Affective State Detection via Facial Expression Analysis within a Human-Computer Interaction Context

Anas Samara · Leo Galway · Raymond Bond · Hui Wang

Abstract The advancement in technology indicates that there is an opportunity to enhance Human-Computer Interaction by way of affective state recognition. Affective state recognition is typically based on ***passive*** stimuli such as watching video clips, which does not reflect genuine interaction. This paper presents a study on affective state recognition using ***active*** stimuli, i.e. facial expressions of users when they attempt computerised tasks, particularly across typical usage of computer systems. A data collection experiment is presented for acquiring data from normal users whilst they interact with software, attempting to complete a set of predefined tasks. In addition, a hierarchical machine learning approach is presented for facial expression-based affective state recognition, which employs an Euclidean distance-based feature representation, conjointly with a customised encoding for users' self-reported affective states. Consequently, the aim is to find the potential relationship between the facial expressions, as defined by Paul Ekman, and the self-reported emotional states specified by users using Russells Circumplex model, in relation to the actual feelings and affective states. The main findings of this study suggest that facial expressions cannot precisely reveal the actual feelings of users whilst interacting with common computerised tasks. Moreover, during ***active*** interaction tasks more variation occurs within the facial expressions of participants than occurs within ***passive*** interaction.

Keywords Emotions · Human Computer Interaction · Facial Expression

Anas Samara
School of Computing, Ulster University, Belfast BT37 0QB, UK
E-mail: samara-a@ulster.ac.uk

Leo Galway
School of Computing, Ulster University, Belfast BT37 0QB, UK

Raymond Bond
School of Computing, Ulster University, Belfast BT37 0QB, UK

Hui Wang
School of Computing, Ulster University, Belfast BT37 0QB, UK

1 Introduction

In human-human interaction, one can intuitively predict the emotional state based on observations about persons facial expressions, body behaviour, and voice intonations (Karray et al., 2008). This ability is essential as humans often adapt their own behaviour based on such inferences. Correspondingly, in computer systems, various input modalities exist that can be utilised to acquire information about the user. Firstly, audio and visual-based input modalities, such as eye gaze tracking, facial expressions, body movement detection, and speech and auditory analysis may be employed as additional inputs. Secondly, physiological input modalities using sensor-based signals, such as electroencephalogram (EEG), galvanic skin response, and electrocardiogram can also be utilised. Alternatively, extra inputs may be gained by interpreting user behaviour with the mouse movements, keyboard keystrokes and content viewing, or even a combination of all of these different modalities (Duric et al., 2002).

This paper describes a Human-Computer Interaction (HCI) data collection study that has been developed to collect data from several input modalities, in order to be used for studying and modelling different characteristics of users across typical HCI context, where users interact with varied computer applications. Subsequently, the paper investigates automatic affective state techniques using analysis of facial expressions from static images, where an ensemble hierarchical classification approach with feature representation based on Euclidean distance has been used alongside various user-based self assessment mapping schemes. In addition, it aims to identify the affective states that users are experiencing whilst interacting with multiple graphical user interface-based applications and attempting different computerised tasks, in conjunction with validating related self-assessment and self-reporting techniques. Accordingly, the work presented herein contributes to the body of literature on automated facial expression recognition.

The sections of the paper are organised as follows. Section 2 gives a brief background on the research area and related work. Section 3 introduces and discusses available facial expression datasets with relevant work on data collection. Section 4 outlines about the methodology employed by the techniques used in the associated experiments carried out, as well as the details of the data collection protocol utilised during the experiments. Section 5 presents the experiments together with their results. Section 6 concludes the paper with discussion on the results and a summary of the findings.

2 Background and Related Work

A variety of models have been developed for modelling innate human feelings; one such model is the *Circumplex Model*, proposed by Russell (Russell, 1980), which asserts that the affective state of human feeling can be considered as a point in two dimensional space, as illustrated in Figure 1. In particular, this model is composed of *valence* and *arousal* intensity dimensions, where-
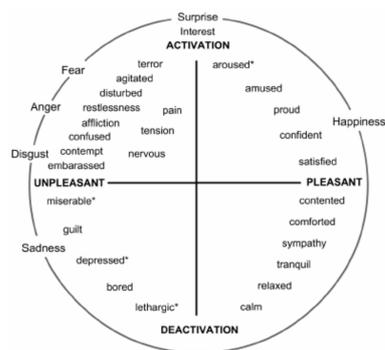
Fig. 1: *Valence* (pleasant-unpleasant continuum) versus *Arousal* (activation-deactivation continuum), in the Circumplex Model (Russell and Lemay, 2000).

