (578 x 6)
Size of test set	540 (90 videos x 6)	2.332	2442 (407 x 6)
Size of a validation set	Not applicable	Not applicable	2298 (383x 6)

3. RESULTS AND DISCUSSION

This section presents the results of the experiments with key discussions. The discussions cover prospective ELS or applications alike that can be enriched with human emotion recognition capability to facilitate better human- computers-interaction.

3.1. Experimental results and discussion

In the computation of results, the parameter setting for the MKL classifier follows the work of [15]. The work of Researchers was also used to implement MKLDT-WFA, which can be realized by just utilizing Gabor features in the feature extraction analysis stage in the FERT pipeline. The final decision of the classifier is done based on majority voting over all the emotion estimates of the test set. System performances are measured using three metrics namely, classification accuracy, confusion matrix, mean average precision (MAP) as well as receiver operating characteristics (ROC) graph. Table 2 shows a comparison of recognition accuracy across feature descriptors. It could be noticed that the FERT pipeline built on the $PCA + LDA + Gabor$ combination achieves as much as 94.20%, 88.00% and 91.00% on the SLE, AFEW and CK+ datasets respectively. Among individual performance, the Gabor features outperforms PCA and LDA features which justifies its recommendation in previous studies [15], [25], [47]. The overall accuracy of results of $PCA + LDA + Gabor$ in Table 2 appears to somewhat contrast an earlier study [23] because the elimination of theunwanted components (eigenvectors of the PCA) was not taken into account thereby possibly accommodating redundancy.

Statistical inferences about the differences in feature descriptors performances were drawn by implementing non-parametric procedures [50] applying them individually to each of the three categories of datasets. Friedman test (a non-parametric variant of the repeated-measures analysis of variance) was used to test the null hypothesis that there is no significant difference in the performances accuracies of the descriptors.

The Friedman test results showed a significant difference in the accuracies ($\chi^2 = 14.619, df = 5$, $p = 0.01212$) for all the models at 95% confidence interval (CI). This indicates that the accuracy of at least one of the models is significantly different from others, hence the null hypothesis that all descriptors' performances are the same is rejected. Nemenyi's [51] post hoc test of the average rank of accuracies was performed with a critical difference (CD) of 5.1308. The top three performing descriptors were, $PCA + LDA + Gabor$, $PCA + Gabor$ and Gabor in that order, while PCA was the worst-performing model with an average rank of 6.0. LDA earned an average rank of 5.0, $PCA + LDA$ (3.67), Gabor (3.3) and $PCA + LDA + Gabor$ yielded an average rank of 1.0. Similarly, in terms of datasets, there was also a significant difference in the accuracies of descriptors in each of the datasets ($\chi^2 = 12, df = 2, p = 2.48 \times 10^{-3}$) with a $CD = 1.4997$ at 95% CI. SLE dataset had the highest average accuracy in all classifiers with an average rank of 1.0 followed by Ck+ ranked 2nd while AFEW depicted lower in the performance.

Tables 3-5 reflect the confusion matrices of FERT across the three datasets (SLE, CK+, and AFEW). Note how the learning emotion of 'happy' is recognized considerably higher precision (96%) compared to the rest. This following in precision scores are the two emotions of engagement and boredom. The lowest precision is achieved on the surprise emotion with 86%. Notably, the FERT scheme fared better on the SLE and CK+ datasets as well as emotion categories compared to its performance on AFEW dataset. Figure 3, reflects the ROC curve of FERT on SLE dataset. By observing the geometric appearance of the ROC convex hull for each emotion class, happy, engagement, and boredom emotions appear to have a considerably higher area compared to the rest. Figures 4 (a and b) reflect how the FERT compares in terms of MAP scores against other emotion recognition schemes on the same datasets. In Figure 4(a), the MAP score reported by most methods including, EDR-PCANet [52] and MKLDT-WFA [15] appears skewed and imbalanced across emotion classes. For instance, one could note how the EDR-PCANet shows a higher MAP on disgust and happiness but also performs poorly on anger, fear, and sadness. As a result, FERT outperforms these methods in terms of average recognition accuracy. In Figure 4(b) however, the FERT outperforms other methods for every emotion class on AFEW dataset. Table 6 presents performances of different methods on CK+ and AFEW datasets respectively with FERT as the control. It can be seen that FERT fares considerably better than other methods on each dataset. Findings in this result can further provide insight on prospective merits of the FERT scheme in future effective educational learning settings to facilitate personalization.

