Received: 30 September 2022 | Revised: 7 March 2023 | Accepted: 16 March 2023

DOI: 10.1111/psyp.14303

PSYCHOPHYSIOLOGY SPR WILEY

ORIGINAL ARTICLE

Prediction of moment-by-moment heart rate and skin conductance changes in the context of varying emotional arousal

Harisu Abdullahi Shehu[1] | **Matt Oxner**[2,3] | **Will N. Browne**[4] | **Hedwig Eisenbarth**[3]

[1]School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand

[2]Institute of Psychology, University of Leipzig, Leipzig, Germany

[3]School of Psychology, Victoria University of Wellington, Wellington, New Zealand

[4]School of Electrical Engineering and Robotics, Queensland University of Technology, Brisbane, Queensland, Australia

Correspondence
Harisu Abdullahi Shehu, School of Engineering and Computer Science, Victoria University of Wellington, Wellington 6012, New Zealand.
Email: harisushehu@ecs.vuw.ac.nz

[Correction added on April 22, 2023 after first online publication: The citations for supporting information have been included in the article.]

Abstract

Autonomic nervous system (ANS) responses such as heart rate (HR) and galvanic skin responses (GSR) have been linked with cerebral activity in the context of emotion. Although much work has focused on the summative effect of emotions on ANS responses, their interaction in a continuously changing context is less clear. Here, we used a multimodal data set of human affective states, which includes electroencephalogram (EEG) and peripheral physiological signals of participants' moment-by-moment reactions to emotional provoking video clips and modeled HR and GSR changes using machine learning techniques, specifically, long short-term memory (LSTM), decision tree (DT), and linear regression (LR). We found that LSTM achieved a significantly lower error rate compared with DT and LR due to its inherent ability to handle sequential data. Importantly, the prediction error was significantly reduced for DT and LR when used together with particle swarm optimization to select relevant/important features for these algorithms. Unlike summative analysis, and contrary to expectations, we found a significantly lower error rate when the prediction was made across different participants than within a participant. Moreover, the predictive selected features suggest that the patterns predictive of HR and GSR were substantially different across electrode sites and frequency bands. Overall, these results indicate that specific patterns of cerebral activity track autonomic body responses. Although individual cerebral differences are important, they might not be the only factors influencing the moment-by-moment changes in ANS responses.

KEYWORDS

autonomic nervous system, EEG, feature selection, machine learning, peripheral physiology, psychophysics

1 | INTRODUCTION

Autonomic nervous system (ANS) responses such as heart rate (HR) and galvanic skin response (GSR) are important tools for tracking emotional responses. They have been linked with the processing, as well as discussed as the basis of emotions (Damasio, 1996; Damasio et al., 1991). For instance, while an elevated heart rate might signal fear (Craske & Rachman, 1987; Roy et al., 2013), a higher GSR was found to be related to more aggression and hostility (Beauchaine et al., 2008; Gordis et al., 2010).

Changes in these measures have been linked with brain activity, that is, previous research has shown that a change in brain activity (e.g., in the Anterior Cingulate Cortex) contributes to changes in autonomic cardiac activity (Critchley et al., 2003). Similarly, brain activity was found to be correlated with GSR during gambling (Patterson et al., 2002) even though there is variation in timing courses of responding across ANS and brain activity captured via electroencephalogram (EEG).

A large amount of the research on the effect of emotion on ANS responses (or vice-versa) considered emotion as a summative experience. For instance, Loggia et al. (2011) accessed HR and GSR of healthy male participants' responses to heat pain. Although both measures increased with more intense pain stimulation, HR was found to be a better predictor of pain than GSR. Similarly, Mackersie and Calderon-Moultrie (2016) investigated the GSR and HR of participants during speech repetition tasks. A significant increase in GSR and a decrease in HR were observed with an increased speaking rate. However, whether these changes occurred due to listening effort or other forms of person-related factors such as stress or motivation remains unclear.

