In Search of Learning Indicators: A Study on Sensor Data and IAPS Emotional Pictures

Haeseon Yun[1], Albrecht Fortenbacher[1], René Helbig[1] and Niels Pinkwart[2]
[1]University of Applied Science, Berlin, Germany
[2]Humboldt University, Berlin, Germany

Keywords: Emotion Detection, Learning Indicators, Sensor Data, Machine Learning.

Abstract: The goal of our research presented in this paper is to relate emotions to sensor data (heart rate and skin conductivity), to interpret them in a learning context (academic emotions) and finally derive learning indicators. For this purpose, we collected sensor data from 27 participants during an emotional picture experiment provided by IAPS (International Affective Picture System). The collected data included EDA signals (electrodermal activity), heart rate and derived data such as skin conductance response, skin conductance level, heart rate variability and instantaneous heart rate labeled by IAPS reference rating and participants' self ratings. The processed data were analyzed using qualitative and quantitative methods as well as machine learning. Furthermore, we applied a human-machine combined approach, namely fuzzy logic reasoning. Our results show that the change of EDA when emotion is induced may serve as a feature to distinguish the intensity of emotion (arousal). Also, classifying EDA signals using a random forest approach shows the best accuracy. In search of learning indicators, we have attempted various tracks of analysis in this study which revealed novel findings, limitations and future steps to consider.

1 INTRODUCTION

Learning involves (among others) mental and emotional cycles. Learners need to memorize, understand, evaluate and even create as part of their learning activities, which involves mental planning, controlling and reflection. In addition to mental work, learners face emotional changes constantly, while facing various situations in learning. Students may feel excited at the beginning of the semester and they can also get frustrated when being involved in some learning tasks that they have trouble with. Learners with a good emotional regulation can drive themselves back from states of frustration to a more positive affective state to overcome challenges. As such, effective learning involves not only domain specific knowledge and strategies, but it also involves regulation of emotion.

In a traditional classroom face-to-face learning environment, the emotional state of a learner can be detected by teachers, instructors and peers, and once it is detected, appropriate feedback can be provided to them (e.g., teachers change their instructional strategy if they detect that most of the class is currently bored). In technology enhanced learning settings however, such a response to student's emotional state is more difficult. Using modern educational technology, learning can happen anywhere and anytime and learners are more than ever free to roam around in different knowledge domains. While most technology enhanced learning environments are not deeply involved in emotional support of learners, wearable technology capable of collecting physiological data of users gain in importance. With wearable sensors, learners' physiological data and surrounding data can be collected and analyzed to provide learners with a personalized environment through context aware feedback and adaptation. As physiological data is related to emotion, it is plausible that student's emotional regulation can be supported by technology enhanced learning designs that make use of wearable sensors.

Motivation for the research presented in this paper came from our research project LISA[1] - Learning Analytics for Sensor-based Adaptive Learning - funded by the German government[2]. In LISA, we developed a wearable sensor device (wrist band) which collects physiological data (skin conductance, heart rate,

[1]http://tel.f4.htw-berlin.de/lisa
[2]16SV7534K

Yun, H., Fortenbacher, A., Helbig, R. and Pinkwart, N.
In Search of Learning Indicators: A Study on Sensor Data and IAPS Emotional Pictures.
DOI: 10.5220/0007734301110121
In *Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019)*, pages 111-121
ISBN: 978-989-758-367-4

skin temperature) and environmental data (TVOC, CO_2). A LISA learning companion, implemented as a tablet app, visualizes and analyzes sensor data received from the wristband. Visualization of sensor data provides awareness about a learner's physiological state and learning environment. Based on goals set by the learner, and based on learning indicators derived from sensor data, the learning companion can give feedback and provide a learning history.

Mirroring back raw physiological signal values such as skin conductance may not be too helpful. An aggregation of this data – translating raw data into information about the learning environment (e.g. a high CO_2 level) or to emotional states – is more appropriate. Many studies on emotion recognition have been conducted (for an overview, see section 2), with different settings, sensors, methodologies, but none of them provides an easy-to-use "recipe" to recognize emotions from EDA[3] and ECG[4] sensors. Yet, a reliable detection of emotions is a requirement for designing guidance mechanisms for students based on the detected emotional states. Therefore, we decided to conduct a study where participants were exposed to IAPS[5] emotional pictures (i.e., pictures that are known to evoke typical emotions), while recording EDA and ECG sensor data. First results from this study will be shown in this paper, and will further be used to calibrate the LISA learning companion.

Section 2 gives an overview on recognition of (academic) emotions, focusing on electro-dermal activity and heart rate. Then, our study design is presented, including the IAPS emotional picture experiment, as well as recording, processing and annotating sensor data during the experiment. Section 4 on data analysis provides statistical features derived from sensor data with respect to the stimuli induced by emotional pictures, and methods like machine learning and fuzzy logic reasoning to classify and predict emotions.