by *valence* represents the intrinsic attractiveness or averseness of an emotion, and can be presented as a pleasant-unpleasant continuum (Frijda, 1986), and *arousal* is the physiological and psychological state that activates the alertness, consciousness and attention as a reaction to stimuli, and can be presented as an activation-deactivation continuum (Coull, 1998; Robbins, 1997). Consequently, results from the experiments carried out and presented in this paper were analysed by means of *valence* and *arousal* separately and conjointly. Within the research literature, facial expression recognition is an active area of research despite the number of successes in this domain. For instance, studies where machine learning techniques trained on features extracted using different approaches achieved a reasonable classification accuracy such as the work of (Liew and Yairi, 2015) who achieved accuracy of 91.2% on CK-Plus dataset, using Support Vector Machine trained on Histogram of Oriented Gradients. Furthermore, recent progress with the evolution of Deep Learning techniques has produced a high classification accuracy of 96.76% on the same dataset, using Convolutional Neural Networks, as shown in the work presented by (Lopes et al., 2017). Consequently, the work presented herein may be considered as comparable in terms of facial expression classification accuracy, whereas this work has been validated against multiple benchmark datasets.

3 Datasets

In terms of facial expression data, there are plenty of available datasets. For instance, Lucey et al. published a benchmark dataset entitled as Extended Cohn-Kanade (CK-Plus) database, which comprises 593 sequences taken from 123 subjects expressing the emotional states of: *angry, contempt, disgust, fear, happy, sadness* and *surprise* (Lucey et al., 2010). In a similar manner, Lundqvist et al. published the Karolinska Directed Emotional Faces dataset (KDEF), which consists of 4900 pictures captured from 70 subjects, whereby each subject acted seven different affective states: *afraid, angry, disgust,*

happy, neutral, sad and *surprise* (Lundqvist et al., 1998). Furthermore, Siddiqi et al. had collected a dataset in which the expressions represent real life situations including emulated, semi-naturalistic, and naturalistic expressions (Siddiqi et al., 2017). Moreover, another multimodal dataset entitled DEAP, is a dataset for Emotion Analysis using EEG, Physiological and Video Signals for 32 subjects (frontal face video was recorded for only 22), who individually watched 40 one-minute music videos of different genres as a stimulus, to induce different affective and emotional states (Koelstra et al., 2012). Generally, such datasets could be used for the analysis of affective states, along with the assessment and self-reports made by the subjects themselves, such as using the Self-Assessment Manikin (SAM) scales (Bradley and Lang, 1994), to typify the awareness of subjects' feelings. Consequently, the work presented herein investigates the relationship between facial expressions, as defined by Paul Ekman, and affective states. Particularly, the work focuses on self-reporting using Russell's Circumplex model that is used as a means to represent the actual experienced affective state (Russell, 1980).

4 Methodology

4.1 Feature Extraction and Distance-Based Representation

Geometric-based techniques for facial expression analysis are based on locating the facial points and determining the location and the shape of associated facial components, including the eyebrows, eyes, nose, lips and mouth. The study presented in this paper used the "*Chehra*" tool in order to extract the location of 49 facial points, as illustrated in Figure 2(a) (Asthana et al., 2014). *Chehra* is a publicly available facial landmark detector based on discriminative facial deformable models, trained using a cascade of linear regressions. In addition, this tool is capable to locate the facial points automatically in real time and handle faces under uncontrolled natural setting. Subsequently, the detected landmarks are represented as the Cartesian coordinates of the extracted facial points, which produce a 98-dimension feature vector that is used in the classification stage of the facial expression analysis pipeline, as mentioned in the experiment section.

Feature representation using Cartesian coordinates yielded a good classification accuracy. Nevertheless, this approach is not robust enough to permit recognition of facial expressions not provided in the training data. This is due to the fact that the constellation of these points varies among the myriad of facial shapes that comprise different facial morphologies (Salah et al., 2010). Therefore, we investigated another feature representation, which is primarily based on finding the *configural* features that represent intra-facial component distances, as shown in Figure 2(b) (Martinez, 2011). Subsequently, resultant features are represented by finding the *Euclidean* distances among all facial landmark points. Consequently, the facial expression is finally represented as a 1176-dimension feature vector, resulting from 49 Cartesian coordinate combinations (Samara et al., 2016).
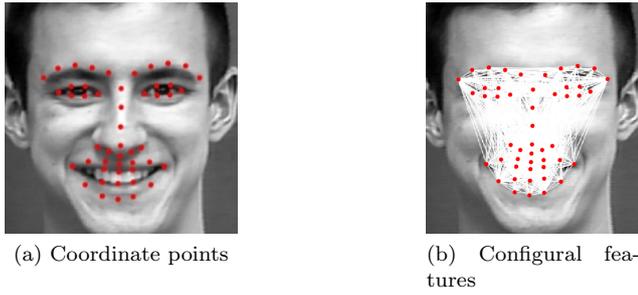
(a) Coordinate points

(b) Configural features

Fig. 2: Configural features of facial point coordinates.

