Chapter 3

Machine learning-based affect detection within the context of human-horse interaction

Turke Althobaiti[1], Stamos Katsigiannis[2], Daune West[2], Hassan Rabah[3] and Naeem Ramzan[2]

This chapter focuses on the use of machine learning techniques within the field of affective computing, and more specifically for the task of emotion recognition within the context of human-horse interaction. Affective computing focuses on the detection and interpretation of human emotion, an application that could significantly benefit quantitative studies in the field of animal assisted therapy. The chapter offers a thorough description, an experimental design, and experimental results on the use of physiological signals, such as electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG) signals, for the creation and evaluation of machine learning models for the prediction of the emotional state of an individual during interaction with horses.

3.1 Introduction

Affective computing is a field within computing that focuses on emotion recognition and interpretation, entailing a variety of methods that allow the use of affect information according to user-specific needs [1]. Among the available approaches for emotion recognition, the analysis of physiological signals has gained prominence in human-computer interaction applications. Various physiological signals originate from the central nervous system (CNS), as well as from the peripheral nervous system (PNS), and studies have shown that they contain information associated with the affective state of an individual [2]. Such information can be further exploited for the task of emotion recognition, with various studies having linked physiological signals to the Valence and Arousal dimensions of emotion [2, 3, 4]. Apart from human-computer interaction, the detection of the affective state of an individual could potentially benefit various other fields, such as the field of animal-assisted intervention (AAI). The use of AAIs as complement to conventional mental health treatment has gained a lot of attention in recent years [5, 6], although the lack of sufficient research

[1]Nothern Border University, Saudi Arabia
[2]University of the West of Scotland, UK
[3]University of Lorraine, France

on the effectiveness of such techniques has hindered their acceptance. However, more and more sophisticated studies are being conducted [7].

A widely documented case of AAI is the use of horses for mental health treatment. Equine assisted therapy (EAT) has many forms, including therapeutic riding, hippotherapy, equine-facilitated therapy, and equine-assisted learning therapy [8]. Evidence of the recognition of horses as agents of healing can be found in early mythology, with physicians suggesting that horse riding would "raise spirits" in people suffering from conditions that could not be treated [9]. Similar examples can be found throughout history. For example, reports from the 18th century indicate that the Pope's physician suggested horse riding as a remedy to the Pope's health problems [10], while Mayberry [11] reported a recommendation by a Scottish physician from 1870 that the riding of a spirited horse "stimulated life forces" and should thus be recognised as a treatment for people suffering from depression. The assertion that horse riding may have therapeutic benefits for people with disabilities has been gaining momentum since the mid-twentieth century [12].

It should be noted that riding horses for therapy differs from recreational or sports horse riding as it incorporates activities which are equine-orientated and designed to promote some of the following positive outcomes: physical, emotional, behavioural, social, cognitive, and educational objectives. The methods used by organisations who provide EAT and therapeutic treatment involving horses varies considerably. The belief that horses have therapeutic benefits for humans derives from historical observations of the emotional and physical health gains that relationships between the two entities has brought to human lives [8]. According to Kendall et al. [8], the sensitivity of horses to non-verbal communication of humans and other animals in their environment and their natural lack of socio-cultural standards and restrictions that communication between humans abides to, offers an environment where a sense of safety, security and trust can be fostered in people, especially those who suffer from disabilities. Based on these observations, EAT is being increasingly introduced in the treatment of both children and adults who suffer from PTSD [13, 14], depression [15], symptoms related to combat trauma [16], etc.

Scientific studies on EAT have been carried out by several researchers, for example, [17], [18], and [19]. However, few of these research studies have concentrated on gaining deeper knowledge of the complex emotions that horses seem to elicit in their riders and handlers. Although evidence of this emotional response in humans during human-horse interaction has been recorded in historic documents on horsemanship [20, 21, 22], these sources offer conclusions drawn from evidence based on subjective experiences and empirical observations. The capabilities of modern computing technology and health sensors have been fundamental in facilitating the measurement and quantitative examination of emotional reactions triggered in humans during activities or interactions. Drawing on analytical methods to assess human-horse interactions quantitatively could potentially be an effective way to gain insightful data which could demonstrate the benefits of EAT.

The aim of this chapter is to examine physiological signal-based affect recognition techniques as a potential solution for assessing the emotional response of humans to interactions with horses. To this end, following a predefined procedure, electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG) signals were captured while humans were interacting with horses. During a variety of activities involving horses, participant emotional states were recorded using a method of self-reporting which required them to identify and select, from a list, the relevant emotions based on their experience during each activity. The emotions reported by the participants were then mapped to their respective Valence and Arousal values according to Russels's Circumplex Model of Affect [23]. The acquired physiological signals and the associated emotion-related labels were used in order to train machine learning models for the task of predicting the respective Valence and Arousal values, thus predicting the emotional state of the respective participant.

The rest of this chapter is divided into six sections. Section 3.2 gives a brief overview of several recent studies within the field of affect recognition and human-horse interaction. Section 3.3 presents a comprehensive description of the experimental approach followed in this study, while Section 3.4 provides a description of the acquired data analysis process and the machine learning approach followed. Experimental results are provided in Section 3.5 and finally the results are discussed in Section 3.6 and conclusions are drawn in Section 3.7.

