Group Synchrony for Emotion Recognition using Physiological Signals

Patrícia Bota, *Student Member, IEEE,* Tianyi Zhang, *Member, IEEE,* Abdallah El Ali, *Member, IEEE,* Ana Fred, *Member, IEEE,* Hugo Plácido da Silva, *Senior Member, IEEE,* and Pablo Cesar, *Senior Member, IEEE*

Abstract—During group interactions, we react and modulate our emotions and behaviour to the group through phenomena including emotion contagion and physiological synchrony. Previous work on emotion recognition through video/image has shown that group context information improves the classification performance. However, when using physiological data, literature mostly focuses on intrapersonal models that leave-out group information, while interpersonal models are unexplored. This paper introduces a new interpersonal Weighted Group Synchrony approach, which relies on Electrodermal Activity (EDA) and Heart-Rate Variability (HRV). We perform an analysis of synchrony metrics applied across diverse data representations (EDA and HRV morphology and features, recurrence plot, spectrogram), to identify which metrics and modalities better characterise physiological synchrony for emotion recognition. We explored two datasets (AMIGOS and K-EmoCon), covering different group sizes (4 vs dyad) and group-based activities (video-watching vs conversation). The experimental results show that integrating group information improves arousal and valence classification, across all datasets, with the exception of K-EmoCon on valence. The proposed method was able to attain mean M-F1 of $\approx$ 72.15% arousal and 81.16% valence for AMIGOS, and M-F1 of $\approx$ 52.63% arousal, 65.09% valence for K-EmoCon, surpassing previous work results for K-EmoCon on arousal, and providing a new baseline on AMIGOS for long-videos.

Index Terms—Emotion Recognition, Physiological Synchrony, Physiological Signals, Machine Learning, Deep Learning, Group Emotion.

1 Introduction

HUMANS are social beings, spending a large amount of time in collective activities, either at work, for leisure or at home [1]. In such contexts, our emotions are adapted to the group and its members [2], [3]. A group can be considered to be any arrangement from a small cluster of two individuals to thousands of people in physical presence or with the ability to interact [4]. The literature on collective emotions [2], [3] reports that during social contexts (e.g. face-to-face encounters or public gatherings) a "macro-level affective process" denoted as collective emotions can emerge. During this process, the dynamics among group members can lead to phenomena such as "emotional cascades", "emotional contagion" or "collective effervescence" [2], [3]. Thus, group behaviour is a component that should be analysed when performing emotion assessment.

Group emotion recognition using audiovisual sources is vastly explored, largely motivated by challenges such as the Emotion Recognition in the Wild (EmotiW), which focused on group emotion analysis using images ([5] in 2018), and audio and video ([6] in 2020). Within this challenge, hybrid approaches – combining information from both the individual-level emotion and environment context – have shown to result in overall higher accuracy and have become the predominant approach. However, these approaches are mostly tested for images/video, which focused on overt (visible) behavioural features, and not on physiological signals which are associated with the autonomic nervous system [7].

The literature on collective emotions [8] reports that in group scenarios, individuals have shown spontaneous and unintended similarities in their physiological and behavioural responses [9] – a phenomena denoted as physiological synchrony [10], i.e. the inter-dependency or temporal interaction between the physiological data of two or more individuals.

To define emotional experiences, we follow the approach by [11], [12], [13], which relies on a "consensual, componential theory of emotion" [11]. In this approach, an emotional response involves subject experience, physiology (peripheral and central nervous system), and a behavioural component, each of these systems being associated with a discrete pattern. We perform emotion recognition based on peripheral physiology patterns as all current major theories of emotion consider physiological responses to be a component of emotion [12], and rely on arousal and valence dimensions from the dimension theory of emotion and core affect to characterize emotional states. Physiological synchrony has been identified in numerous works through peripheral data across different types of relations, such as parent-child, couples, therapist-client, or social interactions [14], [15]. However, there is a gap in the use of unobtrusive physiological signals for emotion recognition exploring

- *P. Bota, A.Fred and H. P. Silva are with Instituto Telecomunicações & Instituto Superior Técnico, Lisbon, Portugal. E-mail: patricia.bota@tecnico.ulisboa.pt*
- *P. Cesar, T. Zhang, are with Centrum Wiskunde & Informatica Amsterdam, The Netherlands & Multimedia Computing Group, Delft University of Technology, 2600AA Delft, The Netherlands. E-mail: {tianyi.zhang, p.s.cesar}@cwi.nl.*
- *A. Ali is with Centrum Wiskunde & Informatica Amsterdam, The Netherlands. E-mail: abdallah.el.ali@cwi.nl.*

Manuscript received April 19, 2005; revised August 26, 2015.

group interactions, which is the open problem addressed by our work.

Physiological data reflects emotional state changes as a result of the *Autonomic Nervous System* (ANS) activity due to endogenous and exogenous conditions [16]. The ANS activity can be assessed through physiological measures such as *Electrodermal Activity* (EDA), related with the *Sympathetic Nervous System* (SNS), or *Heart Rate Variability* (HRV), related with both the SNS and *Parasympathetic Nervous System* (PNS). The use of EDA and HRV-related features shows many advantages, such as: 1) high temporal and amplitude resolution; 2) continuous and unobtrusive data collection over long periods of time in daily living; and 3) have been proven to be an insightful view into the subject emotions [17]. However, when compared to audiovisual content, the field of emotion recognition using physiological data has mostly focused on intrapersonal emotion assessment [17], [18], disregarding group-related phenomena such as emotion contagion and physiological synchrony.

We fill this gap by proposing a novel methodology that explores physiological synchrony, by performing a weighted average of the groups' emotion class labels to predict the label of an unknown subject. The weights are given by the physiological synchrony between the unknown subject and each member of the group. Throughout our work, we address the following research questions:

> **RQ1:** What synchronisation metrics and data representations are most suitable for measuring physiological synchrony for emotion recognition?
> **RQ2:** Does the emotion classification accuracy improve with the inclusion of group-level information?

In this work, we address these research questions and fill the gap identified in the state of the art regarding group emotion recognition based on unobtrusive physiological data (EDA and HRV). Moreover, we study the potential to improve the accuracy of emotion recognition systems by integrating group context information through a novel metric combining weighted group physiological synchrony.

The remainder of this paper is organised as follows: In Section 2 we describe the background and literature on interpersonal emotion recognition. Section 3 describes the overall pipeline of the proposed methodology. In Section 4, we evaluate our methodology against two datasets obtained across different group settings. Lastly, in Section 5, we present our main conclusions along with future work directions.

2 Background

2.1 Group Emotion Recognition

The use of the group context has been successfully employed for emotion recognition in the field of audiovisual content analysis [19], namely through hybrid and top-down approaches [20], [21]. In these approaches, in addition to the subjects' facial expressions, they rely on information from the global scene, skeleton features, and visual attention mechanisms applied to the entire image. These global input sources are combined to perform the emotion classification tasks. Similarly, in the field of emotion recognition from speech, group information has also been taken into consideration, namely in dyad conversations [19], [22].

In [19], the authors apply a *Support Vector Machine* (SVM) to predict the listener/speaker emotion by using the facial features and the emotion prediction of the speaker/listener using a SVM over the listener acoustic features. The classification improved when including cross-subject features. In [23], the authors analyse whether the emotional reaction of one individual can be assessed by the emotional response of their partner, in a dyad cooperation task, exploring physiological and speech data. The models were trained to predict emotional and non-emotional moments using a linear SVM and a Random Forest classifier. The results showed that the emotion classification performance increases when combining information from the two subjects.

In [22], the authors incorporate time-lagged Cosine similarity features on a latent representation from an adapted ResNet architecture performing emotion recognition during dyad conversations using video and audio data. The experimental results showed that the interpersonal method outperformed the model based on individual features only.

The aforementioned works confirm the success of applying group information for emotion assessment, however, they are based on audio-visual or speech features and focus mostly on dyads. The study of interpersonal features extracted from physiological data in groups larger than two was only found in the work by [18]. In [18], the authors assess the individual's multi-label categorical emotional state using speech and *Photoplethysmography* (PPG) during group tasks used as input to a transformer encoder block with positional encoding, followed by a bi-*Long Short-Term Memory* (LSTM) model. The method takes into consideration the group atmosphere given by the aggregation of each group-member score in a Self-supervised Graph Attention Networks (SuperGAT), surpassing all the baseline methods in the NTUBA dataset [24]. The literature lacks further validation as it was only tested for a few datasets/use cases (e.g. of three-person small group conversations in [24]). Additionally, the latter work relies on external displays of affection (speech) and does not explore alternative similarity metrics to Cosine similarity.