4.2 Hierarchical Parallelised Binary Support Vector Machines

Within the research literature, a range of classification techniques have been used in Affective Computing and emotion recognition using facial expressions as input vectors (Corneanu et al., 2016). In this study, we used a Support Vector Machine (SVM), specifically the C-Support Vector with a linear kernel, available from the LIBSVM library developed at National Taiwan University (Chang and Lin, 2001; EL-Manzalawy, 2005). Furthermore, in this study we have investigated an approach that combines a set of SVM classifiers in order to improve the overall performance of the system. Initially, a number of labelled datasets (equal to the number of labels, i.e. classes) were produced from the original dataset. Accordingly, each dataset contains a pair of labels: either one of the labels from the original dataset, which corresponds to an emotional state, or *others*. Afterwards, each dataset generated is used to train a single binary classifier. Subsequently, a number of binary classifiers are trained individually on the individual two-class datasets, which comprise the first stage of the framework. The second stage of the framework then utilises a multi-class classifier that provides the final result from the overall classification task.

Thus, the resulting classification framework, Hierarchical Parallelised Binary Support Vector Machines (HPBSVM), operates in two stages as given in Figure 3. During the first stage, binary SVM models are constructed from annotated data, with one model employed for each emotional state within the dataset. Furthermore, during the second stage, a multi-class SVM model is constructed to predict the state based on the combination of the decisions given from the binary SVM models of the first stage. The output from the first stage, which is a multiple components vector produced by the set of binary classifiers, represents the intermediate feature vector that is used to train a posteriori second stage multi-class classifier, entitled the Aggregation Classifier, which gives the final decision on the detected emotional state. This approach facilitates more efficient and accurate classification of the whole system. The whole system benefits from the advantages of making some features more discriminative for specific classes. In other words, this framework decomposes

the overall problem into smaller micro-decisions that are made by specialised classifiers, which are trained differently.
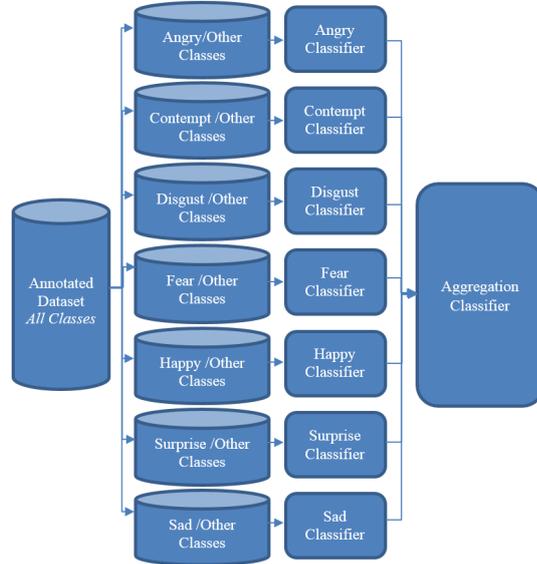

Fig. 3: Hierarchical parallelised binary support vector machines (HPBSVM) for emotion classification.

4.3 Facial Expression Datasets

Two different facial expression datasets have been selected for validating the suggested techniques: CK-Plus and KDEF, as they are considered to be benchmark datasets for facial expression analysis. These two datasets employ different labels for their annotated facial expressions. For CK-Plus, the labelling process was carried out over the peak frame using the Facial Action Coding System (FACS) (Ekman et al., 2002), and only 327 sequences meet the criteria to be labelled as a specific expression. For KDEF, the labels were given according to the emotion state that the subject was asked to attitudinise. In the experiments described later, our interest is to look at the facial expressions that may entail different affective states. Therefore, we have used another variation of CK-Plus by using *neutral* state snapshots. Therefore, three datasets have been utilised with the classifiers: CK-7, in which the sequences were divided into seven groups: *angry (45), contempt (18), disgust (59), fear (25), happy (69), sadness (28)* and *surprise (83)*; CK-8, which expands upon CK-7 by incorporating 112 additional images annotated as *neutral*; KDEF, in which we selected the frontal view of the snapshots, thereby producing an equal number of the seven states: *afraid, angry, disgust, happy, neutral, sad* and *surprise*.

4.4 Data Collection Study

The aim of the data collection study is to collect features from different input modalities, which is used to reason about users' affective states whilst interacting with common computer software and attempting to complete typical computer-based tasks. A total of 42 participants took part in the study. Participants were either staff or a student at Ulster University. There was no specific inclusion or exclusion criterion other than being a current student or staff member of the university, as the study is interested in identifying affective states in a generic HCI context, hence participants experiences could vary from novice to expert computer users.