These studies, plus many other studies, investigated the overall effect of emotion on ANS responses. However, questions remain on whether these measures would reveal the same or a different pattern in moment-by-moment emotional responses. This would reflect that emotions are not overall, but rather a gradual and constantly changing experience (Brown, 2011; Radford, 2004).

Emotional states can vary in intensity and might not follow the emotion categories that have been underlying emotion research for the longest time (Keltner, 2019). Therefore, we need to improve linear outcome prediction in order to account for the non-categorical nature of emotional states (Azari et al., 2020).

As the GSR response time is usually found to be slower compared with other physiological measures (i.e., impact of latency) (Venables et al., 1980), a number of researchers investigated the impact of shifting the GSR signal on the prediction of autonomic response to emotional stimuli. For instance, Lemche et al. (2008) reasoned that due to a slow GSR, shifting the latency windows of the GSR with 1.2–3.3 s post-stimulus would ensure that electrodermal activity was not contaminated by nonspecific skin conductance response discharges prior to the autonomic response to the actual emotional stimuli. Contrary, Sierra et al. (2002) tested the effect of latency (i.e., time shift) in the prediction of autonomic response to emotional stimuli. They found that the latency of responses to specific stimuli was significantly prolonged. However, latency of reactions to nonspecific stimuli was significantly shorter.

Despite decades of research, individual differences in physiological responding can still not be accurately predicted (Pat et al., 2022). The intensity at which people experience an emotion may vary across different individuals based on their past experience or other behavioral characteristics (Madrigal, 2003). More so, while evidence suggests that emotional arousal increases as the difficulty of an experimental task increases (Andreassi, 2010), the intensity this difficulty is experienced might vary across different people, even for within the same context. Therefore, it is important to investigate whether a consideration of individual differences by individually modeling each participants' data improves the prediction of ANS activity.

The relationship between moment-by-moment ANS responses and brain activity has been investigated in the context of social threat (Eisenbarth et al., 2016). Specifically, they analyzed how stress-related autonomous changes are reflected in brain activation and deactivation pattern using fMRI. The study found that despite sharing some common representation in the brain during induced threat, the brain patterns that predicted ANS measures were largely distinct. They also found that predictions of HR and GSR within participants resulted in a lower prediction error than a prediction that included all participants' data in the training of the algorithm (i.e., the impact of individual difference). However, no other study has investigated the relationship between emotional arousal-related ANS changes over time and its association with the brain activity captured using EEG, which provides a higher time resolution.

This study aims to address this question by investigating moment-by-moment brain activity to predict the HR and GSR of participants.

Generally, there are two different ways to investigate the relationship between two variables (e.g., brain activity and the ANS) from a neuroscientific point of view. One way is correlational, whereas the second and most effective way is using a machine learning algorithm to model the relationship between brain activity and ANS. In contrast to correlation, which uses intrinsic properties of features via statistics, machine learning algorithms aim to find optimal performance and can generally achieve higher prediction and reliable results in comparison with

correlation analysis (Tetereva et al., 2022; Xue et al., 2012). More so, while a correlational analysis assumes a linear dependency between variables (Pat et al., 2022), machine learning allows different types of relationships (both linear and non-linear), which gives an open way of approaching this prediction problem. (Ballester & Mitchell, 2010; Svetnik et al., 2003). For that reason, this study will model the relationship between brain activity and the ANS using machine learning algorithms as it is particularly suited for this task.

To understand the relationship between brain and ANS activity, we need to consider how methods of prediction draw information from brain activity and ANS measures to make such predictions. Previous studies used different approaches, that is, the mass univariate approach to learn more about which brain regions as captured using fMRI contribute to predicting different physiological measures (Antonelli et al., 2019; Candia & Tsang, 2019; Helwig, 2019). While such technique can also be used for EEG, it is based on statistical inferences rather than a learning algorithm (Groppe et al., 2011), which would ensure an improvement or at least an equivalent result. However, the approach using a learning algorithm has not been fully explored. Therefore, investigating the way that these methods draw information from brain activity and ANS measures by using a feature selection algorithm will allow us to understand how such learning algorithms work in this context.