3.2 Background

Emotion recognition is one of the main applications in affective computing and has attracted of lot of attention from the research community. The use of physiological signal analysis for emotion recognition has shown great potential and thus has been and still is extensively studied [24, 25]. The use of features extracted from physiological signals with the aim of training machine learning models for the prediction of the Valence and Arousal dimensions of an emotion is one of the most promising solutions. The performance of Support Vector Machines (SVM) using features extracted from peripheral physiological signals and eye gaze data was studied by Soleymani et al. [26] for affect recognition when film clips are used for affect elicitation. Peripheral physiological signals along the Naive Bayes classifier were also examined by Koelstra et al. [2] using music video clips as stimulus, while the performance of connectivity-based and channel-based EEG features was examined on the same data by Arnau-González et al. [27]. The performance of features extracted from various physiological signals was studied by Abadi et al. [28], while the use of EEG and ECG-based features, and EEG, ECG, and Galvanic Skin Response (GSR)-based features when film clips are used as stimulus and low-cost portable devices used for signal acquisition were examined by Katsigiannis and Ramzan [3] and Correa et al. [29] respectively.

While the use of physiological signals for affect recognition has been studied extensively, very limited research has been conducted for the use of physiological signals for assessing the emotional responses associated with human-horse interaction. Hama et al. [30] concluded in an early study that grooming horses leads to a reduction in tension both for the human subjects and the horses. Many years after that study, Chen et al. [31] examined the relationship between autism spectrum disorder (ASD) and resting frontal EEG brain activity in young children when interacting with a horse. Results showed that children with ASD exhibited higher left frontal dominance during the baseline condition, but right frontal dominance while grooming the horse, indicating that this change was a result of the interaction with the horse. The reliability of wearable physiological signal sensors for monitoring physiological signals in horses for the purpose of assessing human-horse interaction was examined by Guidi et al. [32], concluding that it was viable to quantitatively assess human-horse interaction in such a manner. In another study, Lanata et al. [33] examined the use of ECG signals for assessing human-horse interaction before the interaction, during visual-olfactory interaction and while grooming the horse, with the purpose of using machine learning to detect the interaction activity, achieving an accuracy of 70.87% using the Nearest Mean classifier and 90.95% using an SVM classifier on the same data during a later study [34]. Machine learning approaches and ECG-based features were also employed by Althobaiti et al. [35] for distinguishing between negative and positive emotions during human-horse interaction, reporting an accuracy of 74.21%. In another study, Althobaiti et al. [36] examined the use of ECG, EMG, and EEG-based features in combination with various classification algorithms for the detection of the Valence and Arousal dimensions of the emotional state of humans during human-horse interaction.

Despite the aforementioned approaches for examining human-horse interaction via physiological signals and machine learning, it is evident that more extensive studies are required in order to establish robust methods for the evaluation of human-horse interaction.

3.3 Experimental protocol

In this work, the emotional responses elicited though human-horse interaction were studied through an experiment designed to accommodate such interaction while physiological signals were recorded. Participants were selected and participated in activities involving interaction with horses, while their emotional responses were documented via self-reporting.

3.3.1 Field experiment setting

A small livery yard situated in the county of Ayrshire, Scotland, UK, was selected as the location for data acquisition, with the experiments taking place from late May until early August 2018. Prior to the experiment taking place, a consent form was signed by participants who were also briefed on the experimental procedure and were given instructions about the experiment. During the briefing, they were given an opportunity to ask questions about the processes involved, to ensure that they were not in any doubt of what would be required of them. Instructions on the handling of the horses and personal safety were also provided by the horse handler. After the physiological sensors were attached to the participants by the supervising researcher and the quality of signal acquisition was verified, the experiment commenced.

Two healthy horses, Max, who was 20 years old, and Braga, who was 8 years old, were selected by the handler based on their history of having a calm and friendly temperament in the company of humans who were both familiar and unfamiliar to them. The experimental procedure was repeated two times by each participant, one time interacting with Max and one time interacting with Braga, and was divided in three activities conducted in consecutive order within a small indoor sanded arena:

1. *Observing*. Observing was the first activity which lasted for a period of 4 min. During this activity, the participants were required to sit on a chair within the sanded arena while the horse freely moved around. The aim of this activity was to provide conditions where a sense of familiarity could be created between the participant and the horse, as well as to give the horse an opportunity to become acquainted with the surroundings which also included the research team and their equipment [32].
2. *Grooming*. Grooming was the second activity, lasting for a period of 2 min, which required the participants to groom the horse with a brush. Previous research has reported a reduction in heart rate in both humans and horses during similar activities when they are both comfortable [30]. It should, however, be noted that the horse was tied to a pole throughout the duration of this activity.
3. *Leading*. Leading was the third and final activity of the experiment, during which the participants were required to lead the horse around a predetermined route within the sanded arena. The time frame of this particular activity was determined by the experience of the individual participant and their ability to control the horse, although a maximum duration of 4 min was set.

On completion of the three activities, the participants were asked to complete a questionnaire regarding their emotional state during each activity.

A ten minute interval between the experiment with each horse was factored in to the experiment and provided the handler with sufficient time to escort the next horse into the arena. This also allowed sufficient time for physiological sensors to be placed on participants, if required.

3.3.2 Experimental data acquisition

Electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG) signals were captured throughout the whole duration of the experiment using wireless, portable low-cost and low-weight sensors, while a laptop computer was utilised for signal recording and visualisation. The portable sensors were selected for their convenience during outdoor use as other types that require cables may restrict user movement during interaction with the horses. Furthermore, the small form-factor of the sensors and being wearable made them invisible to the participants. This fact, along with setting the laptop used for signal recording to not emit any sounds or show movement in the screen, ensured that there was minimal external interference due to the presence of equipment, thus avoiding bias.