4.4.1 Material

The material for the tasks used throughout the study can be classified into four main categories: (1) basic operating system tasks; (2) online shopping tasks; (3) Excel spreadsheet manipulation tasks, and (4) game-based tasks. Consequently, the themes of these had been chosen according to a study of computer usage statistics carried out by Thomas Beauvisage (Beauvisage, 2009), which presented the average distribution of individual weekly computer usage. Notably, the study showed that the four categories identified and selected for the study presented herein, occupy the highest percentages of time spent on computer usage for both households and individuals. Accordingly, the set of tasks represent ***active*** interaction with a computer in which the participant has a predefined task to carry out within a limited time of 5 minutes at most. Regarding the operating system task, the participant was asked within to change the desktop background, screen saver, time zone, and add a new input language to the system within the predefined time limit. For the online shopping task, the participant was asked to search online for a Tablet-PC with specific properties using their preferred Internet browser application. In terms of the spreadsheet manipulation task, the participant was asked to modify an existing Excel spreadsheet to insert new data into the existing records, sort the data in ascending order, use an aggregation function (i.e. *Average*), and draw a line graph of the data. For the game-based task, the participant was asked to play a version of the arcade game Pacman (namely Deluxe Pacman 2 - Release v16) for a total of 3 minutes.

4.4.2 Procedure

Firstly, each participant was given an information sheet describing the flow of the experiment, and then he/she was asked to sign a consent form that his/her participation in the study is completely voluntary. At this stage, the participant commenced the first task within the designated time. On task completion, the participant was asked to complete SAM scales (Bradley and Lang, 1994) to self-report the levels of *valence* and *arousal* during that task.

During the experiment, the video of participant's face was recorded using a typical webcam placed at the top of the screen.

4.5 Approach and Analysis Design

This work aims to develop a validated technique for automatic classification of emotions from facial expressions by analysing video frames that were acquired whilst users attempted and interacted with each of the computer based tasks. Hence, the main objective of the classification is to model the affective states of users within a HCI context, and the association between the self-reporting and the facial expression. Subsequently, we categorised the tasks into two groups as **active** and **passive** interactions. Firstly, **active** interaction, which represents the situations where the user is working and undertaking an effort to carry out the task. These situations include: general operating systems tasks (tagged as **OS**), online shopping tasks (tagged as **Online**), and spreadsheet tasks (tagged as **Excel**), and entertainment task playing Pacman (tagged as **Pacman**). Secondly, **passive** interaction that represents situations where the user does not exert effort, which refers to the recordings contained within the DEAP dataset (Koelstra et al., 2012), where the subjects passively watched videos on **Youtube** without explicit interaction.

Moreover, as previously described, each participant carried out a self-reported assessment after performing each task using the SAM scale; this is the case in both the dataset obtained from the study and the DEAP dataset. Thus, each recorded video from each task is associated with a *valence* and *arousal* scores. Consequently, the relationship between facial expression percentages and the self-reported *valence* and *arousal* scores given by the participants for each task was further investigated. However, during analysis, rather than using the actual SAM ranges, i.e. [1.0-9.0], a mapping of the reported values into three ranges (*Low, Medium* and *High*) was used. Correspondingly, classification performance improves during supervised learning when the number of target classes is reduced (Aha, 1992), especially when the combined classes have common properties and similarity, as is the case with our dataset. Therefore, a transformation was applied to the rounded values, as illustrated in Figure 4, whereby values within the range [1.0-3.0] are labelled as *Low*, values within the range [4.0-6.0] are labelled as *Medium*, and values within the range [7.0-9.0] are labelled as *High*. In addition to this mapping, the labels corresponding to the self-reported values of *valence* and *arousal* have been used to represent the four quadrants of the Circumplex Model, as illustrated in Figure 5, during the analysis carried out in this study.

5 Results

This section will summarise the experiment results achieved throughout the current study. Firstly, classification experiments applied on existing facial expression datasets. Secondly, percentages of facial expression that automatically
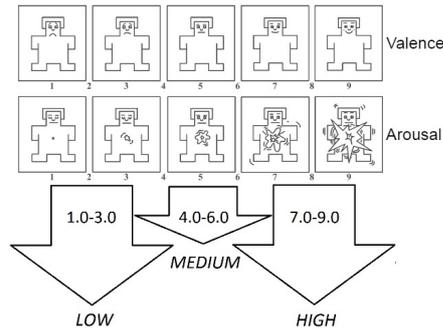
Fig. 4: Mapping from SAM scale value ranges into three labels (scores from [1.0-3.0] mapped to *Low*, [4.0-6.0] mapped to *Medium*, and [7.0-9.0] mapped to *High*).
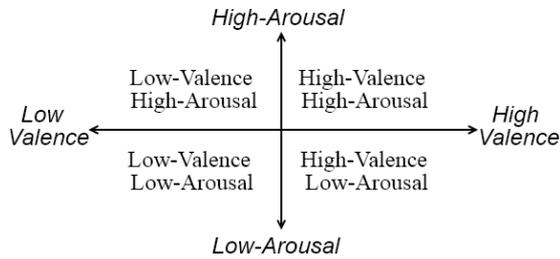