Table 2. Comparison of recognition accuracy across feature descriptors

	Dataset		
Method	SLE	AFEW	Cohn Kanade +
PCA	83.3	72.4	80.4
LDA	85.0	76.1	82.0
Gabor	90.1	77.9	86.7
$PCA + LDA$	89.1	79.0	86.0
$PCA + Gabor$	91.4	86.2	88.3
$PCA + LDA + Gabor$	94.2	88.0	91.0

Table 3. Confusion matrix of FERT on SLE

		Prediction					
		Eng	Hope	Bor	Hap	Fru	Sur
Truth	Eng	**387**	50	19	0	0	0
	Hop	22	**386**	25	0	21	0
	Bor	5	0	**449**	0	0	0
	Hap	0	0	0	**417**	20	37
	Fru	0	0	0	9	**419**	41
	Sur	0	0	0	7	23	**416**
precision		93	89	91	96	87	86
Accuracy				**90.3**			

Table 4. Confusion matrix of FERT on CK+

		Prediction						
		Dis	Fea	Hap	Ang	Sad	Sur	Con
Truth	Dis	**491**	0	0	25	30	1	9
	Fea	28	**492**	0	19	1	7	10
	Hap	0	2	**567**	0	0	4	1
	Ang	2	3	0	**470**	20	20	6
	Sad	0	18	9	7	**465**	0	55
	Sur	2	12	6	0	0	**551**	16
	Con	37	0	2	0	30	2	**470**
precision		88	93	97	90	85	94	83
Accuracy				**90.0**				

Table 5. Confusion matrix of FERT on AFEW

		Prediction						
		Dis	Fea	Hap	Ang	Sad	Sur	Neu
Truth	Dis	**303**	1	7	18	21	3	12
	Fea	13	**259**	12	13	8	16	17
	Hap	0	17	**306**	0	0	10	13
	Ang	19	13	0	**277**	29	11	1
	Sad	14	1	2	11	**258**	4	56
	Sur	8	10	8	6	9	**312**	2
	Neu	12	2	4	3	10	2	**308**
precision		82	85	90	84	77	87	75
Accuracy					**82.8**			