ECG signals were captured at a 256 Hz sampling rate using a SHIMMER v2 [37] sensor that utilised four standard electrodes positioned on both lower ribs and clavicle. EMG signals were also captured at a 256 Hz sampling rate using a SHIMMER v2 sensor that utilised three standard electrodes positioned on the upper trapezius muscles. A 14-channel EEG signal was captured at a 256 Hz sampling rate using the Emotive EPOC+ wireless headset [38] that utilises 16 gold plated contact sensors which are fixed to flexible plastic arms placed against the head of the user in areas aligned with the AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, M1 and M2 locations of the international 10-20 system [39], with the M1 and M2 electrodes used as reference. Furthermore, all the recorded ECG, EMG, and EEG samples were accompanied by timestamps with millisecond precision

In addition to the captured physiological signals, a video recording of the experimental procedure was captured for all the participants in order to provide reference material for validation and for allowing the researchers to accurately trace and record the timestamps associated with each activity.

3.3.3 Self-reporting of emotional state

On completion of the three activities with each horse, participants were given a questionnaire to provide an assessment of their emotional state during the activities they conducted. In that questionnaire, participants were asked to report the emotions they felt during each activity out of a list of 28 emotions, shown in Table 3.1. These 28 emotions were selected due to their available mapping in the Valence and Arousal dimensions, as proposed by Russel in the *Circumplex Model of Affect* [23] and is shown in Figure 3.1. According to Russel's Valence/Arousal model [40], the Valence and Arousal dimensions correspond to the main aspects of human emotion, with Valence being a measurement of the positiveness of an emotion and Arousal a measurement of the excitement associated with an emotion. This model allows for each perceived emotional state to be depicted on a 2-dimensional plane, with Valence and Arousal at each axis respectively, as shown in Figure 3.1. It must be noted that participants were only instructed to select the emotions that they felt and did not have to use the

Table 3.1 The emotions listed in the self-reporting questionnaire arranged in terms of Valence and Arousal.

Low Arousal Negative Valence	**Low Arousal Positive Valence**	**High Arousal Negative Valence**	**High Arousal Positive Valence**
Miserable	Content	Alarmed	Astonished
Sad	Satisfied	Afraid	Excited
Depressed	At ease	Angry	Aroused
Gloomy	Serene	Tense	Happy
Bored	Calm	Frustrated	Delighted
Droopy	Relaxed	Annoyed	Glad
	Sleepy	Distressed	Pleased
	Tired		

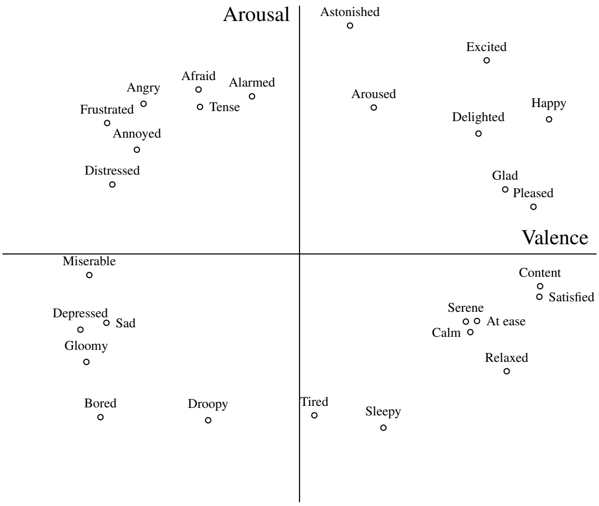

Figure 3.1 Positioning of emotions in the Valence/Arousal space according to Russel's Circumplex Model of Affect [23]

Valence/Arousal scale, thus ensuring that any bias due to a misunderstanding of the rating scale was avoided.

3.3.4 Participants

Out of the 23 participants recruited for the field experiments, only the data from 19 were considered for further analysis, as the captured data for 4 participants were unusable due to erroneous data recording. These 19 participants consisted of 12 males and 7 females that reported having normal health and were between 19 and 64 years old, with the average age being 38.05 years with a standard deviation of 13.14 years. Furthermore, 8 participants reported having no prior experience with horses, 6 participants reported having prior experience with the horses used in this study, and 5 participants reported prior experience with other horses.

3.4 Analysis of captured data

3.4.1 Pre-processing of physiological signals

The video recordings of the experiments were used in order to extract the timestamps referring to each activity and participant. Using these timestamps, the captured ECG, EMG, and EEG signals for each participant were segmented into six segments, each referring to a particular activity and horse. Then, the following denoising procedures were applied to the acquired segments in order to reduce the effects of noise and artefacts that originate from the participants' movement during signal capturing:

- **ECG:** Baseline wander is a common artefact in ECG signals, originating from body movement, respiration, and high frequency noise [41]. The effects of baseline wander in the captured ECG signals were reduced as proposed in [42], by first applying a median filter with a 200 ms window, then a median filter with a 600 ms window, followed by the subtraction of the filtered signal form the original signal. Baseline wander reduction was then followed by a bandpass filter (0.7-20 Hz) in order to further reduce noise.
- **EMG:** The Augsburg Biosignal Toolbox (AuBT) [43] was used for the pre-processing of EMG signals. The process followed consisted of cutting the peaks in the EMG signal with values within the 3% of the lowest or highest signal values, applying a 3rd order Butterworth FIR lowpass filter with a 0.4 Hz cutoff frequency, and finally the normalisation of the filtered signal to the range $[0,1]$.
- **EEG:** A a Butterworth bandpass filter (0.4-65 Hz) was first applied to the EEG signal, followed by the PREP EEG data pre-processing pipeline [44] applied using the EEGLAB toolbox [45].