Fig. 5: Representation of Circumplex Model quadrants using combinations of *Valence* and *Arousal* score mappings, which utilises only the mappings corresponding to the *High* and *Low* labels.

applied on the video frames from the dataset obtained from the current work, which are recordings of subjects interacting with typical software interfaces. Thirdly, percentages of facial expressions across self-reporting ratings given by subjects, that are presumed to represent their actual feeling during each task, as well as, facial expression percentages according to combined ratings that are mapped to the Circumplex model quadrants. Experiments carried out herein used 10-fold cross validation. In addition, a 95% confidence interval of classification results has been used in order to show the lower and upper limits along with the statistical significance of the results obtained.

5.1 Expression Datasets Classification

We have validated the performance of HPBSVM against the normal, single classifier scheme (i.e. SVM), using both the distance-based feature descriptor and a feature vector comprising Cartesian point coordinates. Basically, distance-based feature descriptor and HPBSVM outperforms point coordinates with SVM. Yet the improvement caused by the HPBSVM is substantial, with

Table 1: Classification accuracy of different datasets using Point-coordinates/Distance-based with SVM/HPBSVM classification models. Distance-based feature with HPBSVM outperforms Point-coordinates with SVM with statistically significant improvement (P<0.001).

Dataset	Point-coordinates & SVM	Distance-based & SVM	Point-coordinates & HPBSVM	Distance-based & HPBSVM
CK-7	80.75% ± 4.28	82.86% ± 4.09	96.02% ± 2.24	96.94% ± 2.01
CK-8	76.76% ± 3.95	78.36% ± 3.85	95.22% ± 2.06	95.67% ± 1.98
KDEF	78.47% ± 2.57	81.84% ± 2.42	85.71% ± 2.20	96.22% ± 1.22

comparison to the improvement resulting from using distance-based descriptor. In brief, as given in Table 1, HPBSVM with distance-based descriptor achieves the higher accuracy rates across all three datasets. Consequently, this technique assures a credible result to be used with other data such as the one collected in our study and DEAP dataset. Therefore, we adopted this scheme in the experiments of next section.

5.2 Expression Classification during Human-Computer Interaction

HPBSVM with distance-based feature was used to train two models using the validated datasets CK-8 and KDEF, so that each resulting model can automatically classify the video frames of the recordings obtained during the computer-based tasks at a frequency of 1 frame per second. Figure 6 shows the percentages of facial expressions using the earlier mentioned trained models, which is achieved by automatically classifying the video frames of the user recording while carrying out the Online task. Figure 6(a) presents the expression percentages using a classification model that is trained on CK-8 dataset. Similarly, Figure 6(b) presents the expression percentages using a trained classifier using KDEF dataset.

As depicted in Figure 6, different percentages of each expression have been found for both models. However, one can view these percentages differently by considering the fact that some expressions are much more precisely recognised than others. Generally, detecting states such as *happy* and *surprise* is comparably superior than detecting other states such as *contempt, neutral, fear, angry, sadness* and *disgust*, which is possibly due to the similarity in the geometric shape of these expressions (Samara et al., 2016). Moreover, the work presented in (Joho et al., 2009) underlined this type of grouping, by devising the pronounce level of the associated expressions, where these expressions belong to a *Low Pronounced* level. Therefore, the expression labels *angry, contempt, disgust, fear* and *sadness* from the CK-8 dataset used in our analysis can be combined together to represent the *negative* state. Likewise, the labels *afraid, angry, disgust* and *sad* from the KDEF dataset can be combined together to represent the *negative* state. Additionally, from the Circumplex Model it may be observed that there is a common aspect among these expressions, in

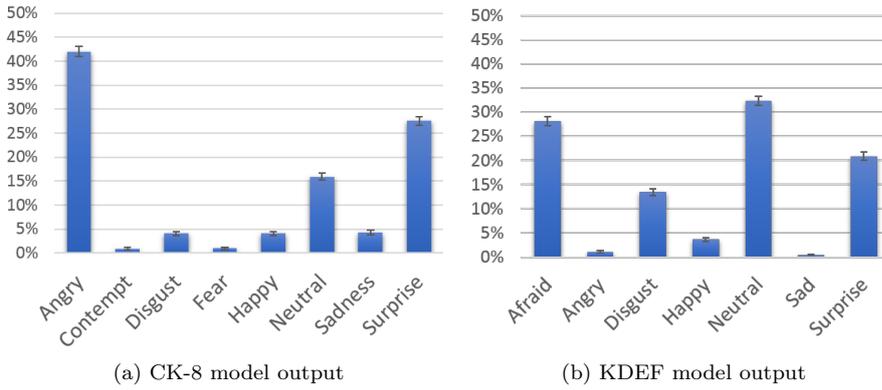(a) CK-8 model output   (b) KDEF model output