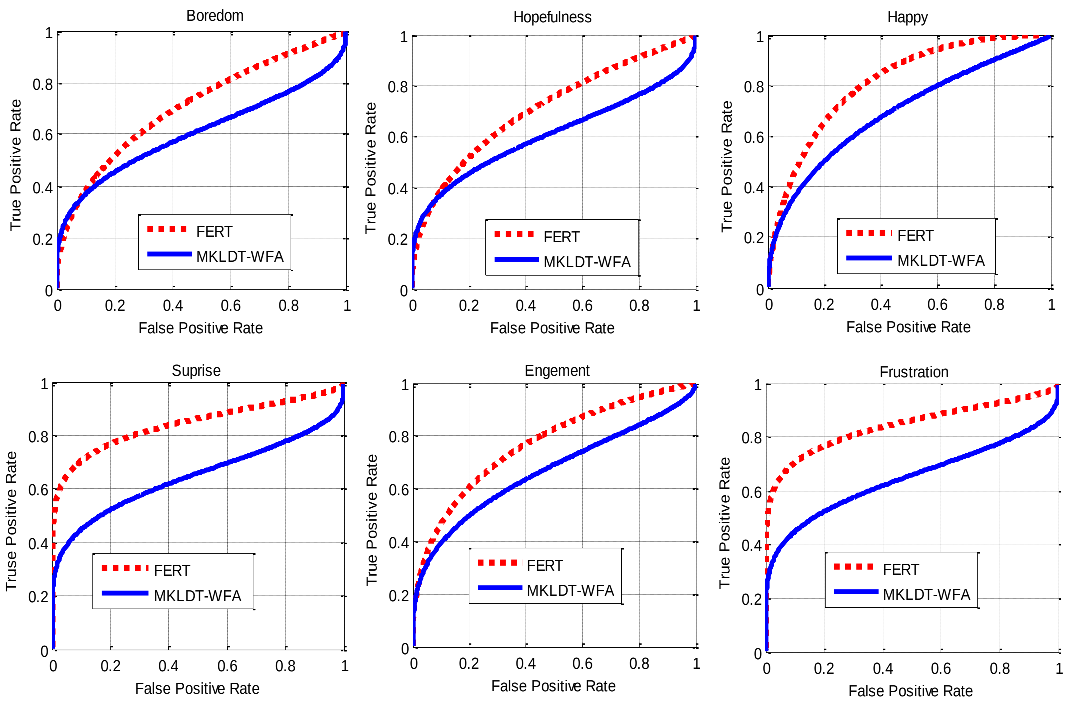
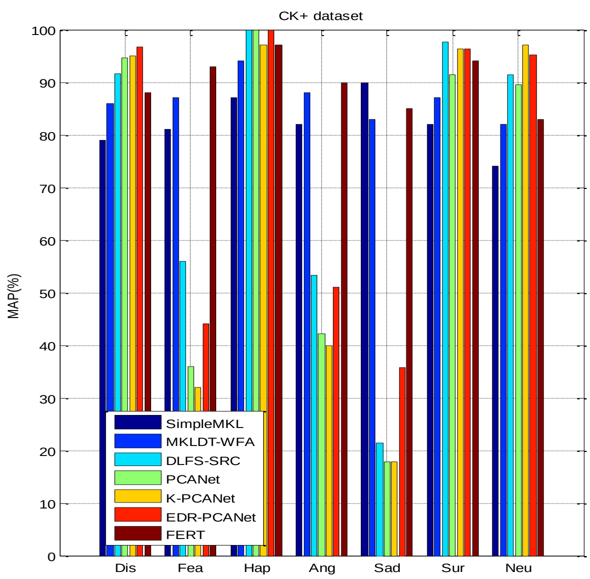
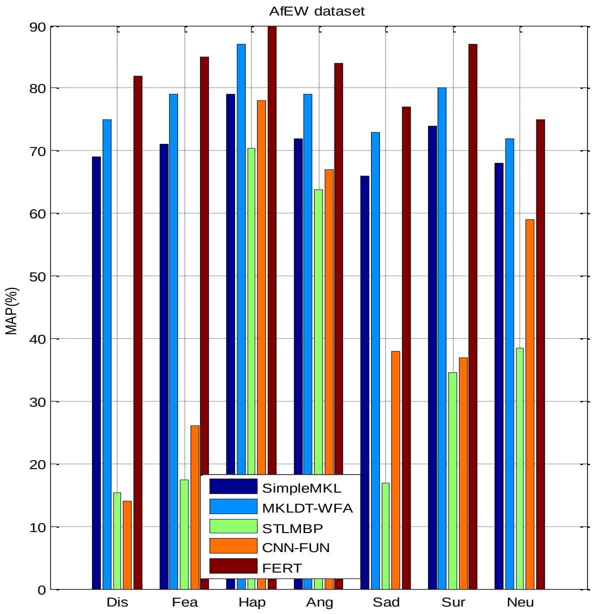


Figure 3. The ROC curve of FERT on the SLE dataset



(a)



(b)

Figure 4. Comparison of MAP among methods on (a) CK and (b) AFEW datasets

Table 6. Recognition accuracies of different methods on AFEW datasets

Method	Accuracy on AFEW
Method of Dhall *et al*. [48]	33.60
Method of Huang *et al*. [53]	43.40
MKLDT-WFA [15]	77.90%
Method of [54]	46.6%
STLMBP [53]	41.52%
CNN-FUN$_{DPM}$ [55]	51.60%
FERT	**82.80%**