The review of the literature shows that the integration of group information in emotion classification tasks can improve the classification performance, namely in audiovisual and dyad conversations. Collective emotion recognition based on physiological data is still largely unexplored, and there is a lack of information regarding which synchronisation metrics and data representations better describe physiological synchrony, and whether they are replicable across different group-related activities (i.e. conversation versus watching a movie). In this paper, we address each of these issues by: 1) proposing a novel approach integrating group context for emotion recognition using physiological data collected unobtrusively; 2) performing a diverse analysis of physiological measures and data representations; and 3) applying our method across two datasets acquired under different group use cases.

2.2 Metrics for Physiological Synchrony

In the literature, a broad range of physiological signals have been used to analyse physiological synchrony including

[16]: cardiovascular (e.g. *Electrocardiography* (ECG), PPG and HRV-related); respiratory (e.g. respiratory rate, respiratory volume time); electrodermal activity (e.g. *Electrodermal Response* (EDR), *Electrodermal Level* (EDL)); and thermal (e.g. skin temperature). In this work, we follow the approach adopted by [16] and focus on physiological signals aligned with the physiological constructs of interpersonal synchronisation. Furthermore, we focus on signals that can be obtained unobtrusively and continuously from a group, and that have low latency so that they can be applied in real-time and in daily living. Our selected signals are the EDA and cardiovascular activity namely, HRV.

In a survey analysing over 61 works [16], the authors report high ambiguity in using EDA to assess physiological synchrony, with many papers identifying synchrony in dyads both using skin conductance [25], and response [26], while others not (in skin conductance) [27]. Similar findings are obtained for inter-group analysis, with [28] identifying synchronisation even between strangers, unlike the authors in [29]. Although the experimental results are not definite, there is evidence of physiologically-related synchrony through EDA measures [15].

Cardiovascular activity can be assessed through *Heart Rate* (HR) [30], [31], inter-beat interval [32], [33], HRV-related features [33], [34]. Similarly to the EDA signal, there is still little consensus in the literature. Physiological synchrony is identified in dyad conversations through differential equations in [35], audience members and dancers [36] in R-peaks through regression. The opposite is described in [37], where no synchrony was identified in groups of 10 individuals at rest and listening to music. For a more detailed description of the works and findings in the literature pertaining physiological synchrony and related areas, we refer the reader to [16].

Overall, although the literature on physiological synchrony is unclear due to factors such as the diversity of signal sources, analysis metrics, protocol setups, or also likely due to the task itself or activity that was studied, numerous papers confirm the existence of physiological synchrony in EDA and HRV [16], [35].

2.3 Datasets

The literature [8] reports that group dynamics such as emotion contagion can occur even without face-to-face interactions or non-verbal clues (e.g. social media). Nummenmaa et al. [38] describe five types of physiological synchrony in groups: a) independent units (group sharing physical presence but in independent tasks); b) externally driven (e.g. group watching a movie); c) leader-follower/sequential interaction (e.g. meeting); d) dynamic interaction (e.g. conversation); and e) group interaction (e.g. group cooperative tasks) [39].

In this article, we perform an analysis of two of these conditions (dynamic interaction in dyad conversation; and externally driven by video watching) by using different datasets. The datasets were selected according to the following requirements: 1) contain group data (2 individuals); 2) contain unobtrusive physiological data, namely EDA and cardiovascular data; and 3) are continuously annotated in terms of arousal and valence for long-duration naturalistic scenarios (i.e. 10 minutes) to elicit group-emotion related phenomena.

The selected datasets were AMIGOS and K-EmoCon. AMIGOS [40] contains physiological (EDA, ECG, *Electroencephalography* (EEG)) and audio-visual (face and full-body video) data collected both in groups (4 individuals) and individual settings. The dataset includes data from 37 subjects watching 4 long videos (> 14 minutes) and 16 short-video clips (< 4 minutes). The data was continuously annotated 2 [-1, 1] in valence/arousal by 3 experts at 20-second intervals. Data from the short-videos experiment, which was acquired only individually was removed. K-EmoCon [41] contains physiological (EDA, PPG, EEG), and audiovisual (face, gesture, speech) data collected during naturalistic dyad conversations, namely a debate on social issues. The dataset contains 16 sessions of approximately 10 minutes each. The data was annotated by self-report, debate partner, and external annotator in a [1, 5] scale at 5 seconds intervals, both using a valence/arousal space and 18 categorical emotions.

Herein, we use the external annotations to maintain the coherency between the two datasets. A summary description of the datasets is shown in Table 1. The labels were divided in binary classification, taking -1: < 0; 1: >= 0 for valence and -1: <= 0; 1: > 0 for arousal.

TABLE 1: Datasets' labels summary. A: Arousal; V: Valence. Group Homogeneity refers to the % of samples in which all members in the group (except the unknown subject) have equal class labels. A sample size of 20 seconds was considered for both datasets.

Label	AMIGOS		K-EmoCon	
	A	V	A	V
-1	81.12	71.23	80.67	08.08
1	18.88	28.76	19.32	91.92
Num. samples	43327		3192	
Group Homogeneity	70.50 ± 7.10	69.19 ± 7.40		

3 METHODOLOGY

We denote our proposed approach as *Weighted Group Synchrony* (WGS), described in Equation 1. In a group context of N subjects, the label of an unknown subject ($\hat{y}_s$) is given by the weighted average of the remaining group members' labels:

$$\hat{y}_s = \sum_{i=1}^{N} {}^{1} W_i y_i \tag{1}$$

Where W are the weights denoting the synchronisation between the unknown individual s and each of the remaining group-member i. The physiological synchronisation is given by:

$$W_i = S(h_i; h_s) \tag{2}$$

Where h is the data representation and S is the similarity metric used to obtain the synchronisation between two subjects. When the similarity metric returns a distance instead of correlation, it is converted by $S = \frac{1}{1+distance}$. The $\hat{y}_s$ 2 f 1, 1g, consists of a binary problem. When a negative correlation occurs, the subject s is given the opposite label

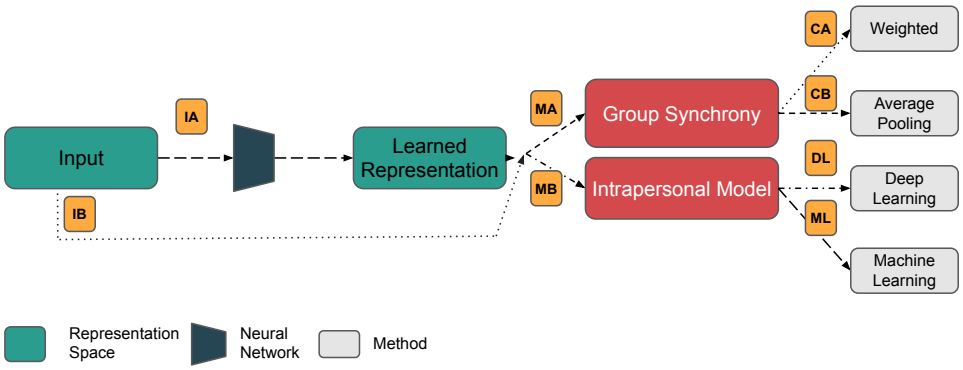


Fig. 1: Pipeline of the implemented methodologies. The squares in orange refer to alternative pathways. The IA and IB paths refer to two alternatives for the input formats: with and without crossing a neural network. The MA and MB pathways refer to the application of interpersonal and intrapersonal models, respectively. When applying MA (interpersonal model), we tested both a weighted and non-weighted (average pooling) approach. For the intra-personal model, the application of deep learning (DL) and machine learning (ML) was tested.

of subject i while, for a positive correlation, the label of the subject i is given. Then, the assigned labels are added and weighted by the synchronisation value of each subject for the computation of the weighted average of the subject s label. When the group consists of a dyad, W is always 1 or 1, and the unknown subject is given the label of the other member in case of a positive correlation, or the opposite label for a negative and zero correlation.

3.1 Pipeline

An overview of our tested methods is shown in Fig. 1. We start by testing two alternative inputs: The *IA* pathway, where the EDA morphology (EDA, EDR and EDL signals), EDA and HRV hand-crafted features, and images (EDA spectrogram and recurrence plot) are used as input to a neural network where a higher representation is learned to compute the subjects' emotion label in the *MA* or *MB* steps. Alternatively, in path *IB* no latent representation is used and the data inputs for the neural network (namely, EDA morphology – EDL and EDR; and EDA and HRV hand-crafted features) are used as input for the *MA* or *MB* steps to obtain the subjects' emotion label.

After the input representation is defined (*IA* vs *IB*), two pathways are proposed, *MA* and *MB*, in which the interpersonal and intrapersonal models are tested, respectively. If the *MA* path is taken (interpersonal model), group synchrony is performed where the subject emotion classification is performed based on the synchronisation between the unknown subject sample and the group members samples for each timestamp (WGS method). The group synchronisation method was tested in two pathways: weighted (*CA*) – where a weight is given according to the synchronisation value; and average pooling (*CB*) – where a non-weighted average is performed so that the synchronisation metrics are not considered.