Fig. 6: Percentages of expressions on CK-8 and KDEF trained models applied on recordings of Online task context respectively.
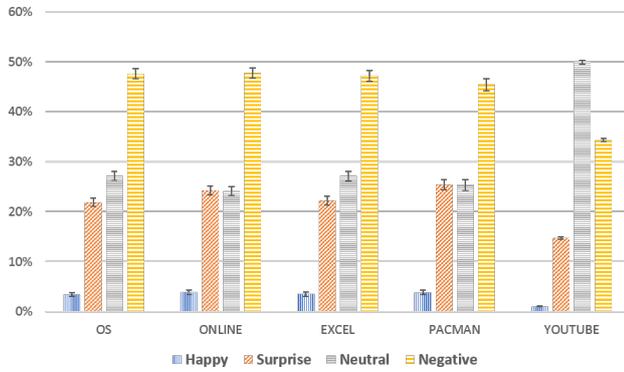

Fig. 7: Percentages of facial expressions across tasks by averaging outputs of two trained models on CK-8 and KDEF datasets.

that such *negative* labels occur on the negative side of the pleasant-unpleasant continuum, e.g. the *valence* axis, as previously shown in Figure 1. Therefore, the negative states grouping (to be within the *negative* state) applied to CK-8 results as well as KDEF results. After that, the resultant percentages obtained by averaging the output of the two trained models. Accordingly, the results depicted in Figure 7, show percentages of each expression obtained from the videos recorded during each task using the average percentages across both the CK-8 and KDEF trained models.

From Figure 7, it can be observed that *neutral* and *negative* expressions occupy the highest percentages across the different tasks, with a greater number of *negative* expressions being shown during the tasks involving *active* interaction by participants. By contrast, in the *passive* interaction Youtube task,

Table 2: Facial expression percentages obtained based on classification of video frames using the average of the two trained models versus self-reported values of *Valence*(Val) and *Arousal*(Aro).

Stimuli	Score	Happy(%)		Surprise(%)		Neutral(%)		Negative(%)	
		Val	Aro	Val	Aro	Val	Aro	Val	Aro
OS	Low	2.64	8.31	28.33	14.54	11.65	39.99	57.38	37.17
	Med	3.72	2.53	20.07	15.65	33.84	32.56	42.38	49.26
	High	3.38	3.76	18.12	31.91	17.27	17.23	61.23	47.1
Online	Low	4.58	2.96	23.83	17.39	13.71	27.19	57.88	52.47
	Med	4.08	4.51	21.91	23.7	25.83	25.93	48.18	45.85
	High	2.53	3.08	32.94	26.85	27.62	20.42	36.91	49.65
Excel	Low	0.81	3.02	27.69	25.75	26.49	34.44	45.01	36.79
	Med	4.72	5.1	19.19	17.55	27.38	28.87	48.71	48.49
	High	2.08	1.03	26.31	28.82	26.82	21.86	44.79	48.29
Pacman	Low	2.75	4.67	18.44	32.45	23.82	31.25	54.99	31.63
	Med	1.59	3.63	28.17	15.22	26.3	30.66	43.94	50.49
	High	5.1	3.83	25.82	28.3	25.28	22.24	43.8	45.63
Youtube	Low	0.5	0.97	16.84	15.05	47.62	54.83	35.03	29.15
	Med	0.65	1.04	13.08	12.99	53.35	50.38	32.92	35.59
	High	1.91	1.12	15.38	17.61	47.2	45.28	35.51	35.99

a greater level of *neutral* expression can be observed. Moreover, although a small percentage of *happy* expression may be observed in all tasks, within the Youtube task, the lowest percentage of *happy* expression is found. Therefore, it may potentially be surmised that, during *active* interaction tasks more variation occurs within the facial expressions of participants, than occurs within *passive* interaction tasks. However, such expressions might not reflect the actual feelings of the participants, as will be exposed later in this paper.

5.3 Expressions versus Reported Valence and Arousal

We also compared the facial expression percentages with the ratings that were self-reported by subjects themselves for both *valence* and *arousal*. Table 2 gives the facial expression percentages obtained from averaging the results of the two trained models, across all the tasks (both *active* and *passive*), using the aforementioned mapping for the self-reported scores of *valence* and *arousal* respectively. In the same manner, combination of these scores as depicted in Figure 5, has been aggregated versus facial expression percentages as given in Table 3. From the results given in the tables, it is apparent that the lowest percentage of frames, across all tasks, show the *happy* expression. However, somewhat surprisingly, tasks where participants self-reported *High valence* values correspond to the facial expression percentages where the *happy* expression is lowest too.