3.4.2 Extraction of features from physiological signals

Each segment of the pre-processed physiological signals was used in order to extract features that would later be used for the training of machine learning models. It must be noted that only the last 30 s of each segment were used for feature extraction in order to remove any bias stemming from the varying duration of each activity. Inconsistent duration of activities is a common issue in affective computing studies that is usually addressed by either analysing signals using a moving window of fixed duration (e.g. [46]) or by only considering a fixed-length window from the beginning or ending of the signal (e.g. [3, 2, 28]). The later option was used in this study since the moving window approach is more suitable for real-time applications. The following features were extracted from the captured physiological signals:

- **ECG:** 84 ECG-based features, commonly used in affect recognition studies [2, 3, 35, 36], were extracted using the Augsburg Biosignal Toolbox (AuBT) [43]. The extracted features consisted of the maxima, minima, mean, median, standard deviation and range from the raw signal and the derivative of PQ, QS and ST complexes of the ECG signal, the maxima, minima, mean, median, standard deviation and range from the Heart Rate Variability (HRV) histogram, the number of intervals with latency $> 50\,ms$ from HRV, and the Power Spectral Density (PSD) from HRV between the intervals $[0\,,0.2]$, $[0.2\,,0.4]$, $[0.4\,,0.6]$ and $[0.6\,,0.8]$. The computed features were then concatenated in order to create the final feature vector.
- **EMG:** 21 EMG-based features, commonly used in affect recognition studies [35, 36], were extracted using the Augsburg Biosignal Toolbox (AuBT) [43]. The following 7 features were extracted from each of the raw EMG signal, its first derivative and its second derivative: mean, median, standard deviation, minima, maxima, and the number of times per time unit that the signal reached the minima and the maxima. The 21 computed features were then concatenated in order to create the final feature vector.

- **EEG (Spectral):** 5 spectral features were computed from each of the theta (θ), alpha (α), beta (β), and gamma (γ) bands of each of the 14 EEG signal's channels, as described in [47], resulting in a total of 280 features. The computed features consisted of the *Spectral Flatness*, the *Spectral Bandwidth*, the *Ratio f50 vs f90*, the *Spectral Crest Factor*, and the *Spectral Roll-off*. Finally, the 280 computed features were concatenated in order to create the final feature vector.
- **EEG (MFCC):** Mel Frequency Cepstral Coefficients (MFCCs) have been used for EEG-based affect detection in various studies [48, 49, 50]. In this study, 12 MFCC features were extracted, as proposed in [48], from each EEG channel using 18 filterbanks that led to 12 cepstral coefficients. The 168 computed features were then concatenated in order to create the final feature vector. Four different MFCC feature vectors were computed over the following frequency bands of the EEG signal: [0.5-40 Hz], [4-40 Hz], [0.5-30 Hz], and [4-30 Hz].
- **EEG (PSD):** Power Spectral Density (PSD)-based EEG features have been extensively utilised for the detection of patterns in EEG signals that relate to human emotion [2, 26, 36, 49, 50]. In this study, the logarithm of the PSD of the theta (θ), 4-8 Hz, low alpha ($\tilde{\alpha}$), 8-10 Hz, alpha (α), 8-13 Hz, beta (β), 13-30 Hz, and gamma (γ), 30-64 Hz, bands of each of the 14 EEG signal's channels was computed using Welch's estimate of spectral power, leading to 5 features per channel. The 70 computed features were then concatenated in order to create the final feature vector.
- **Feature fusion:** The performance of feature fusion was also examined by creating feature vectors via concatenating individual feature vectors after normalising them to the range $[0,1]$, thus compensating for the variability in numerical range across different types of features

3.4.3 Emotion labels

The questionnaire answered by each participant of this study was used in order to extract the Valence and Arousal values associated with each activity. The mapping of the emotions included in the questionnaire to their associated Valence and Arousal values followed the one proposed in Russel's *Circumplex Model of Affect* [23], as shown in Figure 3.1. To this end, the vector (V,A), with V referring to the Valence value and A referring to the Arousal value was computed for each emotion within the questionnaire. V and A took values within the range $[-1,1]$ with negative values referring to negative valence or low arousal respectively, and positive values referring to positive valence or high arousal respectively. Since participants were allowed to select multiple emotions for each activity, the final vector (V,A) that referred to a participant's emotional state during a specific activity with a specific horse was computed as the sum of the vectors (V_k,A_k), $k=1,2,...,N$, with k being the k-th reported emotion and N being the number of different emotions reported. Finally, thresholding was used in order to convert the numerical V and A values to emotion labels. Valence was set to *Positive* for $V>0$ and to *Negative* for $V<0$, while Arousal was set to *Low* for $A<0$ and *High* for $A>0$. As a consequence, the problems of predicting the Valence and Arousal dimensions of the emotional state of a participant

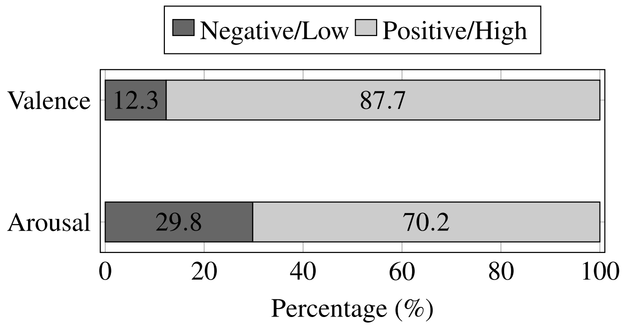

Figure 3.2 Distribution of Valence and Arousal labels.

were converted to binary problems. As a result, each feature vector in this study was associated with a Valence and an Arousal label.