2 STATE OF THE ART

2.1 Emotions in a Learning Context

According to Linnenbrink-Garcia and Pekrun (2014), emotions in education can be defined as multifaceted phenomena which involve affective, cognitive, physiological, motivational and expressive components. Learners' uneasiness feeling before a test is an example of an affective component, being worried is a cognitive component and wanting to avoid the situation is a motivational aspect. The unhappy facial expression is an example of an expressive component, whereas hand sweating or increasing heart rate are physiological components of emotional phenomena. Terms such as mood and affect are widely used in similar areas of research, with mood being emotion with lower intensity and affect as a more comprehensive construct which includes non-cognitive components (Linnenbrink-Garcia and Pekrun, 2014). Emotions in a learning context are also referred to as achievement emotions that learners experience when they succeed or fail on an academic task (Pekrun and Perry, 2014) and as epistemic emotions that are process oriented (Muis et al., 2018). More comprehensively, academic emotions were also theoretically defined as emotions in a learning context mapped onto the two dimensional valence and arousal model in previous studies in autonomic response and emotions in a learning context (Yun et al., 2017).

Regardless of the specific constructs and terms used for emotion in a learning context, the effect of emotions on learners is significant. For example, positive emotions that learners experienced during learning can boost learners to persist in learning and furthermore help learners regulate their learning processes by planning, monitoring, controlling and reflecting on their learning. On the other hand, learners who experience with frequent negative emotions may give up when facing difficult tasks and even drop out completely from the learning course or path. The emotions in a learning context are not only important for students but also for instructors as their roles are not limited to knowledge transfer to students but to encourage and motivate students. Teachers with positive emotion and passion in their roles can exude their positivity to students which in turn can affect learners to have a positive perspective of learning.

2.2 Emotion Detection in a Learning Context

Emotions in a learning context, as stated, are positioned as one of the core components that affect learning progress and experience. Various research attempts to delve into measuring emotions using self-reports, observation and sensor instruments. For instance, self-report instruments have been widely used as they can measure a comprehensive range of emotions with detailed descriptions to distinguish between different construct with economical benefits to administer easily toward massive number of participants (Pekrun and Bühner, 2014). However, these

[3] Electro-dermal activity

[4] Electro-cardiogram

[5] International Affective Picture System

instruments rely heavily on the honest response of learners on their emotional state and limit the complex nature of emotion into a few constructs. Another emotional recognition method is through observation which is mainly used to derive emotional states based on facial expressions. Based on a certain nominal scale, a few observers code the state of the participants who are engaging in a learning task and the coded results are cross-referenced for its reliability (Reisenzein et al., 2014). This method provides rich data with mostly high reliability and validity. However, the observational method takes an enormous amount of time and costs. Furthermore, depending on the cultural background of the observers, the accuracy of classifying emotions changes and detecting subtle emotional changes is difficult with this method (Nelson and Russell, 2013). Another method adopted for emotional recognition is to observe the physiological changes in cardiovascular, electrodermal and respiratory systems using sensors, while emotions are induced. Image/picture, auditory or video stimulus are used to induce specific emotion, in addition to instructing people to relieve the past experience.

2.3 Sensor Data for Emotion Detection

Physiological signals such as heart rate, blood pressure, electrodermal reactions and respirations rate show observable changes due to the effects of emotional stimuli(Bradley and Lang, 2007b). Among various sensor data, electrodermal activity (EDA) and cardiovascular activities derived from ECG were most investigated in relation to emotions (Kreibig, 2010). Various researchers associated skin conductance signals such as SCL [6] and SCR [7] with positive or negative emotion (valence) (Levenson et al., 1990; Cacioppo et al., 2000) and with respect to intensity of emotion (arousal) (Chanel and Mühl, 2015; Lang, 1995). Cardiac signals such as heart rate, heart rate acceleration and heart rate variability also change with respect to the valence of emotion (Mandryk and Atkins, 2007; Vrana et al., 1986; Libby Jr et al., 1973; Ekman et al., 1983). Also combinations of EDA and data derived from ECG signals are used to describe the valence and arousal states of emotions (Winton et al., 1984; Picard et al., 2001a; Gruber et al., 2015; Malkawi and Murad, 2013). Feature extraction is a more practical approach to select appropriate features when using machine learning to classify emotions using heart rate, skin conductance and respiration rate. Aggregated sensor values (e.g. mean absolute normalized first difference of heart rate, first difference of the smoothed skin conductance, three higher frequency bands of the respiration signal) are related to emotions that a person experiences (Picard et al., 2001a).

[6] Skin Conductance Level

[7] Skin Conductance Response

Despite various research efforts and theories between sensor data and emotion, the results are observed to be context-dependent and ambiguous to apply as a ground-truth to recognize learner's emotional state in a context of learning. Therefore, our approach began by focusing on general emotions which can be mapped into an analyzable construct. To do this, we have adapted the emotional picture experiment using IAPS and SAM rating (Bradley and Lang, 2007b) toward students in a higher education institute while recording physiological sensor data (EDA and ECG). The collected sensor data were used to apply qualitative, quantitative and machine learning approaches to find basic relationship between sensor data and emotions, specifically EDA and ECG.

3 EMOTIONAL PICTURE EXPERIMENT

3.1 IAPS Emotional Pictures

IAPS consists of images inducing emotions with normative ratings for valence and arousal levels of emotion, rated by a wide range of people. By using the circumplex model of affect (Russell, 1980), each picture can be mapped onto the grid as shown in Figure 1.

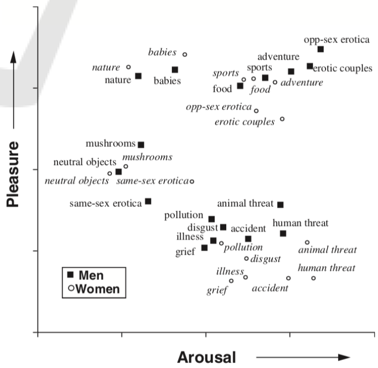

Figure 1: 2-dimensional affective space with IAPS picture (Bradley and Lang, 2007a).