Table 3: Facial expression percentages versus combination of subject self-reported values of *Valence* and *Arousal* together.

Stimuli	Valence/Arousal	Happy(%)	Surprise(%)	Neutral(%)	Negative(%)
OS	Low/Low	0	0	0	0
	Low/High	0.13	41.59	3.25	55.03
	High/Low	0	0	0	0
	High/High	2.43	20.56	27.25	49.76
Online	Low/Low	0	0	0	0
	Low/High	6.84	18.9	10.04	64.22
	High/Low	6.67	30	31.4	31.93
	High/High	1.28	30.36	25.84	42.51
Excel	Low/Low	1.95	15.61	51.46	30.98
	Low/High	0.81	26.74	25.58	46.86
	High/Low	6.65	18.97	42.36	32.02
	High/High	1.59	32.34	21.55	44.52
Pacman	Low/Low	2.45	46.57	12.25	38.73
	Low/High	2.7	13.41	24.3	59.58
	High/Low	6.49	32.15	39.23	22.12
	High/High	4.66	30.15	22.39	42.79
Youtube	Low/Low	0.61	16.5	56.9	26
	Low/High	0.77	18.2	40.48	40.55
	High/Low	1.47	13.68	52.02	32.83
	High/High	1.48	18.65	46.21	33.67

6 Discussion and Conclusion

By merging *Low Pronounced* facial expressions such as *contempt, neutral, fear, angry, sadness* and *disgust* as a single facial expression label, *negative*, makes the analysis of the relationships between the facial expressions and affective states of users performing computer-based tasks much more persuasive. Especially, distinguishing between these facial expressions automatically is a non-robust job (Joho et al., 2009). We have explored a number of different analysis approaches in pursuit of understanding the potential relationships between self-reported affect and the corresponding recorded facial expressions over a number of tasks. Firstly, it was anticipated that there is a difference in the facial expression percentages between *passive* and *active* interaction contexts, due to the significant difference between the nature of the contexts.

Although a general inference is that both *active* and *passive* contexts are similar with regard to the relationships between the self-reported measures and the observed facial expressions, one difference found was that the facial expression alternation that occurs within an *active* context is marginally increased over that found within a *passive* context where an expression mostly remains as it is. In addition, there is a weak correlation, and inconsistency in some cases, between the individual and combined reported values of *valence* and *arousal* with facial expression found in the recordings. Subsequently, this potentially supports the hypothesis that facial expressions do not reflect the actual feelings of users within a HCI context. On the other hand, it may

potentially indicate the fact that when participants attempted to assess and self-report their actual feelings, they were unable to accurately distinguish and report on different emotions and feelings. This point of view certainly agrees with the argument given in (Picard, 2003) that humans often do not know how to articulate their actual feelings and affective states due to ambiguity and mixed mental activities. In addition to the significant differences in *valence* and *arousal* values that were reported for the various tasks, simultaneously the same facial expressions were still portrayed.

Moreover, the results presented reveal that the accuracy of automatically detecting affective states using facial expression analysis, particularly within a HCI context, is not comparable to the accuracy achieved by facial expression analysis when acted and performed deliberately as commonly found in most work within the Affective Computing literature. Consequently, this could be due to the nature of the relationship between humans and computers. Ultimately, humans do not (currently) exchange emotions and feelings with machines in the same manner as they do when interacting with each other.

For these reasons, researchers look to other technologies for the purpose of detecting human emotions and states (Jaimes and Sebe, 2007). Some of these technologies have very good recognition accuracy rates for certain states such as stress, which can be reflected through physiological responses such as heart rate and blood pulse volume, though they could be obtrusive and too noisy to be employed for generic HCI purposes. Additionally, some of the technologies may produce an extra effect on the users and cause additional feelings for humans, rather than detecting actual feelings they are attempting to measure (Hernandez et al., 2014). Therefore, HCI approaches and designs begin, especially within the last decade, to shift the focus from the behaviours and procedures of user experience, towards psychology and sociology concepts, which take into account human factors, emotions, cognitive aspects and individuals behaviours (Harper et al., 2008).