After computing the Valence and Arousal labels, their distribution was examined in order to establish the class balance of the created dataset. From Figure 3.2, it is evident that the dataset is moderately unbalanced for Arousal since 70.2% of the samples refer to High Arousal and only 29.8% to Low Arousal. Furthermore, the dataset is highly unbalanced for Valence as only 12.3% of the samples refer to Negative Valence, while 87.7% refer to positive Valence. The distribution of Valance labels in this study is consistent with the outcomes of the Hama et al. study [30] which suggested that human-horse interaction is usually a pleasant experience for humans that leads to positive emotions.

The distribution of Valence and Arousal labels was also examined in relation to the participants' prior experience with horses. As can be seen in Figure 3.3, the distribution of Arousal labels was similar for both participants with and without prior experience with horses, with $\sim$70% reporting High Arousal and $\sim$30% Low Arousal. Contrary to that, participants with prior experience reported a higher percentage of ratings referring to Positive Valence (92.4%) compared to participants without prior experience (81.2%).

Figure 3.4 shows the distribution of Valence and Arousal labels in relation to the horse-related activity in order of occurrence. As shown in that figure, the vast majority of participants (94.74%) reported emotions with Positive Valence for the first activity, Looking. The percentage of samples referring to Positive Valence decreased gradually for the next two activities, reaching 89.47% for the second activity, Grooming, and 76.32% for the third activity, Leading. It is evident that the first two activities (Looking and Grooming) elicited mostly pleasant emotions to the participants. However, the percentage of reported pleasant (Positive Valence) emotions decreased for the third activity, Leading. This can be attributed to the fact that "handling" a horse can be a challenging or even scary task for inexperienced people. Contrary to Valence, Arousal exhibited the opposite behaviour. For the first activity, Looking, the percentage of Low and High Arousal ratings was almost balanced (42.11% vs 57.89%), with the percentage of High Arousal ratings increasing for the next two activities, reaching 76.32% for Grooming and 73.68% for Leading.

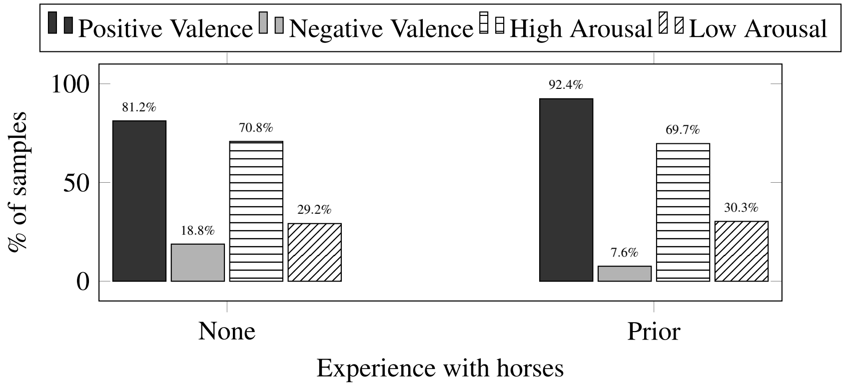

Figure 3.3 Distribution of Valence and Arousal labels in relation to the participants' prior experience with horses.

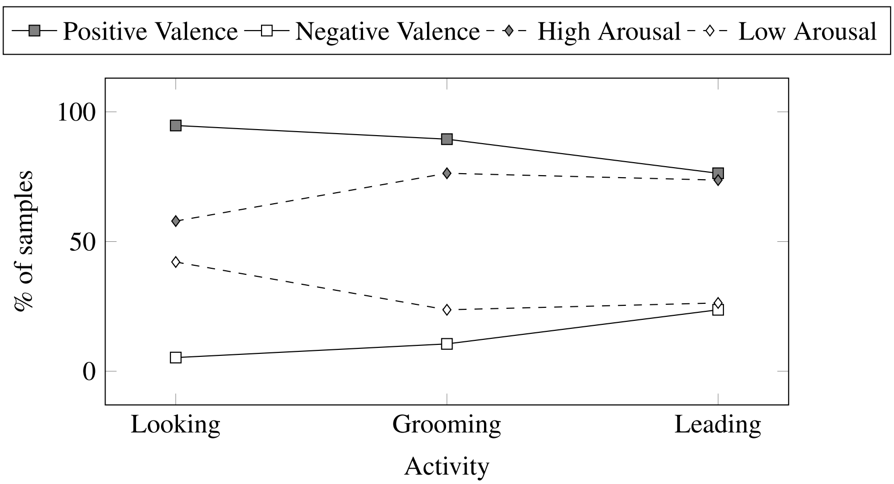

Figure 3.4 Distribution of Valence and Arousal labels in relation to the performed activity.