The valence indicates a range from positive to negative and arousal indicates the intensity of the emotion, for example calm/boring to stimulating/exciting. Currently, IAPS includes more than 1000 pictures and these images serve as good standards to evoke specific emotion which is well evaluated by a large number of people from various cultural backgrounds. Most importantly, the images can be plotted on a two dimensional affective space based on its normative valence and arousal value (Lang and Bradley, 2007).

Furthermore, IAPS can be used with SAM rating, a pictorial rating instrument which allows to record viewers' perception in addition to collecting physiological data. Specifically, skin conductance is found to covary with arousal value of both negative and positive valence image and heart rate reacts to positive valence and deceleration is shown when unpleasant images are presented. Comprehensive pictures are found to be relative safe and also effective by providing targeted emotional stimuli yet not detrimental to the viewers.

3.2 Study Design

1182 IAPS pictures labeled by mean and standard deviation of valence, arousal and dominance ratings were provided by the center for the study of emotion and attention[8] . Based on Lang et al. (1997)'s previous results, three main criteria were applied to make a 45-minute experiment. First, to equally distribute the number of pictures for each intended emotion, 24 pictures for each category were selected. Second, to minimize the gender effect on the experiment, pictures that have no significant statistical difference in ratings (valence and arousal) between genders were selected using independent t-tests. For these pictures, the difference in mean valence rating between female and male participants was less than 1. The mean arousal rating difference between female and male was less than 0.8. This resulted in 735 pictures. Third, to select pictures that are targeted to the specific category, pictures containing explicit violence and sexually explicit pictures were excluded and pictures with rating value higher than 6 and rating value less than 4 were used respectively and lastly standard deviation for the respective category was maintained to be as low as possible (less than 2.5).

For high valence and high arousal pictures (HVHA), pictures with value greater 6 were filtered for valence and arousal. 42 pictures of high valence and high arousal pictures were yielded, out of 42 pictures, 7 sexually explicit pictures (IAPS numbers 4656, 4666, 4672, 4680, 4687, 4690, 4695) were excluded. Pictures with high standard deviation were excluded to result in a set of pictures. In sum, pictures selected in HVHA have a valence mean range from 6.07 to 7.74 and the arousal mean range from 6.01 to 7.35. The range of standard deviation for valence remained less than 2 and for arousal between 1.7 to 2.21.

For high valence and low arousal pictures (HVLA), pictures with values greater then 6 were filtered for valence and pictures with arousal value less than and equal to 4 were filtered, which resulted in 42 pictures of high valence and low arousal. Also, pictures with high standard deviation were excluded. For the remaining 24 pictures, mean range of valence is from 6.03 to 7.52 and of arousal is from 2.51 to 3.97. The range of standard deviation for HVLA stimuli pictures is less than 2 for valence and up to 2.2 for arousal.

For low valence and low arousal pictures (LVLA), selecting pictures with less than 4 for valence and arousal resulted in 9 pictures. Therefore, valence and arousal value less than 4.3 were applied which resulted in 29 pictures. The values with highest standard deviation were excluded for both valence (less than 2) and arousal (up to 2.23).

For low valence and high arousal pictures (LVHA), pictures with valence less than 4 and arousal greater than 6 were filtered and resulted in 53 pictures. Out of 53 pictures, 12 pictures with explicit violence (IAPS numbers 2352.2, 3010, 3030, 3059, 3060, 3069, 3071, 3080, 3131, 6021, 6022, and 9252) and 2 redundant pictures (3010, 6570.1) were excluded. Pictures with highest standard deviation for valence (greater than and equal to 2) and arousal (greater than 2.23) were excluded.

To administer the experiment, once the participant arrived at the experiment setting, the subject was guided to sit in front of the computer screen. The general aim and procedure of the experiment were explained and verbal consent was received. The written consent was provided and signed by the subject. Afterwards, the electrodes to measure EDA and heart rate were attached. While EDA and heart rate signals were verified for accurate recording, baseline task and the emotional picture rating task were explained with examples. When the participant had questions, they were answered. Once the participant was ready, a baseline task which involves heartbeat perception tasks was administered for about 5 minutes. Heartbeat perception tasks were performed 3 times with a 25, 35 and 45 second interval respectively. When instructed, the subject was asked to count his or her heart beats just by concentrating on his or taking his or her own pulse or trying any other physical manipulation which might facilitate the detection of heart beats. After the termination of each interval, the subject was requested to report the counted or estimated number of heart beats. After the baseline task, the

[8]https://csea.phhp.ufl.edu

34 minute emotional picture experiment began with the screen message “Get ready to rate the next slide” shown for 5 seconds to initiate the emotional picture rating task. A total of 96 pictures were shown randomly. Each picture was shown for 6 seconds, then a SAM[9] rating, shown in Figure 2, appeared and the subject had 10 seconds to rate valence and arousal. After 10 seconds, an empty screen with a + sign in

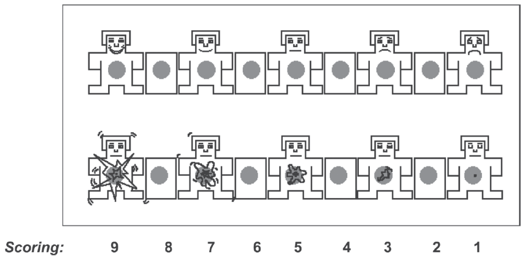

Figure 2: 9-point scale SAM Rating for valence and arousal (Bradley and Lang, 2007a).

the center was presented for 5 seconds followed by the next IAPS picture. After the experiment, participants were directed to a short relaxing video clip to avoid any negative effects of emotional pictures on participants.