The work presented in this paper comprises deploying an approach for user modelling and affective state detection via facial expression analysis within HCI context. Employing a hierarchical ensemble model, the classification decision is decomposed into smaller micro-decisions that are individually made by specialised binary classifiers, thereby facilitating higher accuracy of the overall model. In addition, a data collection study, along with its corresponding findings from data analysis were discussed. The same analysis approach was also applied to data from another published dataset, namely DEAP dataset. Subsequently, from analysing video frames collected within different applications, it is concluded that facial expressions cannot precisely reveal the actual feelings of users. Experiments presented in this paper examined automatic affect recognition during common computer usage with results suggesting that facial expressions doubtfully indicate the actual feelings of users during interactions with computers. Consequently, work is needed to determine much more appropriate and effective techniques that reason upon users' experiences during interaction with computers in order to facilitate the generation of intelligent and adaptive systems. While the work presented within this paper

investigates facial expression variation across different user interaction tasks, future work could undertake more in-depth experimentation and analysis of related cognitive load.

Although a lot of attention is paid towards deploying facial expression analysis in Affective Computing, the challenges encountered in these endeavours may not only be in terms of technical issues. On the contrary, one potential and significant challenge that may need to be addressed, for example, might be in humans' perception of computers; the perception that the computer is a machine that is a task oriented tool, which is inexpert to reason upon human feelings with the same intellect as that of another human. Thereupon, work is progressing in different directions within various disciplines to reinforce the link between the human and the computer.

References

David W Aha. Generalizing from case studies: A case study. In *Proc. of the 9th International Conference on Machine Learning*, pages 1–10, 1992.

A Asthana, S Zafeiriou, S Cheng, and M Pantic. Incremental face alignment in the wild. 2014.

Thomas Beauvisage. Computer usage in daily life. 2009.

Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. 1994.

Chang and Lin. LIBSVM - A Library for Support Vector Machines, 2001. URL `http://www.csie.ntu.edu.tw/%5C cjlin/libsvm/`.

C Corneanu, M Simon, J. Cohn, and S Guerrero. Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition. 2016.

J. T. Coull. Neural correlates of attention and arousal: Insights from electrophysiology, functional neuroimaging and psychopharmacology, 1998.

Z Duric, W D Gray, R Heishman, A Rosenfeld, M J Schoelles, C Schunn, and H Wechsler. Integrating perceptual and cognitive modeling for adaptive and intelligent human-computer interaction. 2002.

P Ekman, WV Friesen, and JC Hager. Facial action coding system: The manual on cd-rom. instructors guide. 2002.

Yasser EL-Manzalawy. WLSVM, 2005. URL `http://www.cs.iastate.edu/%5C yasser/wlsvm/`.

Nico H Frijda. *The Emotions*. Cambridge University Press, 1986.

Edited Richard Harper, Tom Rodden, Yvonne Rogers, Abigail Sellen, Being Human, et al. Human-computer interaction in the year 2020. 2008.

J Hernandez, P Paredes, A Roseway, and M Czerwinski. Under pressure: sensing stress of computer users. 2014.

Alejandro Jaimes and Nicu Sebe. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*, 108:116–134, 2007.

Hideo Joho, Joemon M. Jose, Roberto Valenti, and Nicu Sebe. Exploiting facial expressions for affective video summarisation. *CIVR '09*, page 1, 2009.

F Karray, M Alemzadeh, J Saleh, and M Arab. Human-computer interaction: Overview on state of the art. 2008.

S Koelstra, C Muhl, M Soleymani, J Lee, A Yazdani, T Ebrahimi, T Pun, A Nijholt, and I Patras. Deap: A database for emotion analysis; using physiological signals. 2012.

Chun Fui Liew and Takehisa Yairi. Facial Expression Recognition and Analysis: A Comparison Study of Feature Descriptors. 2015.

André Teixeira Lopes, Edilson de Aguiar, Alberto F De Souza, and Thiago Oliveira-Santos. Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order. 2017.

Lucey, Cohn, Kanade, Saragih, Ambadar, and Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit. 2010.

Daniel Lundqvist, Anders Flykt, and Arne Ohman. The Karolinska directed emotional faces (KDEF). 1998.

Aleix M. Martinez. Deciphering the face. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2011.

Rosalind W Picard. Affective computing: challenges. *International Journal of Human-Computer Studies*, 59(1):55–64, 2003.

Trevor W. Robbins. Arousal systems and attentional processes. In *Biological Psychology*, volume 45, pages 57–71, 1997.

J Russell and G Lemay. Emotion Concepts. Handbook of Emotion. MH-J. Lewis, M. New York, 2000.

James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

A Salah, N Sebe, T Gevers, et al. Communication and automatic interpretation of affect from facial expressions. 2010.

Anas Samara, Leo Galway, Raymond Bond, and Hui Wang. *Sensing Affective States using Facial Expression Analysis*. 2016.

M Siddiqi, M Ali, M Eldib, A Khan, O Banos, A Khan, S Lee, and H Choo. Evaluating real-life performance of the state-of-the-art in facial expression recognition using a novel youtube-based datasets. 2017.