3.5 Experimental results

The computed feature vectors and their respective Valence and Arousal labels were used for the creation of machine learning models in order to evaluate whether the extracted features are able to characterise the emotional state of the participants when they interacted with the horses. Supervised classification experiments were conducted for the prediction of the Valence and Arousal labels using the following classification algorithms: k-Nearest Neighbour (kNN), $k = 1,3,5$, Linear Discriminant Analysis (LDA), Decision Trees (DT), Linear Support Vector Machines (LSVM), and Support Vector Machines using a Radial Basis Function kernel (SVM-RBF).

To ensure that the trained models would not be overfitted and that the performance comparison would be fair, the trained models were evaluated by following a *Leave-One-Out* (LOO) cross validation process. During the application of the LOO cross validation, one sample is used for testing the trained model and the rest of the samples for training the model. This process is repeated until all the samples have been used for testing and the overall performance is computed as the average performance across all iterations. Furthermore, due to the unbalanced dataset, classification accuracy would not be suitable as a performance metric. To compensate for the class imbalance, the classification F1-score was selected as a performance metric since it is more suitable in cases of uneven class distribution. Furthermore, the value of the F1-score metric depends on the class that is considered as positive. To this end, the reported F1-scores in this study correspond to the average F1-score between the two classes.

The performance results of the supervised classification experiments for the prediction of Valence and Arousal when using the best performing classifier for each feature set are reported in Table 3.2 in terms of classification F1-score. Apart from

Table 3.2 Valence and Arousal classification F1-score (%) for the best performing classifier for each set of features.

Features	**Valence**		**Arousal**	
	Classifier	**F1-score (%)**	**Classifier**	**F1-score (%)**
ECG	LSVM	58.09 *†‡	1NN	44.62 *†‡
EMG	LDA	52.33 *	5NN	50.75 *†
EEG (PSD)	1NN	**78.27** *†‡	DT	56.71 *†‡
EEG (Spectral)	1NN	71.50 *†‡	DT	64.38 *†‡
EEG (MFCC 4-40 Hz)	DT	68.49 *†‡	LSVM	63.75 *†‡
EEG (MFCC 0.5-40 Hz)	LSVM	68.49 *†‡	LDA	**65.49** *†‡
EEG (MFCC 4-30 Hz)	LSVM	55.21 *†‡	LDA	55.88 †‡
EEG (MFCC 0.5-30 Hz)	DT	68.27 *†‡	LDA	61.99 *†‡
All	LSVM	65.91 *†‡	LSVM	59.29 *†‡
ECG/EMG/EEG (PSD)	1NN	76.72 *†‡	5NN	54.52 *†
ECG/EMG/EEG (Spectral)	1NN	76.72 *†‡	3NN	55.02 *†‡
ECG/EMG	5NN	59.67 *	3NN	49.71 *†‡
EEG (ALL)	LSVM	67.40 *†‡	LSVM	61.62 *†‡
	Random	41.71	Random	47.88
	Majority	46.73	Majority	41.24
	Class Ratio	50.00	Class Ratio	50.00

$\star$ Statistically significant difference compared to random voting ($p < 0.05$)
† Statistically significant difference compared to majority voting ($p < 0.05$)
‡ Statistically significant difference compared to class ratio based voting ($p < 0.05$)

the single modality features, results for some feature fusion approaches, as well as results for voting randomly (50% probability for each class), voting according to the majority class in the training data (100% probability of the majority class), and voting according to the class ratio (the probability of each class is equal to its ratio of samples within the training set), are also reported in Table 3.2. It is evident that the highest classification F1-scores for both Valence and Arousal were achieved using single modality features, with fusion approaches performing slightly worse. Classification F1-score for Valence reached 78.27% using the 1-NN classifier, while an F1-score of 65.49% was achieved for Arousal using the LDA classifier. The best performing fusion approach for Valence yielded an F1-score of 76.72% using the 1-NN classifier using the fusion of the ECG, EMG, and EEG (PSD) features, as well as the fusion of the ECG, EMG, and EEG (Spectral) features. Furthermore, the best performing fusion approach for Arousal yielded an F1-score of 61.62% using the LSVM classifier and the fusion of all the computed EEG features.

Considering the class imbalance within the dataset and to ensure that the acquired results were significantly different from the results for random voting, majority voting, and class ratio based voting, the acquired results were compared to the analytically computed results for these three cases, reported in Table 3.2. An unpaired Kruskal-Wallis test was used to test for significance against the predicted class labels for random voting and for class ratio based voting, whereas a paired Wilcoxon signed-rank test was used to test for significance against majority voting since the predicted class labels could be computed definitely on a one-by-one basis. As shown in Table 3.2, all settings performed significantly better than random voting ($p < 0.05$) for both Valence and Arousal. Regarding majority voting, all settings performed significantly better for Arousal, whereas for Valence, only the EMG-based features and the fusion of ECG and EMG features failed the significance test ($p \geq 0.05$). Regarding class ratio based voting, for Valence, only the EMG-based features and the fusion of ECG and EMG features failed the significance test ($p \geq 0.05$), while for Arousal,

Table 3.3 Classification F1-scores (%) reported in the literature for Valence and Arousal when physiological signals are used.

Approach	Stimulus	Brain signal device	Valence F1-score	Arousal F1-score
AMIGOS [29]	Film clips	Emotiv EPOC (EEG)	56.40	57.70
Arnau et al. [27]	Music videos	Biosemi Active II (EEG)	69.20	66.70
DEAP [2]	Music videos	Biosemi Active II (EEG)	60.80	58.30
DECAF [28]	Film clips	ELEKTA Neuromag (MEG)	59.00	58.00
DREAMER [3]	Film clips	Emotiv EPOC (EEG)	53.05	57.98
MAHNOB [26]	Film clips	Biosemi Active II (EEG)	56.00	42.00
This study	Horses	Emotiv EPOC+ (EEG)	78.27	65.49

Note: F1-scores refer to the highest performance achieved when using only physiological signals only the EMG-based features and the fusion of ECG, EMG, and EEG (PSD) features failed the significance test ($p \geq 0.05$). It must be noted that the best performing settings for both Valence and Arousal performed significantly better than all the three examined cases.