3.2. Application of emotion technologies

It is well accepted that the emotions of learners form a significant part of the learning process. There is, therefore, a need for educational learning applications to recognize human emotions to facilitate smoother interaction between humans and computers. Not only the traditional learning settings can benefit but other areas of online learning can also make use of affective user data since delivering feedback is more personalized when emotions are involved. Nowadays, learners interact regularly with various forms of web-based collaborative learning tools such as social media platforms, massive open online courses, and cloud services which adds to prospective areas of applications.

In the past, automatic recognition of learning emotions has never been well developed. More recently, however, advances in affective computing and pattern recognition domains have presented various possibilities for detecting emotion that could be built upon in the learning domains. Although the suggested FERT scheme can be extended to an analysis of other forms of affective cues such as vocal and body language, this paper restricts to the most dominant cue (face) and provides experimental insight to achieve this goal.

4. CONCLUSION

There is a growing agreement that user interaction with e-learning systems needs to become more natural, humanlike, personalized, or even more learner-centered. One issue often encountered in revamping such capabilities is how to accurately recognize emotions. But more trivial is not how to recognize emotion, rather the need to understand that emotions rely on context. Therefore, experimental dependence on learning context is compulsory for dataset collection and assessment of the generalization (classification) performance of the recognition engine. In this regard, this study has studied a suggested FERT (result of feature analysis) scheme on contextually relevant learning emotion data, the SLE. Besides, two more conventional facial emotion datasets have also been used to evaluate the FERT and compare it side by side with recently published schemes. The results show that FERT is a more promising approach for personalized learning content when compared to other methods. In terms of feature descriptors, hybridization of descriptors enhances classification accuracy. The emotion classification performance on FERT has shown good merits and prospects for future affective educational learning frameworks.

REFERENCES

[1] Y. Zhang, Y. Tian, L. Yao, C. Duan, X. Sun, and G. Niu, "Individual differences matter in the effect of teaching presence on perceived learning: From the social cognitive perspective of self-regulated learning," *Computers & Education*, vol. 179, p. 104427, Apr. 2022, doi: 10.1016/j.compedu.2021.104427.

[2] V. Caputi and A. Garrido, "Student-oriented planning of e-learning contents for Moodle," *Journal of Network and Computer Applications*, vol. 53, pp. 115–127, Jul. 2015, doi: 10.1016/j.jnca.2015.04.001.

[3] M. Imani and G. A. Montazer, "A survey of emotion recognition methods with emphasis on E-Learning environments," *Journal of Network and Computer Applications*, vol. 147, p. 102423, Dec. 2019, doi: 10.1016/j.jnca.2019.102423.

[4] N. Islam, A. Behl, N. Jayawardena, V. Pereira, M. Del Giudice, and J. Choudrie, "Gamification and e-learning for young learners: A systematic literature review, bibliometric analysis, and future research agenda," *Technological Forecasting and Social Change*, vol. 176, p. 121445, Mar. 2022, doi: 10.1016/j.techfore.2021.121445.

[5] X. Sun, X. Zhao, B. Li, Y. Ma, R. Sutcliffe, and J. Feng, "Dynamic Key-Value Memory Networks With Rich Features for Knowledge Tracing," *IEEE Transactions on Cybernetics*, pp. 1–7, 2021, doi: 10.1109/TCYB.2021.3051028.

[6] F. Alqahtani, S. Katsigiannis, and N. Ramzan, "Using Wearable Physiological Sensors for Affect-Aware Intelligent Tutoring Systems," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3366–3378, Feb. 2021, doi: 10.1109/JSEN.2020.3023886.

[7] C. A. Tecson and M. M. T. Rodrigo, "A Self-paced Tutoring System, ILSA: Which System Utilization aligned with Learners' Goal Orientations?," in *2021 9th International Conference on Information and Education Technology (ICIET)*, Mar. 2021, pp. 14–20, doi: 10.1109/ICIET51873.2021.9419633.