If the *MB* path is selected (intrapersonal model), two methods were tested: classification by classic machine learning algorithms (*ML*); or the implementation of a deep learning classifier (*DL*), using the feature extraction layer from the *IA* path with the addition of a sigmoid activation function to get a binary arousal/valence classification.

3.2 Synchronisation Metrics

To measure physiological synchrony, we considered a set of seven synchronisation metrics following two criteria: 1) The six physiological constructs of physiological synchrony identified in [16] – magnitude, sign, direction, lag, timing and arousal. Magnitude is determined through Pearson, Cosine similarity and Euclidean distance. Sign is determined through Spearman correlation. Direction and Lag is determined through *Dynamic Time Warping* (DTW). Arousal is determined by using EDA data; and 2) Synchrony metrics found in previous works on the study of physiological synchrony confirming its existence in EDA and HRV.

1) **Pearson Correlation** *2* **[-1, 1]**: measures the linear correlation between two signals, from negatively correlated to a perfect correlation. Pearson correlation can be interpreted as synchronisation magnitude, being one of the most commonly applied metrics in the literature as shown by [42], [43].
2) **Spearman Rank Correlation** *2* **[-1, 1]**: analyses the rank-correlation between two signals, from negatively correlated to a perfect correlation, enabling the measurement of the synchrony sign value, i.e. whether signals have the same or opposite dynamics and is applied in [44], [45].
3) **Cosine Similarity** *2* **[-1, 1]**: measures the normalised inner product between two signals, and has been used as a magnitude construct of physiological synchrony in [18].
4) **Euclidean Distance** *2* **[0, + *1*]**: measures the pythagorean distance between two signals and has been used as a magnitude construct of synchrony in [46].
5) **Recurrence Plot,** R^6 *2* **[0, + *1*]**: The aforementioned metrics are linear, fitting for stationary data with constant mean and variance throughout time. However, physiological data is non-stationary and can show temporal dependency [16]. Recurrence plots allow the characterisation of temporal cyclic trends in signals, by filling in the times in which a phase-space trajectory is repeated. Recurrence plots can be

found in the literature in [47], [48]. To compare different recurrence plots we extracted six recurrence quantification analysis metrics: Recurrence rate, Determinism, Average diagonal line length, Longest diagonal line length, Divergence and Entropy diagonal lines[1].

6) **DTW** *2* **[0, + *1*]**: computes the distance between two signals, but instead of calculating the vertical Euclidean distance between the signals, calculates the Euclidean distance across the smallest paths, allowing a temporal synchronisation between the signals. The DTW takes into consideration the timing and lag construct of synchronisation. DTW is applied in [49], [50].
7) **Cross-correlation** *2* **[0, + *1*]**: takes into consideration a lag parameter of physiological synchrony to compute the time-shifted correlation between the two signals. We do so by using SciPy correlate function[2] with mode equal to full, from which we obtain the maximum value to identify the moment of maximum synchrony.
8) **Coherence (Spectral Correlation)** *2* **[-1, 1]**: consists of Pearson correlation computed in the frequency domain. The use of spectral metrics is described in the literature to assess synchrony magnitude in [51], [52].

By analysing such diverse similarity metrics that explore different characteristics of the data, our work further expands the state-of-the-art of emotion recognition using unobtrusive physiological signals.

3.3 Data Representation

Given the diversity found in the literature of physiological synchrony and to analyse the state-of-the-art open question (RQ1) of which data representation better expresses group physiological synchrony for emotion recognition, we explore diverse data representations:

Signal Morphology

Corresponds to the cleaned and processed signal morphology. For the AMIGOS dataset [40], we used the processed data given by the authors. Regarding the EDA data, we removed existent spikes using the modified Z-score[3], the signal was filtered using a Butterworth low-pass filter of 4^{th} order with a cut-off frequency of 5 Hz, and a smoother filter with a window of 0.25 seconds. Afterwards, the signal was normalised ($\frac{y}{\_\_}$, : sample mean; : sample standard deviation) for each trial. The ECG signal was filtered using a FIR bandpass filter *2* [3, 45] Hz and the R-peaks were computed using the BioSPPy Hamilton segmenter [53]. For the K-Emocon dataset, we used the PPG to extract the heartbeat peaks, which was filtered using a Butterworth bandpass filter with 1-8Hz cutoff of 4^{th} order. The heartbeat peaks were extracted using the BioSPPy extractor [54][4]. According to [55], the duration of an emotional response ranges from 0.5 to 4 seconds, reproducing changes in physiological data from 3 to 15 seconds [56]. In both datasets, the data were segmented in 20 seconds windows with 75% overlap. The EDA data was decomposed into the EDR and EDL components using the cvxEDA library [57].

Image

We use spectrograms to collect spectral information and recurrence plots which allow to characterise non-periodic and non-stationary signals. Both have been applied for emotion recognition with state-of-the-art results in the works of [58], [59].

Hand-crafted Features

We extracted a total of 26 features from the EDA data based on the work by [60], and 76 features from the ECG interbeat intervals[5]. Redundant features (with > 85% correlation) were removed, resulting respectively in 21, and 31 features for the EDA and HRV (Table 1 in the Appendix Section of the Supplementary Material).

An example with the data representations is shown in Fig. 2. Fig. 2 – a, shows the morphological space representation (EDA, EDR and EDL) for one sample (20 seconds) extracted from the AMIGOS dataset. Fig. 2 – b and c, display a spectrogram and recurrence plot applied on the signal from Fig. 2 – a, using a viridis colormap.

Signal morphology (Fig. 2 – a) and hand-crafted features, were used as the data space input to compute the synchrony between two subjects in our WGS method (*MA* path – interpersonal model). These two spaces, along with the image-based space (Fig. 2 – b and c) were used as input for a binary arousal/valence classifier, for both a final classification in the *MB* path – intrapersonal model; and to learn a higher level representation in which the WGS was applied (*MA* path – interpersonal model).

Table 2 displays a summary with the synchronisation metrics applied for each data type in the *MA* path – interpersonal model. The data types were divided into morphology and feature-based. The morphology space includes the EDA components (EDR and EDL) through path *IA* and *IB* in Fig. 1, while the feature-based includes EDA and HRV features used also in both *IA* and *IB* paths in Fig. 1. The image representation is not included since it was not used to obtain the subjects' synchronisation, only as input for the neural network in the *IA* path in Fig. 1, where a feature-based latent representation was learned and then used to calculate the physiological synchronisation and the emotion classification label.

3.4 Classification Models

The state of the art of emotion recognition based on physiological data relies, mostly, on the use of artificial intelligence algorithms incorporating the individual's data with no group context. In our work, we start by replicating the traditional approach found in the state of the art [17], which does not consider group dynamics in its architecture. We denote this approach as the intrapersonal methodology, which

1. github.com/bmfreis/recurrence_python; Accessed: 26/08/2022
2. docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlate.html; Accessed: 12/01/2023
3. towardsdatascience.com/removing-spikes-from-raman-spectra-\\8a9fdda0ac22; Accessed: 26/08/2022
4. github.com/scientisst/BioSPPy; Accessed: 26/08/2022
5. pyhrv.readthedocs.io/en/; Accessed: 26/08/2022

TABLE 2: Synchrony metrics applied in each data space in the interpersonal model. The data spaces were divided in morphology space (EDR and EDL signal components used in both *IA* and *IB* paths in Fig. 1), and feature space (EDA and HRV features in the *IB* path in Fig. 1; or latent representations from the deep learning models in the *IA* path in Fig 1).

Morphology Space	Feature Space
Pearson	Pearson
Spearman	Spearman
Cosine	Cosine
DTW	Euclidean Distance
Euclidean Distance	
Cross-Correlation	
Coherence	
Recurrence Plot	

we use as a benchmark model. Then, we expand the state of the art by using the learned classification model to obtain a higher-level data representation in which the interpersonal approach is applied. We apply a specific neural network model for each data space.

Signal Morphology

For 1D data (EDA, EDR, EDL – Morphology representation), we rely on the proposed approach from [61], which attained state-of-the-art results for the AMIGOS dataset. The architecture denoted as RTCAN-1D receives as input the three components of the EDA data: EDA, EDR and EDL in three channels. The architecture starts by performing a shallow feature extraction with a convolution layer and batch normalisation. Then, a combination of the three components is performed by an attention module – signal channel attention (SCA): with two convolution layers, followed by a sigmoid activation which is multiplied by the attention weight. Temporal similarities are analysed using a non-local attention mechanism relying on an embedded Gaussian kernel as similarity metric and a 1D convolution layer followed by average pooling with a kernel of 1 to conduct linear embedding – residual nonlocal temporal attention module. In a third step, an adapted ResNet-18 extracts higher-level features, replacing the 2D convolutions with 1D and simplifying the residual block to perform 1D convolution, batch normalisation and a ReLU activation. Lastly, the classification is performed following 3 fully connected layers, 3 ReLU functions, and a Softmax function. A pipeline of the architecture can be found in Fig. 1 of the Supplementary Material.