3.6 Discussion

The classification performance achieved in this study was compared with the results reported in other studies ([2, 3, 26, 27, 28, 29]) that used physiological signals for the task of predicting Valance and Arousal. From Table 3.3, it can be seen that the reported results are consistent with the ones reported in the literature, providing evidence that the utilised physiological signal-based features are suitable for the task of emotion recognition during human-horse interaction and are able to provide performance on par with the state-of-the-art. Furthermore, the fact that this performance was achieved using low-cost portable devices for signal acquisition, indicates that such devices are potentially a suitable alternative to expensive non-portable medical-grade devices such as the ones used in [2, 26, 27], a finding that is also consistent with the findings of [3] and [29]. The use of portable wearable low-weight sensors is of utmost importance for applying affective computing techniques within EAT studies, since in most cases, the study of human-horse interaction requires that the ability of the users to move is not restricted. Being able to monitor the emotional responses of people interacting with horses can potentially be highly beneficial to quantitative studies on equine assisted therapy, by providing the means to study such complex emotional responses.

3.7 Conclusion

This chapter provided an experimental approach for the task of emotional state detection via physiological signal analysis and machine learning. Experimental results showed that the described approach is suitable for the detection of the emotional state of an individual during interaction with horses, as it achieves performance on par with the state-of-the-art in emotion recognition literature. The described methodology can potentially benefit the field of equine-assisted therapy (EAT) as it constitutes a quantitative method for assessing the emotional state of people undergoing EAT, thus providing the means to measure its effects.

References

[1] Picard RW. Affective computing: from laughter to IEEE. IEEE Trans Affect Comput. 2010 Sep;1(1):11–17.

[2] Koelstra S, Muhl C, Soleymani M, et al. DEAP: A Database for Emotion Analysis Using Physiological Signals. IEEE Trans Affect Comput. 2012 Jan;3(1):18–31.

[3] Katsigiannis S, Ramzan N. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE J Biomed Health Informat. 2018 Jan;22(1):98–107.

[4] Ramzan N, Palke S, Cuntz T, et al. Emotion recognition by physiological signals. In: Proc. IS&T International Symposium on Electronic Imaging. Society for Imaging Science and Technology; 2016. .

[5] Morrison ML. Health benefits of animal-assisted interventions. Complementary Health Practice Review. 2007;12(1):51–62.

[6] Amiot CE, Bastian B. Toward a psychology of human-animal relations. Psychol Bull. 2015;141(1):6.

[7] Acri M, Hoagwood K, Morrissey M, et al. Equine-assisted activities and therapies: enhancing the social worker's armamentarium. Social Work Education. 2016;35(5):603–612.

[8] Kendall E, Maujean A, Pepping CA, et al. Hypotheses about the psychological benefits of horses. Explore. 2014;10(2):81–87.

[9] Haskin MR. Therapeutic horseback riding for the handicapped. Arch Phys Med Rehabil. 1974;55:473–474.

[10] Woods D. Horseriding catching on as a therapy for the disabled. Can Med Assoc J. 1979;121(5):631–650.

[11] Mayberry RP. The mystique of the horse is strong medicine: riding as therapeutic recreation. Rehabil Lit. 1978;39(6-7):192–196.

[12] Masini A. Equine-assisted psychotherapy in clinical practice. J Psychosoc Nurs. 2010;48(10):30–34.

[13] Yorke J, Nugent W, Strand E, et al. Equine-assisted therapy and its impact on cortisol levels of children and horses: A pilot study and meta-analysis. Early Child Dev Care. 2013;183(7):874–894.

[14] Earles JL, Vernon LL, Yetz JP. Equine-assisted therapy for anxiety and post-traumatic stress symptoms. J Trauma Stress. 2015;28(2):149–152.

[15] Kemp K, Signal T, Botros H, et al. Equine facilitated therapy with children and adolescents who have been sexually abused: A program evaluation study. J Child Fam Stud. 2014;23(3):558–566.

[16] Schroeder K, Stroud D. Equine-facilitated group work for women survivors of interpersonal violence. Journal for Specialists in Group Work. 2015;40(4):365–386.

[17] Benda W, McGibbon NH, Grant KL. Improvements in muscle symmetry in children with cerebral palsy after equine-assisted therapy (hippotherapy). J Altern Complem Med. 2003;9(6):817–825.

[18] Klontz BT, Bivens A, Leinart D, et al. The effectiveness of equine-assisted experiential therapy: Results of an open clinical trial. Soc Anim. 2007;15(3):257–267.

[19] Selby A, Smith-Osborne A. A systematic review of effectiveness of complementary and adjunct therapies and interventions involving equines. Health Psychology. 2013;32(4):418.

[20] Xenophon. The art of horsemanship (Translation by Morris H. Morgan). Dover Publications Inc.; 2006.

[21] Duarte D. The Royal Book of Jousting, Horsemanship, and Knightly Combat: A Translation Into English of King Dom Duarte's 1438 Treatise Livro Da Ensinança de Bem Cavalgar Toda Sela (The Art of Riding in Every Saddle). Preto AF, editor. The Chivalry Bookshelf; 2010.