[8] B. De Carolis, F. D'Errico, N. Macchiarulo, M. Paciello, and G. Palestra, "Recognizing Cognitive Emotions in E-Learning Environment," in *Communications in Computer and Information Science*, vol. 1344, Springer Science and Business Media Deutschland GmbH, 2021, pp. 17–27.

[9] C.-M. Chen and H.-P. Wang, "Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance," *Library & Information Science Research*, vol. 33, no. 3, pp. 244–255, Jul. 2011, doi: 10.1016/j.lisr.2010.09.010.

[10] M. Bouhlal, K. Aarika, R. AitAbdelouahid, S. Elfilali, and E. Benlahmar, "Emotions recognition as innovative tool for improving students' performance and learning approaches," *Procedia Computer Science*, vol. 175, pp. 597–602, 2020, doi: 10.1016/j.procs.2020.07.086.

[11] S. D'Mello, A. Kappas, and J. Gratch, "The Affective Computing Approach to Affect Measurement," *Emotion Review*, vol. 10, no. 2, pp. 174–183, Apr. 2018, doi: 10.1177/1754073917696583.

[12] X. Huang *et al.*, "Multi-modal emotion analysis from facial expressions and electroencephalogram," *Computer Vision and Image Understanding*, vol. 147, pp. 114–124, Jun. 2016, doi: 10.1016/j.cviu.2015.09.015.

[13] K. Bahreini, R. Nadolski, and W. Westera, "Towards real-time speech emotion recognition for affective e-learning," *Education and Information Technologies*, vol. 21, no. 5, pp. 1367–1386, Sep. 2016, doi: 10.1007/s10639-015-9388-2.

[14] J. Xiaoqing, X. Kewen, L. Yongliang, and B. Jianchuan, "Noisy speech emotion recognition using sample reconstruction and multiple-kernel learning," *Journal of China Universities of Posts and Telecommunications*, vol. 24, no. 2, pp. 1,17-9, Apr. 2017, doi: 10.1016/S1005-8885(17)60193-6.

[15] O. K. Akputu, K. P. Seng, Y. Lee, and L. M. Ang, "Emotion recognition using multiple kernel learning toward E-learning applications," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 14, no. 1, pp. 1–20, Jan. 2018, doi: 10.1145/3131287.

[16] G. Tonguç and B. Ozaydın Ozkara, "Automatic recognition of student emotions from facial expressions during a lecture," *Computers & Education*, vol. 148, p. 103797, Apr. 2020, doi: 10.1016/j.compedu.2019.103797.

[17] M. Kas, Y. El Merabet, Y. Ruichek, and R. Messoussi, "New framework for person-independent facial expression recognition combining textural and shape analysis through new feature extraction approach," *Information Sciences*, vol. 549, pp. 200–220, Mar. 2021, doi: 10.1016/j.ins.2020.10.065.

[18] W. Li, Z. Zhang, and A. Song, "Physiological-signal-based emotion recognition: An odyssey from methodology to philosophy," *Measurement: Journal of the International Measurement Confederation*, vol. 172, p. 108747, Feb. 2021, doi: 10.1016/j.measurement.2020.108747.

[19] D. Huang, S. Chen, C. Liu, L. Zheng, Z. Tian, and D. Jiang, "Differences first in asymmetric brain: A bi-hemisphere discrepancy convolutional neural network for EEG emotion recognition," *Neurocomputing*, vol. 448, pp. 140–151, Aug. 2021, doi: 10.1016/j.neucom.2021.03.105.

[20] S. N. Fatima and E. Erzin, "Use of affect context in dyadic interactions for continuous emotion recognition," *Speech Communication*, vol. 132, pp. 70–82, Sep. 2021, doi: 10.1016/j.specom.2021.05.010.