Image

For the 2D representation (Image – EDA Spectrogram and EDA Recurrence plot), we followed a similar strategy and applied a pre-trained ResNet-18 model [62] (Fig. 2 in the Supplementary material), with state-of-the-art results for the AMIGOS dataset in [58], [59]. The ResNet-18 is based on the addition of residual layers with an identity mapping to the input data, i.e. shortcuts that allow skipping layers. The usage of residual layers has been shown to improve the convergence in deep networks [62].

Hand-crafted Features

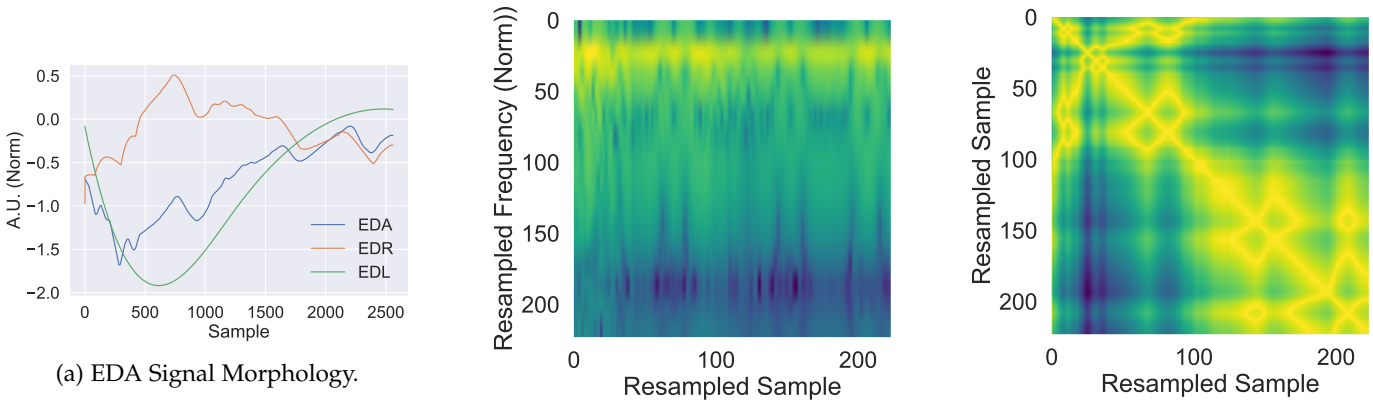
Lastly, for the hand-crafted features space (EDA and HRV features), we maintained the ResNet-18 overall architecture but changed the 2D convolutional layers to linear transformation layers. For all the models, the last layer was changed to set the class number to 1 to perform binary classification.

The models were evaluated using *Leave-One-Subject-Out* (LOSO), with one subject left for the test set while the remaining are used in the training set. One subject from the group training set was randomly selected and used as the validation set. The training, testing and validation configuration is the same for both intra- and interpersonal evaluation.

The models were developed in Python using the PyTorch library[6], and tuned through the Ray tune library[7] using a grid-search space. Table 3 shows the values of the hyperparameters used as search space. The Adam optimiser was used as the optimisation algorithm.

In addition to deep learning models, we used classic machine learning algorithms[8]: *Random Forest* (RF), SVM,

6. pytorch.org/; Accessed: 26/08/2022
7. docs.ray.io/en/latest/tune/index.html; Accessed: 26/08/2022
8. scikit-learn.org; Accessed: 26/08/2022



(a) EDA Signal Morphology.

(b) EDA Spectrogram.

(c) EDA Recurrence plot.

Fig. 2: Illustration of the data representations for the AMIGOS dataset on a 20 seconds sample. The image data in (b) and (c) were resampled to a 224x224 size to fit the input to the deep learning models.

TABLE 3: Hyperparamters space values. Dim – Dimension space in arousal (A), and valence (V). The parameters are shown per dataset for AMIGOS (A) and K-Emocon (K). Nomenclature: Grace Per. – Grace period. Learn. Rate – Learning rate; Weight Dec. – Weight decay.

	Dim.	Feature		Image		Morphology
		EDA	HRV	Spect.	RP	EDA-EDR-EDL
Batch size	A	16 (K, A), 128 (K, A)	16 (K, A), 128 (K, A)	16 (K), 128 (K, A)	16 (K), 128 (K, A)	16 (K), 128 (K), 256 (A)
	V					16 (K), 256 (K, A)
Epoch	A	800 (K, A)	800 (K, A)	300 (K), 800 (A)	800 (A), 1000 (K)	800 (K), 1000 (A)
	V	600 (K), 800 (A)	600 (K), 800 (A)	60 (K), 800 (A)	60 (K), 800 (A)	
Gamma	A	0.5 (A), 1.5 (K)	0.5 (A), 1.5 (K)	0.5 (K)	1.5 (K)	1.5 (K)
	V			1.5 (K)		
Grace Per.	A	0 (A), 10 (K), 50 (K)	10 (K), 50 (K), 60 (A)	0 (K), 10(A), 20 (K)	0 (K, A), 50 (K)	10 (K), 100 (K), 900 (A)
	V	0 (A), 10 (K), 100 (K)	0 (A), 10 (K), 100 (K)	0 (K), 10 (A), 50 (K)	0 (K), 10 (A), 40 (K)	10 (K), 200 (K), 900 (A)
Learn. Rate	A	1e-3 (K, A), 1e-5 (K, A)	1e-3 (K, A), 1e-5 (K, A)	1e-3 (K), 1e-5 (K), 1e-6 (A)	1e-5 (K, A), 1e-3 (K)	1e-5 (K, A), 1e-3 (K)
	V			1e-6 (K, A), 1e-3 (K)	1e-6 (K, A), 1e-3 (K)	
Patience	A	6 (A), 10 (K)	6 (A), 10 (K)	5 (A), 10 (K)	5 (A), 10 (K)	10 (K), 50 (A)
	V					
Weight Dec.	A			0.01 (K, A)		
	V			0.01 (K, A)		

and a *Naive Bayes* (NB), representing a non-linear, non-probabilistic and linear, and probabilistic model, respectively. The SVM, and RF hyperparameters were tunned using a 4-fold *Cross-Validation* (CV) grid-search.

4 RESULTS

The results are divided into two sub-sections. We start by presenting the results for the benchmark intrapersonal model (Section 4.1) where no group information is embedded into the model architecture. The obtained models are then used to get an additional higher-level space used as input data representation for the interpersonal method (Section 4.2).

4.1 Intrapersonal Model

We divide our results by group use case:

AMIGOS Dataset Table 4 shows the intrapersonal models results for the arousal and valence dimensions on the AMIGOS dataset. Due to the heavy data imbalance, we consider two metrics: The weighted F1-score (W-F1), which can be found in the emotion recognition literature, e.g. [55], weights the results by their count value to consider data imbalance. While the macro F1-score (M-F1) performs an unweighted mean of the label predictions. Through the rest of our work, we focus our analysis on the macro F1-score[9], while still leaving for observation the weighted F1-score on the tables, for a more realistic performance in the real world, where data imbalance is expected to be pervasive.

Overall in Table 4, the deep learning models attain similar results or outperform traditional machine learning models. For the arousal dimension, the best performance is obtained for a fully-connected ResNet network using HRV features (59.4%, M-F1). For the valence dimension, the best performance is obtained when combining EDA and HRV features in a fully-connected ResNet (66.44%, M-F1). The use of images (recurrence plot and spectrogram) or raw data (EDA, EDR, EDL – Morp.) did not result improve performance, attaining a F1-score close to random chance in both dimensions. Likewise, for the SVM on EDA data which attained the lowest performance overall (< 40%, M-F1).

9. scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html

TABLE 4: Results for the intrapersonal approach applied on the AMIGOS dataset. Nomenclature: Morp. – Signal morphology space; FV – Feature vector; Spect. – Spectrogram; RF – Random Forest; NB – Naive Bayes; SVM – Support Vector Machine; W-F1: Weighted F1-score; M-F1: Macro F1-score; RP – Recurrence Plot; NN – Neural Network.