4 DATA COLLECTION AND PROCESSING

The data collected through the emotional picture experiment were EDA, ECG, participant’s self-rating of IAPS pictures (valence and arousal), IAPS picture ID and UNIX timestamps. The IAPS reference values for arousal and valence ratings along with standard deviations of each IAPS picture were provided by the center for the study of emotion and attention and related to sensor data using UNIX timestamps.

The EDA and ECG sensor data were collected using a wearable sensors (BITalino (r)evolution Plugged Kit BT [10]). The device was suitable for explorative research with low cost and raw data acquisition functionality. Both EDA and ECG raw sensor data were sampled and stored with 10bit resolution and converted to μSiemens for EDA and mV for ECG using the transfer function stated in the BITalino’s data sheets. Based on the EDA signal in μS, the standardized EDA values were calculated using z-score formula $\frac{x-\bar{x}}{s}$, where $\bar{x}$ is the mean value, s stands for standard deviation. In addition, as EDA signals can be split into a tonic component (SCL, skin conductance level) and a phasic component (SCR, skin conductance response) (Boucsein, 2012), SCR pulses which are peaks within the SCR signal with a rise time that varies from 0.5 to several seconds and non-specific SCR (NSSCR) which are the pulses in the absence of a stimulus were derived. Heart beats were derived from the ECG signal (mV) along with the time interval between consecutive peaks (RR interval) and its reciprocal value IHR [11].

The next step was to aggregate sensor data and derived data to obtain features which could be attributed to emotional states induced by exposing a participant to a picture. These features were then annotated by IAPS picture ID, IAPS rating of valence and arousal and a participant’s self rating. Data were aggregated during a time window immediately following the stimulus, which is the start of the viewing window. Time windows for viewing, rating and relaxing were 6 seconds, 10 seconds and 5 seconds, respectively, so the maximum interval for aggregation was 21 seconds, which is the time between 2 stimuli. For EDA signals, we chose an interval of 6 seconds (viewing window), which complies with findings in literature (Boucsein, 2012). For heart rates, the interval was 21 seconds, to get statistically meaningful values. For all data intervals, a set of values were computed, from mean, minimum, maximum, gradient to statistical values like standard deviation, kurtosis, moments and skewness.

Further features, which may be closely related to mental processes, can be derived from our sensor data. For instance, in case of EDA, the latency of an SCR pulse (if there is a pulse within 6 seconds after stimulus) together with its gradient could be derived. From ECG signals, the most interesting feature, HRV[12], can be derived. HRV has been used as an indicator for the state of a person’s autonomic nervous system (Camm et al., 1996) and it has been identified as a possible indicator for emotions(Gruber et al., 2015). To analyze HRV, either time or frequency domain can be used. With longer data sets, frequency domain analysis can regard the balance between sympathetic and vagal activity (Heathers and Goodwin, 2017). In our data set, due to the relatively short data intervals, we have decided to just include the time-domain values such as RMSSD[13] in our feature set.

[9] Self-Assessment Manikin

[10] https://bitalino.com

[11] Instantaneous Heart Rate.

[12] Heart Rate Variability.

[13] Root Mean Square of the Successive Differences.

5 METHODOLOGY AND RESULTS

Recording sensor data at 1000 Hz during approximately 45 minutes yields a vast amount of data annotated by stimuli induced by showing a picture. As we have IAPS reference ratings along with self-ratings of each picture from each participant, we have investigated the collected sensor data based on previous research relating emotions and sensor data. In our first explorative approach, statistical methods were applied onto all collected data. By using the statistical approach, we have aimed at finding correlations between features derived from sensor data and emotional states, indicated by both IAPS ratings and self-rating of a participant. Subsequently, we have applied machine learning to our data sets. Specifically, through supervised classification, we aimed at finding classifiers for sensor data with respect to valence and arousal. Finally, we adapted our previous fuzzy logic reasoning approach (Moukayed et al., 2018). The common goal of all methodologies is to find a model which predicts emotional states based on EDA and ECG sensor data.

5.1 Qualitative Analysis

In a first step, we used literature findings on emotion and sensor data and attempted to visually validate the three hypotheses: 1) Skin conductance is related to arousal (Ekman et al., 1983; Lanzetta et al., 1976; Levenson et al., 1990; Cacioppo et al., 2000; Picard et al., 2001b). 2) Heart Rate (magnitude and acceleration) is related to valence (Libby Jr et al., 1973; Lang and Bradley, 2007) and 3) Positive emotion has high EDA and high heart rate whereas negative emotion shows high EDA with depressed heart rate (Winton et al., 1984). To validate the hypothesis qualitatively, four representative pictures were chosen. Picture 5621 was chosen for high valence and low arousal and picture 2035 for positive and passive emotion. For low valence, picture 3170 was selected to represent high arousal and low valence furthermore, picture 2039 was selected for low valence passive emotion.