[21] K. Bahreini, R. Nadolski, and W. Westera, "Towards multimodal emotion recognition in e-learning environments," *Interactive Learning Environments*, vol. 24, no. 3, pp. 590–605, Apr. 2016, doi: 10.1080/10494820.2014.908927.

[22] S. Zhang, X. Tao, Y. Chuang, and X. Zhao, "Learning deep multimodal affective features for spontaneous speech emotion recognition," *Speech Communication*, vol. 127, pp. 73–81, Mar. 2021, doi: 10.1016/j.specom.2020.12.009.

[23] O. K. Akputu and A. O. Adedolapo, "Emotion Recognition for User Centred E-Learning," *IEEE 40th Annual Computer Software and Applications Conference*, 2016, doi: 10.1109/COMPSAC.2016.106.

[24] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1. pp. 71–86, 1991, doi: 10.1162/jocn.1991.3.1.71.

[25] F. Tabassum, M. Imdadul Islam, R. Tasin Khan, and M. R. Amin, "Human face recognition with combination of DWT and machine learning," *Journal of King Saud University - Computer and Information Sciences*, Feb. 2020, doi: 10.1016/j.jksuci.2020.02.002.

[26] S. Liu, S. Guo, W. Wang, H. Qiao, Y. Wang, and W. Luo, "Multi-view laplacian eigenmaps based on bag-of-neighbors for RGB-D human emotion recognition," *Information Sciences*, vol. 509, pp. 243–256, Jan. 2020, doi: 10.1016/j.ins.2019.08.035.