Data	AMIGOS			
	Acc (%)	W-F1 (%)	M-F1 (%)	Training Time (s)
Arousal				
Morp. – NN	66.73 ± 07.84	66.68 ± 09.67	50.49 ± 05.21	11255.11 ± 191.50
EDA FV – NN	61.65 ± 06.93	64.16 ± 07.53	50.93 ± 05.84	301.58 ± 273.83
EDA FV – SVM	37.26 ± 10.85	36.70 ± 15.09	34.14 ± 09.42	5.43 ± 3.69
EDA FV – NB	72.75 ± 06.79	68.92 ± 10.68	50.21 ± 03.82	0.01 ± 0.00
EDA FV – RF	64.51 ± 10.65	65.26 ± 10.16	51.07 ± 07.73	3.58 ± 3.92
EDA Spect. – NN	70.18 ± 08.73	68.25 ± 10.84	50.66 ± 03.09	1203.70 ± 340.90
EDA RP – NN	70.95 ± 11.32	68.19 ± 11.87	49.37 ± 04.25	2076.95 ± 6608.11
HRV FV – NN	**68.51 ± 07.31**	**70.71 ± 07.81**	**59.40 ± 08.08**	**984.82 ± 623.01**
HRV FV – SVM	49.59 ± 11.87	52.45 ± 12.82	45.43 ± 11.99	6.13 ± 1.28
HRV FV – NB	69.18 ± 17.15	66.76 ± 17.49	51.21 ± 11.47	0.01 ± 0.00
HRV FV – RF	66.95 ± 09.19	69.20 ± 10.95	58.79 ± 10.78	2.06 ± 1.59
EDA + HRV FV – NN	61.65 ± 6.93	64.16 ± 07.53	50.93 ± 05.84	301.58 ± 273.83
EDA + HRV FV – RF	64.16 ± 11.06	66.40 ± 11.70	56.54 ± 11.39	5.16 ± 4.35
Valence				
Morp. – NN	55.33 ± 08.56	55.06 ± 08.30	49.65 ± 06.47	11197.13 ± 192.68
EDA FV – NN	55.77 ± 04.51	56.30 ± 04.78	51.78 ± 03.01	283.68 ± 227.95
EDA FV – SVM	41.37 ± 07.15	37.69 ± 09.72	38.59 ± 07.51	5.00 ± 0.98
EDA FV – NB	61.86 ± 06.59	56.95 ± 08.65	49.02 ± 04.12	0.01 ± 0.00
EDA FV – RF	52.46 ± 09.84	52.17 ± 10.19	49.13 ± 08.35	1.96 ± 2.69
EDA Spect. – NN	54.55 ± 05.14	55.13 ± 05.67	50.34 ± 02.99	1140.24 ± 409.72
EDA RP – NN	53.69 ± 05.38	54.28 ± 05.54	49.97 ± 03.77	1168.37 ± 436.49
HRV FV – NN	68.05 ± 09.89	67.98 ± 10.30	64.56 ± 10.57	352.97 ± 328.95
HRV FV - RF	69.04 ± 10.63	68.64 ± 12.10	65.63 ± 11.91	3.30 ± 1.93
HRV FV – NB	61.85 ± 11.63	59.89 ± 11.42	53.82 ± 08.54	0.02 ± 0.00
HRV FV – SVM	51.50 ± 10.05	51.40 ± 10.33	49.54 ± 10.07	5.25 ± 1.18
EDA + HRV FV – NN	**69.16 ± 09.99**	**69.36 ± 10.41**	**66.44 ± 10.05**	**430.14 ± 319.09**
EDA + HRV FV – RF	51.50 ± 10.05	51.40 ± 10.33	49.54 ± 10.07	5.18 ± 1.15

In addition to the macro F1-score, we display the accuracy (which is the most predominant score in the literature), although it does not take data imbalance into consideration, and weighted F1-macro (which shows the expected results for an imbalanced distribution). In both metrics, the best performing methodology is above 69% in both dimensions.

K-EmoCon Dataset Table 5 shows that for the K-EmoCon data, the classification performance is overall lower for the arousal dimension, being below random chance; For the arousal dimension, the best performance is obtained for the NB classifier combining EDA and HRV data. For the valence dimension, the best performance is obtained using the HRV features and a neural network model (73.7%, M-F1). The deep learning morphology-based and image-based (recurrence plot, spectrogram) methods either outperform traditional machine learning algorithms or attain similarly competitive results. Overall, the best method

shows an accuracy superior to 70% and 90% for the arousal and valence dimensions, respectively.

TABLE 5: Results for the intrapersonal approach applied on the K-EmoCon dataset.

Data	K-EmoCon			
	Acc (%)	W-F1 (%)	M-F1 (%)	Training Time (s)
Arousal				
Morp. – NN	59.39 ± 22.63	61.59 ± 22.80	45.04 ± 15.44	169.02 ± 134.94
EDA FV – NN	69.77 ± 14.76	69.10 ± 19.07	45.71 ± 05.24	56.60 ± 83.81
EDA FV – SVM	60.48 ± 19.49	63.49 ± 18.91	45.71 ± 10.00	0.23 ± 0.06
EDA FV – NB	74.48 ± 17.50	70.97 ± 22.12	46.28 ± 07.22	0.00 ± 0.00
EDA FV – RF	57.65 ± 18.99	61.27 ± 19.10	44.90 ± 10.84	0.13 ± 0.11
EDA Spect. – NN	69.62 ± 14.77	67.88 ± 20.69	43.87 ± 05.49	608.62 ± 770.74
EDA RP – NN	71.54 ± 15.63	68.93 ± 21.93	47.49 ± 13.53	1962.41 ± 3416.71
HRV FV – NN	68.81 ± 14.14	68.60 ± 19.58	46.30 ± 06.56	59.19 ± 70.02
HRV FV – SVM	56.20 ± 14.88	61.02 ± 15.58	45.27 ± 09.88	0.40 ± 0.09
HRV FV – NB	72.76 ± 15.10	71.22 ± 19.47	47.62 ± 07.15	0.00 ± 0.00
HRV FV - RF	59.64 ± 13.26	63.93 ± 15.50	46.37 ± 08.26	0.32 ± 0.14
EDA + HRV FV – NN	67.37 ± 14.99	68.13 ± 19.91	47.15 ± 06.59	54.93 ± 57.88
EDA + HRV FV – NB	**70.68 ± 14.05**	**70.63 ± 18.32**	**47.67 ± 05.74**	**0.00 ± 0.00**
Valence				
Morp. – NN	92.86 ± 09.38	90.35 ± 13.23	73.02 ± 27.04	1185.19 ± 875.01
EDA FV – NN	90.45 ± 10.63	88.95 ± 13.02	62.32 ± 24.80	394.79 ± 370.72
EDA FV – SVM	70.87 ± 18.13	76.83 ± 16.86	43.42 ± 07.47	0.22 ± 0.07
EDA FV – NB	84.76 ± 10.03	86.83 ± 11.77	47.97 ± 02.84	0.00 ± 0.00
EDA FV – RF	77.66 ± 15.90	81.79 ± 14.93	50.71 ± 16.88	0.16 ± 0.13
EDA Spect. – NN	93.32 ± 09.34	90.60 ± 13.35	73.14 ± 26.93	479.80 ± 97.12
EDA RP – NN	94.19 ± 07.10	92.96 ± 08.43	71.85 ± 26.40	350.58 ± 118.23
HRV FV – NN	**93.36 ± 09.44**	**90.66 ± 13.41**	**73.70 ± 26.52**	**238.49 ± 280.54**
HRV FV – SVM	83.14 ± 09.59	85.89 ± 11.40	48.80 ± 05.40	0.25 ± 0.08
HRV FV – NB	10.82 ± 10.76	10.66 ± 08.67	10.06 ± 09.79	0.00 ± 0.00
HRV FV – RF	87.56 ± 10.46	87.94 ± 12.13	55.28 ± 18.90	0.34 ± 0.21
EDA + HRV FV – NN	90.89 ± 09.22	89.36 ± 12.93	55.22 ± 18.98	71.32 ± 81.30
EDA + HRV FV – RF	88.16 ± 12.72	88.02 ± 13.61	60.59 ± 23.18	0.42± 0.33

4.2 Interpersonal Model

Table 6 summarises the main results for the interpersonal approach, namely WGS and average pooling. Tables 2 and 3 (for AMIGOS), and Tables 4 and 5 (for the K-EmoCon dataset) in the Appendix of the supplementary material, provide the detailed results obtained across all data representations and similarity metrics for the arousal and valence dimensions, respectively.

Classification Performance: Analysing Table 6, for the arousal dimension, the experimental results show that similarly to what was found with the intrapersonal method, the WGS applied on the HRV features obtained the best performance (72.15%, M-F1) using the Euclidean distance to measure physiological synchronisation. The use of Euclidean distance on EDL data was also able to maintain the M-F1 average above the 72% mark. For the valence, the best performance was obtained for the EDA features using once again the Euclidean distance (81.16%, M-F1), closely followed by HRV features on the learned representation using Cosine similarity (81.11%, M-F1).