The positive relation between skin conductance and arousal was set as the first hypothesis. Lanzetta et al. (1976) recounted the linkage between SCR and the intensity of emotion and Ekman et al. (1983) recounted that passive emotion depicts a larger change in EDA (in their case, EDA was measured as skin resistance in kΩ). In addition, Picard et al. (2001b) also verified the peak level of skin conductance when a participant's perception of the stimulus is also in the high arousal state.

The second hypothesis is the positive relation between heart rate (magnitude and acceleration) and valence. In general, Libby Jr et al. (1973) reported the association between the change of heart rate and valence of emotion and Ekman et al. (1983) specifically reported that negative emotion resulted in a faster heart rate acceleration compared to the acceleration during happiness. Levenson et al. (1990) extended this logic by indicating that the heart rate shows an accelerative state during negative emotional states. Lang (1995) also recounted that the heart rate shows a modest relationship with the valence self-rating.

The third hypothesis is the relation of positive emotion with high EDA and high heart rate whereas negative emotion with high EDA and depressed heart rate. Winton et al. (1984) reported that the extreme pleasantness is characterized by high value of SCR and high heart rate whereas extreme unpleasantness is characterized by high SCR with depressed heart rate. The results of our qualitative investigation supported the relationship between EDA and arousal (hypothesis 1). Specifically, the accelerated EDA is associated with high intensity and decelerated EDA with low intensity of emotion.

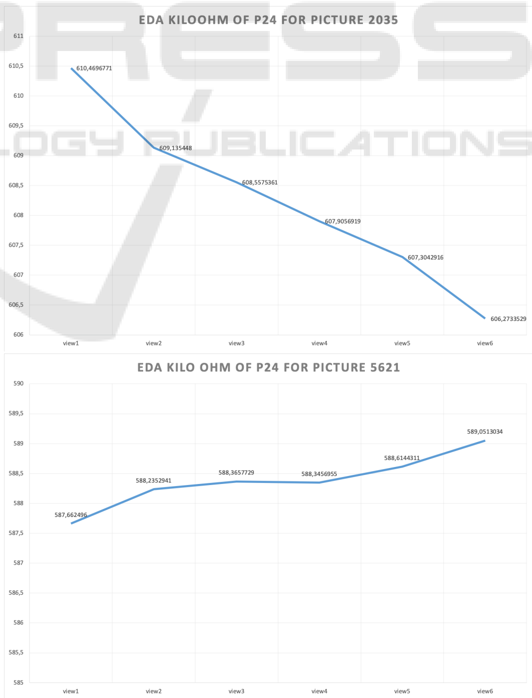

Figure 3: Gradient of EDA (kΩ) of participant 24 for high valence low arousal picture (top), for high valence high arousal picture (bottom).

As shown in Figure 3, the decelerated EDA was observed while the low arousal image was shown and the accelerated EDA was observed when the high arousal image was presented to the participants.

As for the relationship between valence and heart rate (hypothesis 2), our findings showed the opposite results from the previous findings. Levenson et al. (1990) observed the association of accelerated heart rate due to negative emotion, however in our investigation, the heart rate accelerates with positive emotion and decelerates with negative emotion. In addition, the decelerative heart rate for low arousal picture and accelerative trend for high arousal picture were also observed for negative valence, which leads us to further investigate the trend with larger datasets using a statistical method. For the hypothesis 3 indicating the relation of positive emotion with high EDA and high heart rate and the negative emotion with high EDA and depressed heart rate, our visual inspection was not able to find the prominent trend.

5.2 Quantitative Analysis

Based on the qualitative study results, we have further investigated the statistical significance of the findings by applying a quantitative approach on our data set. As we have verified the relation between EDA and arousal (hypothesis 1), in addition to observing the opposite trend in case of the relation between the heart rate and valence (hypothesis 2), we have first taken a step to see the relationship between heart rate acceleration and valence. To do this, we applied F-test and t-test between heart rate acceleration and valence. Afterwards, we have analyzed the association between the accelerative trend of EDA and arousal.

The hypothesis of non-difference of variance and identical averages between high valence and low valence stimulus on heart rate change was set. To confirm the difference in variance, F-test was conducted and the results show that the difference of the variances in heart rate acceleration between high valence situation and low valence situation was not significant. Moving on to find the effect of EDA gradient and arousal level, the hypothesis of non-difference of variance and identical averages between high arousal and low arousal stimulus on EDA change was set. To confirm the difference in variance, F-test was conducted as shown in Table 4.

The difference of the variances in EDA gradient between high arousal situation and low arousal situation is significant with P <0.05, F(2.0629) >Fcrit (1.1334). Furthermore, there was a significant difference in the EDA gradient interval for high arousal stimulus (M=-2.1533×10^{-6}, Std: 5.3111×10^{-5}) and for low arousal stimulus (M=-7.6525×10^{-6}, Std: 3.6978×10^{-5}) conditions with t(1009)=2.89, p = 0.0.0303 (α <0.05) as shown in Table 5.

F-Test Two-Sample for Variances

	HA eda6_gradient_interval	*LA eda6_gradient_interval*
Mean	-2.15334E-06	-7.65257E-06
Variance	2.82084E-09	1.3674E-09
Observations	598	799
df	597	798
F	2.062922451	
P(F<=f) one-tail	8.25096E-22	
F Critical one-tail	1.133490608	

Figure 4: F-test results between high arousal EDA gradient and low arousal EDA gradient.

t-Test: Two-Sample Assuming Unequal Variances

	HA eda6_gradient_interval	*LA eda6_gradient_interval*
Mean	-2.15334E-06	-7.65257E-06
Variance	2.82084E-09	1.3674E-09
Observations	598	799
Hypothesized Mean Difference	0	
df	1009	
t Stat	2.168932719	
P(T<=t) one-tail	0.015160388	
t Critical one-tail	1.6463652	
P(T<=t) two-tail	0.030320776	
t Critical two-tail	1.962317871	

Figure 5: T-test results between high arousal EDA gradient and low arousal EDA gradient.