[27] Y. Yan, Z. Zhang, S. Chen, and H. Wang, "Low-resolution facial expression recognition: A filter learning perspective," *Signal Processing*, vol. 169, p. 107370, Apr. 2020, doi: 10.1016/j.sigpro.2019.107370.
[28] H. A. Q. MAARIF, "Adaptive Language Processing Unit for Malaysian Sign Language Synthesizer," *IAES International Journal of Robotics and Automation (IJRA)*, vol. 10, no. 4, pp. 326–339, Dec. 2021, doi: 10.11591/ijra.v10i4.pp326-339.
[29] K. R. Premlatha and T. V. Geetha, "Learning content design and learner adaptation for adaptive e-learning environment: a survey," *Artificial Intelligence Review*, vol. 44, no. 4, pp. 443–465, Dec. 2015, doi: 10.1007/s10462-015-9432-z.
[30] P. Brusilovsky, S. Somyurek, J. Guerra, R. Hosseini, V. Zadorozhny, and P. J. Durlach, "Open Social Student Modeling for Personalized Learning," *IEEE Transactions on Emerging Topics in Computing*, vol. 4, no. 3, pp. 450–461, Jul. 2016, doi: 10.1109/TETC.2015.2501243.
[31] S. Fatahi, "An experimental study on an adaptive e-learning environment based on learner's personality and emotion," *Education and Information Technologies*, vol. 24, no. 4, pp. 2225–2241, Jul. 2019, doi: 10.1007/s10639-019-09868-5.
[32] S. Graf *et al.*, "Adaptivity and Personalization in Learning Systems based on Students' Characteristics and Context," in *The 1st international symposium on smart learning environment*, 2012, pp. 33–36.
[33] M. Taub, R. Azevedo, R. Rajendran, E. B. Cloude, G. Biswas, and M. J. Price, "How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system?," *Learning and Instruction*, vol. 72, p. 101200, Apr. 2021, doi: 10.1016/j.learninstruc.2019.04.001.
[34] S. Afzal and P. Robinson, "Modelling affect in learning environments: Motivation and methods," in *Proceedings - 10th IEEE International Conference on Advanced Learning Technologies, ICALT 2010*, Jul. 2010, pp. 438–442, doi: 10.1109/ICALT.2010.127.
[35] R. Pekrun, "The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice," *Educational Psychology Review*, vol. 18, no. 4, pp. 315–341, Nov. 2006, doi: 10.1007/s10648-006-9029-9.
[36] C. M. Kim and R. Pekrun, "Emotions and motivation in learning and performance," in *Handbook of Research on Educational Communications and Technology: Fourth Edition*, Springer New York, 2014, pp. 65–75.
[37] J. A. Russell, "A circumplex model of affect.," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.
[38] S. Craig, A. Graesser, J. Sullins, and B. Gholson, "Affect and learning: An exploratory look into the role of affect in learning with AutoTutor," *Journal of Educational Media*, vol. 29, no. 3, pp. 241–250, Oct. 2004, doi: 10.1080/1358165042000283101.
[39] B. Kort, R. Reilly, and R. W. Picard, "An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion," in *Proceedings - IEEE International Conference on Advanced Learning Technologies, ICALT 2001*, 2001, pp. 43–46, doi: 10.1109/ICALT.2001.943850.
[40] O. K. Akputu, Y. Lee, and K. P. Seng, "Comparative analysis of multiple kernel learning on learning emotion recognition," in *Proceedings of the 6th International Conference on Information Technology and Multimedia*, Nov. 2014, pp. 357–362, doi: 10.1109/ICIMU.2014.7066659.
[41] J. G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 7, pp. 1169–1179, Jul. 1988, doi: 10.1109/29.1644.
[42] K. Delac, M. Grgic, and S. Grgic, "Independent comparative study of PCA, ICA, and LDA on the FERET data set," *International Journal of Imaging Systems and Technology*, vol. 15, no. 5, pp. 252–260, 2005, doi: 10.1002/ima.20059.
[43] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, 2004, pp. 41–48, doi: 10.1145/1015330.1015424.
[44] G. V. Sai Prasanna, K. Pavani, and M. Kumar Singh, "Spliced images detection by using Viola-Jones algorithms method," *Materials Today: Proceedings*, Jul. 2021, doi: 10.1016/j.matpr.2021.06.300.
[45] T. Senechal, V. Rapp, H. Salam, R. Seguier, K. Bailly, and L. Prevost, "Facial action recognition combining heterogeneous features via multikernel learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 4, pp. 993–1005, Aug. 2012, doi: 10.1109/TSMCB.2012.2193567.
[46] C.-C. Tsai, Y.-Z. Chen, and C.-W. Liao, "Interactive emotion recognition using Support Vector Machine for human-robot interaction," in *2009 IEEE International Conference on Systems, Man and Cybernetics*, Oct. 2009, pp. 407–412, doi: 10.1109/ICSMC.2009.5346180.
[47] T. Pumlumchiak and S. Vittayakorn, "Facial expression recognition using local Gabor filters and PCA plus LDA," in *2017 9th International Conference on Information Technology and Electrical Engineering, ICITEE 2017*, Oct. 2017, vol. 2018-Janua, pp. 1–6, doi: 10.1109/ICITEED.2017.8250446.
[48] A. Dhall, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon, "Emotion Recognition In The Wild Challenge 2014," in *Proceedings of the 16th International Conference on Multimodal Interaction*, Nov. 2014, pp. 461–466, doi: 10.1145/2663204.2666275.
[49] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings - 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000*, 2000, pp. 46–53, doi: 10.1109/AFGR.2000.840611.
[50] J. Gardner and C. Brooks, "Dropout Model Evaluation in MOOCs," Feb. 2018, [Online]. Available: http://arxiv.org/abs/1802.06009.
[51] M. Janicka, M. Lango, and J. Stefanowski, "Using information on class interrelations to improve classification of multiclass imbalanced data: A new resampling algorithm," *International Journal of Applied Mathematics and Computer Science*, vol. Vol. 29, no. 4, pp. 769–781, Dec. 2019, doi: 10.2478/AMCS-2019-0057.
[52] Z. Sun, Z.-P. Hu, R. Chiong, M. Wang, and W. He, "Combining the Kernel Collaboration Representation and Deep Subspace Learning for Facial Expression Recognition," *Journal of Circuits, Systems and Computers*, vol. 27, no. 08, p. 1850121, Jul. 2018, doi: 10.1142/S0218126618501219.
[53] X. Huang, Q. He, X. Hong, G. Zhao, and M. Pietikainen, "Improved Spatiotemporal Local Monogenic Binary Pattern for Emotion Recognition in The Wild," in *Proceedings of the 16th International Conference on Multimodal Interaction*, Nov. 2014, pp. 514–520, doi: 10.1145/2663204.2666278.
[54] H. Yan, "Collaborative discriminative multi-metric learning for facial expression recognition in video," *Pattern Recognition*, vol. 75, pp. 33–40, Mar. 2018, doi: 10.1016/j.patcog.2017.02.031.
[55] H. Kaya, F. Gürpınar, and A. A. Salah, "Video-based emotion recognition in the wild using deep transfer learning and score fusion," *Image and Vision Computing*, vol. 65, pp. 66–75, Sep. 2017, doi: 10.1016/j.imavis.2017.01.012.