[22] de Pluvinel A. Le Maneige Royal (Translated by Hilda Nelson from a 1626 edition). Franktown, VA, USA: Xenophon Press; 2010.

[23] Russell JA. A circumplex model of affect. J Pers Soc Psychol. 1980;39(6):1161–1178.

[24] Zeng Z, Pantic M, Roisman GI, et al. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Trans Pattern Anal Mach Intell. 2009 Jan;31(1):39–58.

[25] Gunes H, Schuller B. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. Image Vision Comput. 2013;31(2):120 – 136.

[26] Soleymani M, Lichtenauer J, Pun T, et al. A Multimodal Database for Affect Recognition and Implicit Tagging. IEEE Trans Affect Comput. 2012 Jan;3(1):42–55.

[27] Arnau-González P, Arevalillo-Herráez M, Ramzan N. Fusing highly dimensional energy and connectivity features to identify affective states from EEG signals. Neurocomputing. 2017;244:81 – 89.

[28] Abadi MK, Subramanian R, Kia SM, et al. DECAF: MEG-based multimodal database for decoding affective physiological responses. IEEE Trans Affect Comput. 2015 Jul;6(3):209–222.

[29] Correa JAM, Abadi MK, Sebe N, et al. AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups. IEEE Trans Affect Comput. 2018;(Early Access).

[30] Hama H, Yogo M, Matsuyama Y. Effects of stroking horses on both humans' and horses' heart rate responses. Jpn Psychol Res. 1996;38(2):66–73.

[31] Chen CCJ, Crews D, Mundt S, et al. Effects of equine interaction on EEG asymmetry in children with autism spectrum disorder: a pilot study. Int J Dev Disabil. 2015;61(1):56–59.

[32] Guidi A, Lanata A, Baragli P, et al. A wearable system for the evaluation of the human-horse interaction: a preliminary study. Electronics. 2016;5(4):63.

[33] Lanata A, Guidi A, Valenza G, et al. Quantitative heartbeat coupling measures in human-horse interaction. In: Proc. IEEE EMBC. Orlando, FL, USA; 2016. .

[34] Lanata A, Guidi A, Valenza G, et al. The role of nonlinear coupling in Human-Horse Interaction: A preliminary study. In: Proc. IEEE EMBC. Seogwipo, South Korea; 2017. .

[35] Althobaiti T, Katsigiannis S, West D, et al. Affect Detection for Human-Horse Interaction. In: Proc. 21st Saudi Computer Society National Computer Conference (NCC). Riyadh, Saudi Arabia: IEEE; 2018. p. 1–6.

[36] Althobaiti T, Katsigiannis S, West D, et al. Examining Human-Horse Interaction by Means of Affect Recognition via Physiological Signals. IEEE Access. 2019;7:77857–77867.

[37] Burns A, Greene BR, McGrath MJ, et al. SHIMMER - A Wireless Sensor Platform for Noninvasive Biomedical Research. IEEE Sens J. 2010 Sept;10:1527–1534.

[38] Badcock NA, Mousikou P, Mahajan Y, et al. Validation of the Emotiv EPOC EEG gaming system for measuring research quality auditory ERPs. PeerJ. 2013 Feb;1(e38).

[39] Wijayasekara D, Manic M. Human machine interaction via brain activity monitoring. In: Proc. IEEE HSI. Sopot, Poland; 2013. .

[40] Russell JA, Mehrabian A. Evidence for a three-factor theory of emotions. J ResPers. 1977;11(3):273 – 294.

[41] Blanco-Velasco M, Weng B, Barner KE. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. Comput Biol Med. 2008;38(1):1 – 13.

[42] Kannathal N, Acharya UR, Joseph KP, et al. Analysis of electrocardiograms. In: Advances in Cardiac Signal Processing. Springer; 2007. p. 55–82.

[43] Wagner J. Augsburg Biosignal Toolbox (AuBT); 2005. University of Augsburg.

[44] Bigdely-Shamlo N, Mullen T, Kothe C, et al. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Front Neuroinform. 2015;9:16.

[45] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Meth. 2004;134(1):9–21.

[46] AlZoubi O, D'Mello SK, Calvo RA. Detecting Naturalistic Expressions of Nonbasic Affect Using Physiological Signals. IEEE Trans Affect Comput. 2012 Jul;3(3):298–310.

[47] Monge-Álvarez J, Hoyos-Barceló C, San José-Revuelta LM, et al. A Machine Hearing System for Robust Cough Detection Based on a High-Level Representation of Band-Specific Audio Features. IEEE Trans Biomed Eng. 2018;(Early Access), DOI: 10.1109/TBME.2018.2888998.

[48] Piciucco E, Maiorana E, Falzon O, et al. Steady-State Visual Evoked Potentials for EEG-Based Biometric Identification. In: Proc. BIOSIG. Darmstadt, Germany: Gesellschaft für Informatik, Bonn; 2017. .

[49] Arnau-González P, Arevalillo-Herráez M, Katsigiannis S, et al. On the influence of affect in EEG-based subject identification. IEEE Transactions on Affective Computing. 2018;p. 1–1.

[50] Arnau-González P, Katsigiannis S, Arevalillo-Herráez M, et al. Image-Evoked Affect and its Impact on Eeg-Based Biometrics. In: 2019 IEEE International Conference on Image Processing (ICIP); 2019. p. 2591–2595.