Even though the association of the heart rate change with valence was not conclusive, our quantitative approach confirmed that the EDA gradient may serve as an indicator to distinguish the arousal level of emotions. As a next step to find more possible indicators for academic emotions based on sensor data, we applied a machine learning approach for classification of sensor data.

5.3 Machine Learning

Some promising approaches to detect emotions from biometric signals using machine learning can be found in literature. For instance, Conati et al. (2018) applied a machine learning approach using EDA and EMG. Ayata et al. (2017) focused on EDA data and made some deeper research on the feature extraction and selection on specific EDA signals and Ferdinando et al. (2018) used ECG based features such as HRV. Specifically, Ayata et al. (2017) used the circumplex model of affect and the valence/arousal classification (low/high arousal, low/high valence) similar to our emotional picture experiment and achieved an accuracy rate of 81.81% and 89.29% for arousal and valence respectively by using only EDA sensor data.

In our search of learning indicators, we used Weka (Witten et al. (2016)) as machine learning environment for the algorithms proposed by Ayata et al. (2017) and Weka's automatic mode Auto-Weka to check other algorithms for their performance on our dataset. As the very first step, we extracted the EDA features suggested by Ayata et al. (2017) from our dataset and evaluated them with several algorithms such as SVM, J48 and random forest. To check other algorithms, we used Auto-Weka, which also attempts to find good fitting hyper-parameters for the algorithms. The result shows that random forest is the algorithm that provides the best accuracy.

On our EDA data, we were able to achieve results with an accuracy of 90% and above with random forest as machine learning algorithm depending on the hyper-parameters. We were able to achieve this although we did not use the features from the Empirical Mode Decomposition as suggested by Ayata et al. (2017). The accuracy values for J48 (decision tree) and SVM were comparable.

As our aim is to construct a general model which is independent of participants (individual differences), we evaluated the models with 10-fold cross validation and k-1 validation (k : number of participants). As soon as the training data did not include the data of one participant and the test data included the missing participant, the accuracy diminished to around 50% for valence and arousal respectively. This could be caused by the size of our dataset being too small, which might have led to a overfitting model. To verify the findings, we have analyzed our feature files (described in Section 4) and the results were similar. To attain a more generally usable model, we reduced the feature selection to the z-score, normalized features, the features calculated from the z-scored data and more baseline independent features, yet no significantly different results were found. Further research is needed on a generally usable model, which does not depend on previous training for each person. Nevertheless, it appears to be possible to implement emotion detection in a learning environment, if a participant-dependent training of the machine learning model is completed prior to the start of learning. In this case, we suggest a random forest algorithm and a training time of approximately 90 percent of the time of our emotional picture experiment, as this fits to our outcome with 10-fold cross validation.

Based on our qualitative approach, we have noticed interesting trends and associations of physiological data and the dimension of emotion. However, the quantitative and the machine learning approach provided less convincing results. As both research methods did not lead to a general model, we analyzed our data using a mixed-method which is fuzzy logic reasoning.

5.4 Fuzzy Logic

In a human-like manner, a fuzzy logic model uses expert knowledge as sets of simple rules to associate sensor values with the dimension of emotion. In many cases, these rules are insufficient to apply a quantitative approach as they are stated as a relational statement without numerical value. For example, the association between stress and the combination of EDA peak height and the instantaneous heart rate (Setz et al., 2010) can not be directly applied in a quantitative analysis. However, as fuzzy logic allows ambiguity, the association without concrete values can be analyzed. As we have investigated emotion using IAPS pictures, various literature findings could have served as experts' rules. For instance, as emotional valence of 6 or above can be described as happiness or contentment (Lang, 1995), it can be transferred into a simple rule in a fuzzy logic.

To apply these rules in fuzzy logic, the fuzzy membership function was defined based on the scatterplot or histogram of the collected data. The boundaries of each range (e.g. low, mid, high) were set by using the mean and standard deviations of the sample. Then the membership functions transformed the membership of a specific element into a percentage membership in the set of values. The fuzzy logic system weighs each input signal, defines the overlap between the levels of input, and determines an output response. Domain knowledge is modelled as a set of IF/THEN rules which use the input membership values as weighting factors to determine their influence on the fuzzy solution sets. Once the functions are inferred, scaled, and combined, they are de-fuzzified or translated into a solution variable, which is a scalar output (Cox, 1992).