BIOGRAPHIES OF AUTHORS

Dr. Akputu, Oryina Kingsley holds a PhD in Computing from Sunway University Malaysia, a M.Sc. of Computer Science from The University of Nottingham UK and a B.Sc. (Hons) in Mathematics and Computer Science from the Federal University of Agriculture Makurdi, Nigeria. At the time of writing, he was with Admiralty University of Nigeria as Lecturer and Coordinator of the Computer Science Programs. He worked previously with Trinity University Lagos, as a Lecturer and Acting Head (Ag.), Department of Computer and Information Sciences. And with Ritman University as the Ag. Head, Department of Mathematics & Computer Science, and Coordinator, Office of Research and Innovation. He has served as Teaching/ Research Assistance with Sunway University, under the Affective and Assistive Computing Laboratory. His research interests span, Pattern Recognition, Signal Processing, Affective Computing, Information Systems, Adaptive eLearning, Big Data Analytics and UX Evulation. He can be contacted at email: oryinak@gmail.com.

Associate Professor Udoinyang G. Inyang received the B. Sc. Degree in Computer Science from the University of Uyo, Nigeria in 1996. He received the M.Sc. and Ph.D degrees in Computer Science from the University of Technology Akure, in 2005 and 2012 respectively. He is with the Department of Computer Science at the University of Uyo and has many publications in reputable journals. His research interest includes Soft Computing, Data Minning, Knowledge Engineering and Speech Technology. At the time of writing, he was the Director of Tertiary Institution Trust Fund (TETFUND) CENTRE OF EXCELLENCE in Computational Intelligence, University of Uyo Chapter. He is a Member of IEEE, Computer Professionals Registration Council of Nigeria (CPN) and Nigerian Computer Society among other professional bodies. He can be contacted at email: udoinyanginyang@uniuyo.edu.ng, udoiinyang@yahoo.com.

Ortil Msugh recieved a M.Sc. degree in Computer Science from Staffordshire University UK, and a B. Sc. degree in Mathematics/Statistics/Computer Science Education from Federal University of Agriculture Makurdi, Nigeria. His research interests include machine learning, Image Processing and Pattern Recognition. He can be contacted at email: ortil.msugh@gmail.com.

F. T. Mughal did her Bachelors (Hons) in Information Technology. She is doing her graduate studies at the University of Technology Sydney, under the Games Studio. Her Research Interest spans, Geriatric mental health, machine learning and affect recognition. She has publications in reputable journals and conferences. She can be contacted at email: fizatariq.mughal@student.uts.edu.au.

Professor Abel Usoro researches in Information Systems which include e-learning and knowledge management. He worked at School of Computing, Engineering and Physical Sciences at the University of the West of Scotland and at the time of writing was a visiting Professor in Akaw Ibom State University, Visiting Reader in the University of Calabar, all in Nigeria. He is the founder and President of the International Informational Technology and Economic Development Association (IITEDA), a research, development and innovation association that partners universities. He has widely published and served in the scientific committee of several international refereed journals and conferences, some of which he chairs. He can be contacted at email: abel.usoro@uws.ac.uk