Recognizing facial emotions for educational learning settings

Oryina K. Akputu¹, Udoinyang G. Inyang², Msugh Ortì³, Fiza T. Mughal⁴, Abel Usoro⁵

¹Department of Computing Sciences, Admiralty University of Nigeria, Delta State, Nigeria.
²Department of Computer Science, University of Uyo, Uyo, Nigeria.
³Department of Computer Science, FCT College of Education Zuba, Abuja, Nigeria.
⁴Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia.
⁵School of Engineering and Computing, University of the West of Scotland Paisley, Renfrewshire, United Kingdom.

Article Info

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 constrained 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.

Keywords:
Contextual dataset
Educational learning
Emotions recognition
Learning emotions
Personalization

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 effortlessly 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 pre-defined 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 works 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.
2.4.1. Feature extraction using Gabor wavelet

The 2D Gabor filters [41], are utilized to extract features of every detected facial image. Let, \( l(x, x') \) 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, \( S_0 = \{ \psi_{\theta_0, \tau}(l); \mu \in [0,1,2,3,4,5], \tau = [0,1,2,3] \} \), contains the Gabor wavelet filter representation of the image, \( l(x, x') \). 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 \( \times \) 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).

\[
V^{(d)} = \left[(\psi_{0,0}(l), \psi_{0,2}(l)) \right]_{0,0}^{\psi} (l)^T = [V_1, \ldots, V_{24}]_1^{\theta}
\]  

(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., \( \{p_1, p_2, \ldots, p_P\}_{i=1}^{\theta} \), \( p_i = \{x_1, x_2, \ldots, x_n\}_{i=1}^{\theta} \). 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, \( v_i \in \mathbb{R}^{M \times N} \), such that, a resultant scatter of transform features spaces \( \{v_1, \ldots, v_{24}\}_{i=1}^{\theta} \), becomes [42].

\[
W^TS_iW
\]  

(2)

Where \( W \in \mathbb{R}^{N \times M} \) and \( S_i \) 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).

\[
W_o = \arg \max_{W} \left| W^TS_iW \right| = [w_s, w_2, \ldots, w_m]
\]  

(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 \( S_i \). 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,

\[
S_B = \sum_{i=1}^{N} N (\theta_i - \bar{\theta})(\theta_i - \bar{\theta})^T
\]  

(4)

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

\[
S_W = \sum_{i=1}^{n} \sum_{j \in p_i} (x_i - \theta_j)(x_i - \theta_j)^T
\]  

(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).
\[
W_o = \arg \max_W \ |W^T S_b W| \ = [w_1, w_2, \ldots, w_m].
\]

The matrix \(w_i\)s 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).

\[
S_bW_i = \lambda_iS_wW_i, i = 1,2,\ldots,m
\]

### 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 pattern to diverse emotional classes. The objective in context is to designate \(p\)-classes emotion classes. Classification with MKLDT-WFA begins from the root node \(n_1\) of the DT routine. The \(p\)-classes of a father node are divided into binary disjoint clusters or child nodes or non-leaf node \(n_1\). Consider the cluster group a node to be \(\mathcal{C}_i\) and \(\mathcal{C}_j\) each containing at least one class or possibly multiple classes. The approach realized the grouping of the data using the following distance measure [15].

\[
d_{i,j}(x_{e,i}, c_j(n_i)) = \sqrt{K_{x_ix_i} - 2K_{x_ic_j}(n_i) + K_{c_jc_j}(n_i)}, i,j = 1,\ldots,N
\]

Where,

\[
d_{i,j} = \begin{cases} d_{i,j}(x_{e,i}, c_j(n_i)) & \text{if } j \neq i \\ 0, \text{other.} \end{cases}
\]

The \(c_j\) is the cluster center at a node \(n_i\); \(K_{x_ix_i}\) and \(K_{c_ic_j}\) 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).

\[
f(x_{e,i}) = \sum_{q=1}^{Q} \sum_{i,j=1}^{N_E} \sum_{m=1}^{M} d_{i,j}(x_{e,i}, c_j(n_i)) \alpha_j y_j K_m(x_{e,p}, x_{e,i}) + b
\]

The terms \(a^*\) 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 \(d_m, a^*_m\) and \(b\) using the objective function [15].

\[
\max_d \sum_{q=1}^{Q} \sum_{i,j=1}^{N_E} \sum_{m=1}^{M} d_{i,j}(x_{e,i}, c_j(n_i))\]

Where,

\[
j(d) = \begin{cases} \max_{a} \sum_{q=1}^{Q} \alpha_{i,q} - \frac{1}{2} \sum_{j=1}^{N_E} \alpha_{i,q} \alpha_{j,q} y_j y_j \sum_{m=1}^{M} d_m K_m(x_{e,p}, x_{e,i}) & \text{if } j \neq i \\ 0, \text{other.} \end{cases}
\]

Notably,

\[
\sum_{j=1}^{N_E} \alpha_{i,q} y_j = \forall i,j = 1,\ldots,N
\]

### 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.

![Sample images of datasets](image)