For the K-EmoCon dataset, in the arousal dimension, the Cosine similarity on HRV features achieves an equal performance to the use of a non-weighted average pooling (52.63%, M-F1), with the accuracy (83.07%) surpassing random chance. For the valence dimension, the best F1-score is obtained for the EDL representation using Cross-Correlation to measure physiological synchrony (65.09%, M-F1).

Similarity Metric: Looking at each data representation (Morp. – Morphology (i.e. EDA, EDR, EDL) vs FV – feature vector vs LR – learned representation) in Tables 2 and 3 for AMIGOS and Tables 4 and 5 for K-EmoCon in Appendix. Across dimensions, we observe that for the signal morphology spaces (EDR and EDL) the DTW, Euclidean distance, cross-correlation, recurrence plot, and coherence obtain the best performance. With the exception of feature representations, the use of Pearson, Spearman and Cosine similarity often deteriorate the results.

For the AMIGOS dataset, across data representations and dimensions, often Cosine similarity shows the lowest STD on the similarity weights (Weight STD), approximating its results to average pooling. The results for the non-weighted group synchronisation (average pooling) show similar results with the use of synchronisation metrics, outperforming the aforementioned low-quality synchronisation metrics (i.e. Pearson, Spearman, Cosine similarity). Moreover, for the K-EmoCon dataset, average pooling outperforms the remaining. Additionally, we observe that regarding data representations, the EDA and HRV features obtain the most consistent results across synchronisation metrics. In terms of accuracy, the best-performing method attains average results > 80% (arousal and valence) for AMIGOS, and > 70% (arousal) and 90% (valence) for K-EmoCon.

Computational Complexity: Analysing the computation time, overall, a similar order of magnitude (< 0.1 seconds per sample) is obtained across similarity metrics, with the exception of the recurrence plot in the AMIGOS dataset, which shows high prediction times. The recurrence plot computation time is due to the time to compute the recurrence plot and obtain the quantitative analysis features.

4.3 Discussion

Our work focused on group emotion recognition based on unobtrusive physiological data, evaluating whether physiological synchrony can be identified in diverse group use cases such as dyadic conversations as per K-EmoCon dataset [41], and group video-watching interaction as per AMIGOS dataset [40]. Overall, our methodology relying on physiological synchrony holds across datasets attaining competitive results with the state-of-the-art intrapersonal methodology, with the exception of the K-Emocon dataset on the valence dimension. For the arousal dimension on the K-EmoCon dataset, weighted synchronisation on the best-performing metric (52.63%, M-F1) attains a similar result to performing average pooling (52.63%, M-F1). Possibly as a dyad is not enough to create a group atmosphere with emotion contagion; also group-watching is more prone to emotion contagion and physiological synchronisation than conversation. Another possibility is that synchronisation to just one user is more prone to noise than an average over multiple users. Comparing a group of 4 individuals (AMIGOS) to a dyad (K-EmoCon), the WGS performance decreases in the latter (from 72.15% arousal, 81.16% valence to 52.63% arousal, 65.09% valence, M-F1), still maintaining score values above random chance.

Synchronisation Metrics and Data Representations

Our RQ1 focused on analysing which synchronisation metrics and data representations can better measure physiological synchrony for emotion recognition. Our experimental results showed that for the AMIGOS dataset, the use of a learned representation on the HRV features attained the highest performance for arousal, and EDA and HRV features for valence. The results for the EDA data on the valence dimension are opposite to what we canonically expect,

TABLE 6: Best performing data representation and synchronisation metrics for the WGS methodology. The results are shown in terms of accuracy (Acc), weighted-F1 score (W-F1), macro-F1 score (M-F1), computation time per sample (Time), and weights standard deviation (Weight STD). The best results are shown in bold.

Dataset	Dimension	Data	Similarity Metric	Acc (%)	W-F1 (%)	M-F1 (%)	Time (s)	Weight STD
AMIGOS	Arousal	LR HRV FV	Euclidean Distance	**83.07 ± 04.92**	**82.87 ± 06.05**	**72.15 ± 10.11**	**0.007 ± 0.002**	**0.12 ± 0.05**
			Average Pooling	82.65 ± 05.49	82.26 ± 07.05	71.34 ± 10.78	0.010 ± 0.005	0.00 ± 0.00
		HRV FV – NN	Intrapersonal	68.51 ± 07.31	70.71 ± 07.81	59.40 ± 08.08	03.07 ± 01.95	
	Valence	EDA FV	Euclidean Distance	**82.80 ± 06.52**	**83.26 ± 06.09**	**81.16 ± 07.63**	**0.004 ± 0.006**	**00.06 ± 00.01**
			Average Pooling	82.70 ± 06.46	83.19 ± 06.01	81.11 ± 07.58	0.008 ± 0.001	00.00 ± 00.00
		EDA + HRV FV – NN	Intrapersonal	69.16 ± 09.99	69.36 ± 10.41	66.44 ± 10.05	01.34 ± 00.99	
K-EmoCon	Arousal	HRV FV	Cosine Similarity	**70.87 ± 22.50**	**70.80 ± 22.90**	**52.63 ± 21.28**	**0.000 ± 0.000**	
			Average Pooling	70.87 ± 22.50	70.80 ± 22.90	52.63 ± 21.28	0.000 ± 0.000	
		EDA + HRV FV – NB	Intrapersonal	70.68 ± 14.05	70.63 ± 18.32	47.67 ± 05.74	00.00 ± 00.00	
	Valence	EDL	Cross-Correlation (Max)	90.04 ± 08.91	90.16 ± 10.63	65.09 ± 22.55	0.001 ± 0.000	
			Average Pooling	90.04 ± 09.88	90.04 ± 10.94	64.90 ± 22.59	0.000 ± 0.000	
		HRV FV – NN	Intrapersonal	**93.36 ± 09.44**	**90.66 ± 13.41**	**73.70 ± 26.52**	**01.88 ± 02.21**	

since EDA is typically associated with arousal [17] and the SNS. The results are possibly due to a high arousal-valence annotation correlation (around 66%), or lower variability in the data (Weight STD of 0.06) which could bias for a higher synchronisation. For the K-EmoCon dataset, average pooling obtained a similar performance to WGS using Cosine similarity on HRV features for arousal, while for the valence space, Cross-Correlation on the EDL space was the best-performing metric. Overall, we observed a better performance on the feature space compared to morphology or image-based learned representations (from recurrence plots or spectrograms). Across datasets and dimensions, with the exception of feature representation, the use of Pearson, Spearman and Cosine similarity often deteriorate the method's performance. On the whole, the results for the non-weighted group synchronisation (average pooling) show similar results with the use of synchronisation metrics, outperforming the aforementioned lower performance synchronisation metrics (i.e. Pearson, Spearman, Cosine). For both datasets, the valence dimension attained a higher classification F1-score when compared to arousal, being in line with what is expected in the literature [17].

Interpersonal Model vs Intrapersonal Model

Our RQ2 extends the state of the art by analysing whether we can perform emotion recognition from a subject group members' emotion labels. Our experimental results show that the proposed interpersonal methodology outperforms or obtains competitive results comparatively with the state of the art: The authors in [63] report an accuracy of 55.22% (F1-score: 44.86%) for arousal; and 91.04% (F1-score: 87.62%) for valence. It should be noticed that the authors rely on additional multi-modal data such as accelerometer, ECG, and skin temperature. For the AMIGOS dataset, to the best of the authors' knowledge, this is the first paper to use only the group data (long videos) for a direct comparison of results. Our methodology surpasses the state of the art in the K-EmoCon on arousal (52.63%, M-F1 K-EmoCon) and provides novel results for AMIGOS. The literature often relies on intrapersonal models which do not take into account group information. We evaluated our proposed methodology against the baseline intrapersonal models and observed that interpersonal models outperform intrapersonal models for both datasets and dimensions, with exception of the valence dimension on the K-EmoCon dataset: On the AMIGOS dataset, the intrapersonal model on the arousal and valence dimensions evaluated on M-F1 attained 59.40% and 66.44% versus 72.15% and 81.16% for the interpersonal model, respectively; and on the K-EmoCon dataset the intrapersonal model on arousal and valence evaluated on M-F1 attained 47.67%, 73.70% versus 52.63%, 65.09% for the interpersonal model.

A limitation which is not controlled for in our work is the existence of pseudosynchrony i.e., the apparent synchronisation between signals even when those signals are not sharing information due to random coincidence. Pseudosynchronisation was identified in the works of [64] where nonverbal synchrony in psychotherapy is analysed. The authors in [64] control for change by creating pseudogroups with data from interactions that never happened. They do so through permutation of time segments, which they compare to genuine interactions. In WGS, similarly to the previous work, only data from genuine interactions is used for the prediction of the unknown subject label. In the work of [65], the authors examined several methods for surrogate data generation, where the synchronisation was removed (e.g. data shuffling, segment shuffling, data sliding, participant shuffling) to test for pseudosynchrony using window cross-correlation. Across these samples, synchronisation above 0 was observed, denoting the phenomena of pseudosynchrony using window cross-correlation as a similarity metric.