Based on the fuzzy logic approach presented in Mandryk and Atkins (2007), we have started out with a fuzzy logic model for assessing arousal with EDA gradient. We have first customized the membership function according to our data set. The mean of the EDA gradient was 0.004925482 Ω with standard deviation 0.030947348, therefore three boundaries were set based on the shape of our data set as follows:

```
TERM low := (−0.29527918, 1)
    (0.004925482, 0);
TERM mid := trian 0.004925482
    0.026021866  0.03587283;
TERM high := (0.004925482, 0)
    (0.03587283, 1);
```

As our qualitative and quantitative approaches confirmed the relation between EDA changes (gradient) with the arousal level, the following three rule sets were used to de-fuzzify/ translate the level of EDA gradient to the arousal level.

```
RULE 1 : IF
    eda6_gradient_interval_ohm IS
    low THEN arousal IS low ;
RULE 2 : IF
    eda6_gradient_interval_ohm IS
    high THEN arousal IS high ;
RULE 3 : IF
    eda6_gradient_interval_ohm IS
    mid THEN arousal IS mid ;
```

As we did not have a specific boundaries to specify each level of arousal, as the general rule, 0.50 was used as the mid-point for de-fuzzification.

```
TERM low := (0, 1) (50, 0);
TERM mid := (25, 0) (50, 1) (75, 0)
    ;
 TERM high := (50, 0) (100, 1);
```

Our experiment was designed to distinguish two levels of valence and arousal (high or low) whereas the outputs from the fuzzy logic produced three arousal levels (low, mid and high). Upon the results from the fuzzy logic, we have omitted any values with mid arousal classification and compared against IAPS reference classification and also with participants' self-rating classification. As the SAM scale that the participants used to indicate their perception of the image has a range from low, mid and high, more neutral responses (value 4, 5 and 6) were filtered out and the values above 6 were chosen as high and ones with less than 4 were classified as low. Out of 2016 data, classification with mid was filtered out and only 1465 data were available for low and high arousal results. Comparing against the IAPS reference classification, 953 values were in match with de-fuzzified classification, which indicates that the stimulis that are normally classified by a large number of people either as high or low can also be classified accordingly by fuzzy logic approach. Specifically, when using a simple rule set relating EDA gradient and arousal level, the output shows that 65 % of data are in sync with IAPS reference classification. Comparing against the participant's self-rating, 708 values were in match with de-fuzzified classification. This indicates that 48 % stimuli that were perceived high or low by the participants can also be classified accordingly using a fuzzy logic approach. Without changing the de-fuzzification rule, for the analysis purpose, we have applied the stricter boundary by setting values higher than 65 as high and lower than 40 or 35 as low on de-fuzzified results to compare with the IAPS reference value and the participant's self-rating value.

The percentage of matching between IAPS reference and de-fuzzified results remained as 65 % whereas it has increased to 81 % between the participant's self-rating value and de-fuzzified classification. Our current fuzzy logic approach used the general rule (0.50 as the mide-point) for de-fuzzification. However, using a specific rule for the de-fuzzification may result in a better match among all three measures. Our next step will involve applying the resulted fuzzy logic in a new data set, in addition to conducting an optimization similar to our previous study (Moukayed et al., 2018) in search of finding novel learning indicators using sensor data.

6 CONCLUSION

The overall goal in our research project LISA is to investigate learning environments that are able to adapt to the emotional state of a learner as detected by physiological data. To design the system components that derive emotional states based on the sensor data, the goal of the study presented in this paper was to generate sensor data which can be related to well-known stimuli generated by exposing participants to IAPS pictures. After an extensive literature research, we attempted to verify findings from literature using our data qualitatively. Before doing data analysis to understand the autonomic nervous system, more data processing is required (including filtering and derived values). Relevant related research exists, but is very heterogeneous and done differently (both hardware and software) and thus difficult to apply straightforwardly for our purposes. In addition, post data processing is also an important aspect to delve into, as downsampling and averaging over a longer time interval bears the risk of losing information about subtle changes in emotional processes.

Based on intuitive features found in our previous work, we have applied a quantitative approach to look deeper inside our data. Through an exploratory approach, we gained deeper insights into the correlations between stimuli induced by high or low arousal pictures and increase or decrease of EDA values. Our investigation of our data is limited and explorative as various assumptions should be verified and more advanced statistical methods such as MANOVA should be considered for the forthcoming steps. Regarding the machine learning approach, using a random forest classifier, we were able to find an acceptable (good) classification of EDA values with respect to two arousal classes. The findings support the feasibility of using machine learning approaches for emotion detection – yet, we are unable to tell whether the training time of the machine learning model can be shortened to a timespan that is realistic for typical learning situations. Lastly, human and machine combined approaches (fuzzy logic reasoning) are a very promising tool for emotion prediction. However as this approach depends on rather vague "domain knowledge" with relationships between physiological data and emotional state (e.g IF eda IS high THEN arousal IS high), it may not be sufficient to obtain precise arousal or valence values. Better rules which include features like skewness of EDA values, SCR latency or HRV, might be derived from insights obtained by quantitative analysis and/or machine learning – however, for some learning setting it may be perfectly fine to only identify rough tendencies in emotional states instead of fine-grained insights. The required granularity of the detection certainly depends on the intended use within the educational software.

The results of the work presented in this paper directly feed into several industrial learning applications of LISA project partners that design learning analytics dashboards incorporating emotional information[14] or make use of the emotional state in order to adapt difficulty levels of educational games[15]. Our next steps include the use of data collected in a similar study conducted by one of our LISA project partners[16], who combined an emotional picture experiment with a cognitive task, using the identical device to record physiological data (EDA, Heart Rate and Skin temperature). In parallel to getting more insight into emotion detection, we will also try to analyze cognitive states using features from sensor data. Combining indicators for emotional and cognitive states could be a big step towards our goal of finding learning indicators.

REFERENCES

Ayata, D. D., Yaslan, Y., and Kamaşak, M. (2017). Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods. *Istanbul University-Journal of Electrical & Electronics Engineering*, 17(1):3147–3156.