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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SLE</th>
<th>CK+</th>
<th>AFEW 4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of emotion classes</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Size of dataset</td>
<td>1350 (225 videos x6)</td>
<td>5831 images</td>
<td>1,368</td>
</tr>
<tr>
<td>No. of actors</td>
<td>25</td>
<td>123</td>
<td>428</td>
</tr>
<tr>
<td>Size of train set</td>
<td>810 (135 videos x6)</td>
<td>3,498</td>
<td>3,468 (578 x 6)</td>
</tr>
<tr>
<td>Size of test set</td>
<td>540 (90 videos x 6)</td>
<td>2,332</td>
<td>2,442 (407 x 6)</td>
</tr>
<tr>
<td>Size of a validation set</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td>2,298 (383x 6)</td>
</tr>
</tbody>
</table>

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 the unwanted 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 recognition schemes on the same datasets. In Figure 4(a), the MAP score reported by most methods including, EDR-PCANet [52] and MKDLT-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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>PCA</th>
<th>$PCA + LDA$</th>
<th>$PCA + LDA + Gabor$</th>
<th>Gabor</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLE</td>
<td>83.3</td>
<td>80.4</td>
<td>81.1</td>
<td>77.9</td>
<td>82.0</td>
</tr>
<tr>
<td>AFEW</td>
<td>77.9</td>
<td>80.4</td>
<td>81.1</td>
<td>77.9</td>
<td>82.0</td>
</tr>
<tr>
<td>Cohn Kanade+</td>
<td>80.4</td>
<td>81.1</td>
<td>81.1</td>
<td>77.9</td>
<td>82.0</td>
</tr>
</tbody>
</table>

Table 3. Confusion matrix of FERT on SLE

<table>
<thead>
<tr>
<th>Truth</th>
<th>Eng</th>
<th>Hop</th>
<th>Bor</th>
<th>Hap</th>
<th>Fru</th>
<th>Sur</th>
<th>Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng</td>
<td></td>
<td></td>
<td>97</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>99.3</td>
</tr>
<tr>
<td>Hop</td>
<td>88</td>
<td>94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bor</td>
<td>98</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>98.0</td>
</tr>
<tr>
<td>Hap</td>
<td>99</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>99.0</td>
</tr>
<tr>
<td>Fru</td>
<td>98</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>98.0</td>
</tr>
<tr>
<td>Sur</td>
<td>99</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>99.0</td>
</tr>
</tbody>
</table>

Table 4. Confusion matrix of FERT on CK+

<table>
<thead>
<tr>
<th>Truth</th>
<th>Dis</th>
<th>Fea</th>
<th>Hop</th>
<th>Ang</th>
<th>Sad</th>
<th>Sur</th>
<th>Con</th>
<th>Precision</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>79</td>
<td>96.0</td>
</tr>
<tr>
<td>Fea</td>
<td>88</td>
<td>93</td>
<td>97</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Hop</td>
<td>98</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td>96.0</td>
</tr>
<tr>
<td>Ang</td>
<td>99</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>99.0</td>
</tr>
<tr>
<td>Sad</td>
<td>98</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>98.0</td>
</tr>
<tr>
<td>Sur</td>
<td>99</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>99</td>
<td>99.0</td>
</tr>
</tbody>
</table>

Recognizing facial emotions for educational learning settings (Oryina K. Akputu)
Table 5. Confusion matrix of FERT on AFEW

<table>
<thead>
<tr>
<th></th>
<th>Dis</th>
<th>Fea</th>
<th>Hap</th>
<th>Ang</th>
<th>Sad</th>
<th>Sur</th>
<th>Neu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dis</td>
<td>303</td>
<td>1</td>
<td>7</td>
<td>18</td>
<td>21</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Fea</td>
<td>13</td>
<td>259</td>
<td>12</td>
<td>13</td>
<td>8</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Hap</td>
<td>0</td>
<td>17</td>
<td>306</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Ang</td>
<td>19</td>
<td>13</td>
<td>0</td>
<td>277</td>
<td>29</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Sad</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>11</td>
<td>258</td>
<td>4</td>
<td>56</td>
</tr>
<tr>
<td>Sur</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>312</td>
<td>2</td>
</tr>
<tr>
<td>Neu</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>10</td>
<td>2</td>
<td>308</td>
</tr>
<tr>
<td>precision</td>
<td>82</td>
<td>85</td>
<td>90</td>
<td>84</td>
<td>77</td>
<td>87</td>
<td>75</td>
</tr>
</tbody>
</table>

Accuracy 82.8

Figure 3. The ROC curve of FERT on the SLE dataset
Recognizing facial emotions for educational learning settings (Oryina K. Akputu)

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

Table 6. Recognition accuracies of different methods on AFEW datasets

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy on AFEW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method of Dhall et al. [48]</td>
<td>33.60</td>
</tr>
<tr>
<td>Method of Huang et al. [53]</td>
<td>43.40</td>
</tr>
<tr>
<td>MKLDT-WFA [15]</td>
<td>77.90%</td>
</tr>
<tr>
<td>Method of [54]</td>
<td>46.6%</td>
</tr>
<tr>
<td>STLMBP [53]</td>
<td>41.52%</td>
</tr>
<tr>
<td>CNN-FUN [55]</td>
<td>51.60%</td>
</tr>
<tr>
<td>FERT</td>
<td>82.80%</td>
</tr>
</tbody>
</table>

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


Recognizing facial emotions for educational learning settings (Oryna K. Akputu)
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 Evaluation. 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 Mining, 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: udoinyanginya@uniuyo.edu.ng, udoinyang@yahoo.com.

Ortil Msugh received 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.