These works show that due to the phenomena of pseudosynchrony, physiological signals may show higher synchronisation than would be expected if the signals were totally disjoint, leading to a biased (lower accuracy) and less generalization accuracy of WGS than it could be obtained for truly synchronous groups. Thus, pseudosynchrony can explain some of the misclassification errors observed in our work. For the application of hypothesis tests, the creation of pseudogroups (fake synchrony) makes sense since it allows one to compare the pseudogroups to the sample under study, and identify if synchrony exists or not. In our work, on the other hand, we want to infer if WGS can be applied for emotion recognition classification, or if the phenomena of pseudosynchrony results in misclassification errors and makes the model prediction unusable due to an increased classification error.

5 Conclusion

The literature on group emotion recognition [2] reports that under certain conditions, the interaction between the group

members can lead to emotional contagion and physiological synchrony. This phenomenon can be an evolutionary step for emotion recognition systems, which tend to focus on the data of the subjects individually [18] (intrapersonal methods), missing context information provided by the group. In this article, we expand the state of the art by proposing and evaluating a novel method based on the group members' labels according to their physiological synchrony – WGS (interpersonal method). To do so, we perform an analysis of synchrony metrics and data representations to analyse which better capture the physiological synchrony interaction in a group setting for emotion recognition systems. Additionally, the method is evaluated under different group sizes (group of 4 versus dyad) and interaction use cases (video-watching versus conversation).

The experimental results show that the WGS integrating group information (interpersonal) is able to outperform the current state-of-the-art methods based on intrapersonal data (without group context) across datasets and dimensions, with the exception of the valence dimension on the dyads conversation dataset (K-EmoCon). Moreover, our method surpasses the previous works (i.e. [63]) for K-EmoCon on arousal, and provides novel comparable results for AMIGOS. Through the analysis of the WGS results on the two datasets, we conclude that group physiological synchrony contains useful context information for emotion recognition in both use cases, namely dyad conversations and group watching.

Future work may tackle the limitations of our proposed method, namely: 1) requires annotated labels at test time; 2) performs binary classification, and adaptation is required for fine-grained classification (i.e. multi-class); and 3) due to the datasets specifications, it relies on the external annotations by experts, which may not be related with the true underlying emotional experiences. Lastly, we focus on either dyads or groups of 4, although groups can vary widely in size. This limitation arises from the available datasets, but should be considered.

In summary, our study contributes to the field by: 1) introducing interpersonal physiological synchrony using physiological signals to emotion recognition; 2) performing an analysis of feature representation methods for group emotion recognition; 3) improving the accuracy of emotion recognition on group-activity tasks over prior work.

ACKNOWLEDGMENTS

This work was funded by FCT - Fundação para a Ciência e a Tecnologia under grant 2020.06675.BD by the FCT/MCTES through national funds and when applicable co-funded EU funds under the projects UIDB/50008/2020 and PCIF/SSO/0163/2019 "SafeFire", and by IT - Instituto de Telecomunicações.

REFERENCES

[1] E. Sundstrom and I. Altman, "Interpersonal relationships and personal space: Research review and theoretical model," *Human Ecology*, vol. 4, no. 1, pp. 47–67, 1976.

[2] A. Goldenberg, D. Garcia, E. Halperin, and J. J. Gross, "Collective emotions," *Current Directions in Psychological Science*, vol. 29, no. 2, pp. 154–160, 2020.

[3] C. von Scheve and S. Ismer, "Towards a theory of collective emotions," *Emotion Review*, vol. 5, no. 4, pp. 406–413, 2013.

[4] E. A. Veltmeijer, C. Gerritsen, and K. Hindriks, "Automatic emotion recognition for groups: A review," *IEEE Trans. on Affective Computing*, vol. 14, no. 1, pp. 89–107, 2023.

[5] A. Dhall, A. Kaur, R. Goecke, and T. Gedeon, "EmotiW 2018: Audio-video, student engagement and group-level affect prediction," pp. 653–656, 2018.

[6] A. Dhall, G. Sharma, R. Goecke, and T. Gedeon, "EmotiW 2020: Driver gaze, group emotion, student engagement and physiological signal based challenges," in *Proc. of the Int'l Conf. on Multimodal Interaction*, K. Truong, D. Heylen, M. Czerwinski, N. Berthouze, M. Chetouani, and M. Nakano, Eds. United States: Association for Computing Machinery, 2020, pp. 784–789.

[7] M. Iwasaki and Y. Noguchi, "Hiding true emotions: Micro-expressions in eyes retrospectively concealed by mouth movements," *Scientific Reports*, vol. 6, no. 1, p. 22049, 2016.

[8] C. Herrando and E. Constantinides, "Emotional contagion: A brief overview and future directions," *Frontiers in Psychology*, vol. 12, p. 712606, 2021.

[9] A. Strang, G. Funke, S. Russell, A. Dukes, and M. Middendorf, "Physio-behavioral coupling in a cooperative team task: Contributors and relations," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 1, p. 145, 2014.

[10] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: a survey of evaluation methods across disciplines," *IEEE Trans. on Affective Computing*, vol. 3, no. 3, pp. 349–365, 2012.

[11] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and Emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[12] G. Coppin and D. Sander, "Theoretical approaches to emotion and its measurement," in *Emotion Measurement*, H. L. Meiselman, Ed. Woodhead Publishing, 2016, pp. 3–30.

[13] A. Moors, "Theories of emotion causation: A review," *Cognition and Emotion*, vol. 23, no. 4, pp. 625–662, 2009.

[14] V. Misal, S. Akiri, S. Taherzadeh, H. McGowan, G. Williams, J. L. Jenkins, H. Mentis, and A. Kleinsmith, "Physiological synchrony, stress and communication of paramedic trainees during emergency response training," in *Companion Publication of the Int'l Conf. on Multimodal Interaction*. Association for Computing Machinery, 2020, p. 82–86.

[15] A. Karvonen, V.-L. Kykyri, J. Kaartinen, M. Penttonen, and J. Seikkula, "Sympathetic nervous system synchrony in couple therapy." *J Marital Fam Ther*, vol. 42, no. 3, pp. 383–395, 2016.

[16] R. Palumbo, M. Marraccini, L. Weyandt, O. Wilder-Smith, H. McGee, S. Liu, and M. Goodwin, "Interpersonal autonomic physiology: A systematic review of the literature," *Personality and Social Psychology Review*, vol. 21, no. 2, pp. 99–141, 2017.

[17] P. Bota, C. Wang, A. Fred, and H. Silva, "A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals," *IEEE Access*, vol. 7, pp. 140 990–141 020, 2019.

[18] W.-S. Chien, H.-C. Chou, and C.-C. Lee, "Self-assessed emotion classification from acoustic and physiological features within small-group conversation," *Companion Publication of the Int'l Conf. on Multimodal Interaction*, pp. 230–239, 2021.

[19] S. Mariooryad and C. Busso, "Exploring cross-modality affective reactions for audiovisual emotion recognition," *IEEE Trans. on Affective Computing*, vol. 4, no. 2, pp. 183–196, 2013.

[20] X. Guo, L. Polanía, and K. Barner, "Group-level emotion recognition using deep models on image scene, faces, and skeletons," in *Proc. of the ACM Int'l. Conf. on Multimodal Interaction*. Association for Computing Machinery, 2017, p. 603–608.

[21] B. Nagarajan and R. Oruganti, "Group emotion recognition in adverse face detection," in *IEEE Int'l Conf. on Automatic Face Gesture Recognition*, 2019, pp. 1–5.

[22] J. Quan, Y. Miyake, and T. Nozawa, "Incorporating interpersonal synchronization features for automatic emotion recognition from visual and audio data during communication," *Sensors*, vol. 21, no. 16, p. 5317, 2021.

[23] G. Chanel, S. Avry, G. Molinari, M. Bétrancourt, and T. Pun, "Multiple users' emotion recognition: Improving performance by joint modeling of affective reactions," in *Int'l Conf. on Affective Computing and Intelligent Interaction*, 2017, pp. 92–97.

[24] W.-S. Chien, H.-C. Chou, and C.-C. Lee, "Belongingness and satisfaction recognition from physiological synchrony with a group-modulated attentive BLSTM under small-group conversation,"

Companion Publication of the Int'l Conf. on Multimodal Interaction, pp. 220–229, 2021.

[25] G. Chanel, M. Kivikangas, and N. Ravaja, "Physiological compliance for social gaming analysis: Cooperative versus competitive play," *Interacting with Computers*, vol. 24, no. 4, pp. 306–316, 2012.