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Bradley, M. M. and Lang, P. J. (2007a). The international affective picture system (iaps) in the study of emotion and attention. In Coan, J. A. and Allen, J. J. B., editors, *Handbook of Emotion Elicitation and Assessment*, chapter 29, pages 29–46. Oxford University Press, New York.

Bradley, M. M. and Lang, P. J. (2007b). Motivation and emotion. In Cacioppo, J., Tssinary, L. G., and Berntson, G. G., editors, *Handbook of psychophysiology*, chapter 25, pages 581–607. Oxford University Press, New York.

[14]Neocosmo, Saarbrücken.

[15]Serious Games Solutions, Potsdam.

[16]Leibniz Institut für Wissensmedien, Tuebingen.

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., Ito, T. A., et al. (2000). The psychophysiology of emotion. *Handbook of emotions*, 2:173–191.

Camm, A., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., Coumel, P., Fallen, E., Kennedy, H., Kleiger, R., et al. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology. *Circulation*, 93(5):1043–1065.

Chanel, G. and Mühl, C. (2015). Connecting brains and bodies: applying physiological computing to support social interaction. *Interacting with Computers*, 27(5):534–550.

Conati, C., Chabbal, R., and Maclaren, H. (2018). A Study on Using Biometric Sensors for Monitoring User Emotions in Educational Games. Technical report.

Cox, E. (1992). Fuzzy fundamentals. *IEEE spectrum*, 29(10):58–61.

Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, 221(4616):1208–1210.

Ferdinando, H., Seppänen, T., and Alasaarela, E. (2018). Emotion recognition using neighborhood components analysis and ECG/HRV-based features. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Gruber, J., Mennin, D. S., Fields, A., Purcell, A., and Murray, G. (2015). Heart rate variability as a potential indicator of positive valence system disturbance: a proof of concept investigation. *International Journal of Psychophysiology*, 98(2):240–248.

Heathers, J. and Goodwin, M. (2017). Dead science in live psychology: A case study from heart rate variability (hrv).

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, 84(3):394–421.

Lang, P. and Bradley, M. M. (2007). The international affective picture system (iaps) in the study of emotion and attention. *Handbook of emotion elicitation and assessment*, 29.

Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American psychologist*, 50(5):372.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International affective picture system (iaps): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, pages 39–58.

Lanzetta, J. T., Cartwright-Smith, J., and Eleck, R. E. (1976). Effects of nonverbal dissimulation on emotional experience and autonomic arousal. *Journal of Personality and Social Psychology*, 33(3):354.

Levenson, R. W., Ekman, P., and Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology*, 27(4):363–384.

Libby Jr, W. L., Lacey, B. C., and Lacey, J. I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology*, 10(3):270–294.

Linnenbrink-Garcia, L. and Pekrun, R. (2014). *International handbook of emotions in education*. Routledge.

Malkawi, M. and Murad, O. (2013). Artificial neuro fuzzy logic system for detecting human emotions. *Human-Centric Computing and Information Sciences*, 3(1):3.

Mandryk, R. L. and Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International journal of human-computer studies*, 65(4):329–347.

Moukayed, F., Yun, H., Fortenbacher, A., and Bisson, T. (2018). Detecting academic emotions from learners? skin conductance and heart rate: Data-driven aproach using fuzzy logic. In Schiffner, D., editor, *Proceedings der Pre-Conference-Workshops der 16. E-Learning Fachtagung Informatik co-located with 16th e-Learning Conference of the German Computer Society (DeLFI 2018)*, volume 2250. CEUR Workshop Proceedings (CEUR-WS.org).

Muis, K. R., Chevrier, M., and Singh, C. A. (2018). The role of epistemic emotions in personal epistemology and self-regulated learning. *Educational Psychologist*, pages 1–20.

Nelson, N. L. and Russell, J. A. (2013). Universality revisited. *Emotion Review*, 5(1):8–15.

Pekrun, R. and Bühner, M. (2014). *Self-report measures of academic emotions*. Routledge Handbooks Online.

Pekrun, R. and Perry, R. P. (2014). Control-value theory of achievement emotions. *International handbook of emotions in education*, pages 120–141.

Picard, R. W., Vyzas, E., and Healey, J. (2001a). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23(10):1175–1191.

Picard, R. W., Vyzas, E., and Healey, J. (2001b). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE transactions on pattern analysis and machine intelligence*, 23(10):1175–1191.

Reisenzein, R., Junge, M., Studtmann, M., and Huber, O. (2014). Observational approaches to the measurement of emotions. *International handbook of emotions in education*, pages 580–606.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., and Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417.

Vrana, S. R., Cuthbert, B. N., and Lang, P. J. (1986). Fear imagery and text processing. *Psychophysiology*, 23(3):247–253.

Winton, W. M., Putnam, L. E., and Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology*, 20(3):195–216.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 4th edition.

Yun, H., Fortenbacher, A., Pinkwart, N., Bisson, T., and Moukayed, F. (2017). A pilot study of emotion detection using sensors in a learning context: Towards an affective learning companion. In Ullrich, C. and Wessner, M., editors, *Proceedings der Pre-Conference-Workshops der 15. E-Learning Fachtagung Informatik co-located with 16th e-Learning Conference of the German Computer Society (DeLFI 2017)*, volume 2092. CEUR Workshop Proceedings (CEUR-WS.org).