[26] S. Järvelä, J. M. Kivikangas, J. Kätsyri, and N. Ravaja, "Physiological linkage of dyadic gaming experience," *Simulation & Gaming*, vol. 45, no. 1, pp. 24–40, 2014.

[27] R. Reed, A. Randall, J. Post, and E. Butler, "Partner influence and in-phase versus anti-phase physiological linkage in romantic couples," *Int'l. Journal of Psychophysiology*, vol. 88, no. 3, pp. 309–316, 2013.

[28] R. Silver and R. Parente, "The psychological and physiological dynamics of a simple conversation," *Social Behavior and Personality: An Int'l Journal*, vol. 32, no. 5, pp. 413–418, 2004.

[29] D. Shearn, L. Spellman, B. Straley, J. Meirick, and K. Stryker, "Empathic blushing in friends and strangers," *Motivation and Emotion*, vol. 23, no. 4, pp. 307–316, 1999.

[30] I. Konvalinka, D. Xygalatas, J. Bulbulia, U. Schjødt, E.-M. Jegindø, S. Wallot, G. Orden, and A. Roepstorff, "Synchronized arousal between performers and related spectators in a fire-walking ritual," *Proc. of the National Academy of Sciences*, vol. 108, no. 20, pp. 8514–8519, 2011.

[31] P. Mitkidis, J. McGraw, A. Roepstorff, and S. Wallot, "Building trust: Heart rate synchrony and arousal during joint action increased by public goods game," *Physiology & Behavior*, vol. 149, pp. 101–106, 2015.

[32] A. Strang, G. Funke, S. Russell, A. Dukes, and M. Middendorf, "Physio-behavioral coupling in a cooperative team task: Contributors and relations," *Journal of experimental psychology. Human perception and performance*, vol. 40, 2013.

[33] E. Montague, J. Xu, and E. Chiou, "Shared experiences of technology and trust: An experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters," *IEEE Trans. on Human-Machine Systems*, vol. 44, no. 5, pp. 614–624, 2014.

[34] V. Müller and U. Lindenberger, "Cardiac and respiratory patterns synchronize between persons during choir singing," *PLOS ONE*, vol. 6, no. 9, pp. 1–15, 2011.

[35] E. Ferrer and J. L. Helm, "Dynamical systems modeling of physiological coregulation in dyadic interactions," *Int'l. Journal of Psychophysiology*, vol. 88, no. 3, pp. 296–308, 2013.

[36] A. Bachrach, Y. Fontbonne, C. Joufflineau, and J. L. Ulloa, "Audience entrainment during live contemporary dance performance: physiological and cognitive measures," *Frontiers in Human Neuroscience*, vol. 9, p. 179, 2015.

[37] E. Codrons, N. Bernardi, M. Vandoni, and L. Bernardi, "Spontaneous group synchronization of movements and respiratory rhythms," *PLOS ONE*, vol. 9, no. 9, pp. 1–10, 2014.

[38] L. Nummenmaa, J. M. Lahnakoski, and E. Glerean, "Sharing the social world via intersubject neural synchronisation," *Current Opinion in Psychology*, vol. 24, pp. 7–14, 2018, social Neuroscience.

[39] S. Järvelä, "Physiological synchrony and affective protosocial dynamics," Ph.D. dissertation, University of Helsinki, 2020.

[40] J. Miranda-Correa, M. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. on Affective Computing*, vol. 12, no. 2, pp. 479–493, 2021.

[41] C. Park, N. Cha, S. Kang, A. Kim, A. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee, "K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations," *Scientific Data*, vol. 7, no. 1, p. 293, 2020.

[42] E. Jun, D. McDuff, and M. Czerwinski, "Circadian rhythms and physiological synchrony: Evidence of the impact of diversity on small group creativity," *Proc. ACM Hum. Comput. Interact.*, vol. 3, pp. 1–22, 2019.

[43] Z. Li, M. Sturge-Apple, S. Liu, and P. Davies, "Parent-adolescent physiological synchrony: Moderating effects of adolescent emotional insecurity," *Psychophysiology*, vol. 57, no. 9, p. e13596, 2020.

[44] E. Prochazkova, E. Sjak-Shie, F. Behrens, D. Lindh, and M. Kret, "Physiological synchrony is associated with attraction in a blind date setting," *Nature Human Behaviour*, vol. 6, no. 2, pp. 269–278, 2022.

[45] S. Gashi, E. Di Lascio, and S. Santini, "Using students' physiological synchrony to quantify the classroom emotional climate," in *Proc. Int'l Joint Conf. and Int'l Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. Association for Computing Machinery, 2018, p. 698–701.

[46] I. Gordon, S. Wallot, and Y. Berson, "Group-level physiological synchrony and individual-level anxiety predict positive affective behaviors during a group decision-making task," *Psychophysiology*, vol. 58, no. 9, p. e13857, 2021.

[47] D. Richardson and R. Dale, "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension," *Cognitive Science*, vol. 29, no. 6, pp. 1045–1060, 2005.

[48] A. Knight, D. Kennedy, and S. McComb, "Using recurrence analysis to examine group dynamics." *Group dynamics: Theory, research, and practice*, vol. 20, no. 3, p. 223, 2016.

[49] A. Guidi, A. Lanata, P. Baragli, G. Valenza, and E. Scilingo, "A wearable system for the evaluation of the human-horse interaction: A preliminary study," *Electronics*, vol. 5, no. 4, 2016.

[50] P. Chikersal, M. Tomprou, Y. Kim, A. Woolley, and L. Dabbish, "Deep structures of collaboration: Physiological correlates of collective intelligence and group satisfaction," in *Proc. of the ACM Conf. on Computer Supported Cooperative Work and Social Computing*. Association for Computing Machinery, 2017, p. 873–888.

[51] O. Oullier, G. Guzman, K. Jantzen, J. Lagarde, and J. Kelso, "Social coordination dynamics: Measuring human bonding," *Social Neuroscience*, vol. 3, no. 2, pp. 178–192, 2008.

[52] E. Delaherche and M. Chetouani, "Multimodal coordination: Exploring relevant features and measures," in *Proc. of the Int'l Workshop on Social Signal Processing*. Association for Computing Machinery, 2010, p. 47–52.

[53] P. Hamilton, "Open source ECG analysis," in *Computers in Cardiology*, 2002, pp. 101–104.

[54] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions," *PLOS ONE*, vol. 8, no. 10, pp. 1–11, 2013.

[55] T. Zhang, A. E. Ali, C. Wang, A. Hanjalic, and P. Cesar, "Weakly-supervised learning for fine-grained emotion recognition using physiological signals," *IEEE Trans. on Affective Computing*, 2022.

[56] S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan, "Physiological signals based human emotion recognition: A review," *IEEE Int'l Colloquium on Signal Processing and its Applications*, vol. 1, pp. 410–415, 2011.

[57] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. on Biomedical Engineering*, vol. 63, no. 4, pp. 797–804, 2016.

[58] R. Elalamy, M. Fanourakis, and G. Chanel, "Multi-modal emotion recognition using recurrence plots and transfer learning on physiological signals," in *Proc. of the IEEE Int'l Conf. on Affective Computing and Intelligent Interaction*, 2021, pp. 1–7.

[59] Siddharth, T. Jung, and T. Sejnowski, "Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing," *IEEE Trans. on Affective Computing*, vol. 13, no. 1, pp. 96–107, 2022.

[60] J. Shukla, M. Barreda-Ángeles, J. Oliver, G. Nandi, and D. Puig, "Feature extraction and selection for emotion recognition from electrodermal activity," *IEEE Trans. on Affective Computing*, vol. 12, no. 4, pp. 857–869, 2021.

[61] G. Yin, S. Sun, D. Yu, D. Li, and K. Zhang, "A multimodal framework for large-scale emotion recognition by fusing music and electrodermal activity signals," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 18, no. 3, 2022.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on computer vision and pattern recognition*, 2016, pp. 770–778.

[63] P. Gupta, S. A. Balaji, S. Jain, and R. K. Yadav, "Emotion recognition during social interactions using peripheral physiological signals," in *Computer Networks and Inventive Communication Technologies*, S. Smys, R. Bestak, R. Palanisamy, and I. Kotuliak, Eds. Springer Singapore, 2022, pp. 99–112.

[64] F. Ramseyer and W. Tschacher, *Nonverbal Synchrony or Random Coincidence? How to Tell the Difference*. Springer Berlin Heidelberg, 2010, pp. 182–196.

[65] R. Moulder, S. Boker, F. Ramseyer, and W. Tschacher, "Determining synchrony between behavioral time series: An application of surrogate data generation for establishing falsifiable null-hypotheses," *Psychological Methods*, vol. 23, 2018.