First, we will test whether prediction within or across participants is more accurate (i.e., will result in achieving a much lower error rate). Initially, we will perform the analysis by shifting the time scale of the GSR, taking into account different shifts, to test the physiological delay. Second, we will investigate whether within and across participants' training leads to better accuracy in predicting GSR and HR using the electroencephalogram (EEG), emotional arousal, and other physiological measures (i.e., blood volume and respiration). It was anticipated that not all electrode sites and peripheral measures have the same predictive power. For that reason, with this approach, we aim to unpack temporal changes with specific physiological measures as different patterns of brain activity likely mediate different physiological responses (Levenson, 2014; Norman et al., 2014). Therefore, as a final objective, we will investigate which electrode sites and frequency bands (features) play a role in predicting HR and GSR using a feature selection algorithm.

2 | METHOD

The study is a secondary data analysis of the Emotional Arousal Pattern (EMAP) data set (Eisenbarth et al., in press). The EMAP data set can be accessed online via: https://www.wgtn.ac.nz/emap/.

2.1 | Emotional arousal pattern data set

2.1.1 | Participants

The study included 145 participants (137 right-handed, 93 female, and 48 male). All participants were recruited at Victoria University of Wellington and have received two movie tickets for their participation. The study was approved by the ethics committee at Victoria University of Wellington.

2.1.2 | Data set description

The Emotional Arousal Pattern (EMAP) data set is a multimodal data set (of 3434 trials) of human affective states. The data set includes continuous arousal ratings, electrophysiological activity at the scalp using a 64-channel electroencephalogram (EEG) and peripheral physiological signals such as skin conductance, respiration, heart rate, and blood volume (plethysmography). The peripheral physiological measures were collected via a PowerLab 16/35 amplifier and sampled at 1000 HZ by LabChart recording software. These included respiration via a respiration belt around the ribcage; galvanic skin response (skin conductance level) through electrode plates placed on the index and ring finger of the non-dominant hand; and blood volume through an IR plethysmograph attached to the middle finger of the inferior hand. Heart rate (BPM) was calculated online in LabChart using default parameters for human ECG. The continuous arousal ratings for each trial were upsampled from 85 to 250 HZ via zero-order hold sampling, whereas the peripheral measures were downsampled by a factor of four from 1000 to 250 HZ. The trials with missing data were excluded from this analysis.

The data set is provided in two forms, that is, the raw data set with little modification and the clean data set with further processing of EEG channels to eliminate noise and artifacts. No further pre-processing was performed on the peripheral channels. This study used the clean data set and extracted features (see Section 2.2.1) to predict the ANS responses.

2.1.3 | Stimulus material

The stimulus material included 24 emotion-provoking video trials selected from https://www.alancowen.com/

film (Cowen & Keltner, 2017). The selected sound-less videos ranged from 13 to 19 s long ($M = 15$ s, $SD = 1$ s) with three videos in each of the following eight categories: low arousal positive, low arousal negative, low-mid arousal positive, low-mid arousal negative, mid-high arousal positive, mid-high arousal negative, high arousal positive, and high arousal negative.

2.2 | Modeling continuous peripheral physiological measures

2.2.1 | Feature extraction

To be able to model the data, we initially used Welch's method (Welch, 1967) to calculate the power spectra with a window size of 125. Each EEG electrode was then decomposed into four frequency bands (i.e., theta, alpha, beta, and gamma) using Simson's rule (Kepler, 2008), 4–8 Hz, 8–13 Hz, 13–0 Hz, and 30–60 Hz respectively. This resulted in a total of 256 EEG features plus four peripheral channels such as the galvanic skin response (GSR), respiration, heart rate (HR), and blood volume, that is, a total of 260 features and the moment-by-moment arousal ratings.

Figures 1 and 2 show GSR and HR of an exemplar participant from the EMAP data set.

2.2.2 | Machine learning

1. Long short-term memory: The data were modeled using a commonly used recurrent neural network (RNN) architecture called the long short-term memory (LSTM) model.
   LSTM is a type of RNN capable of learning dependence (for a long amount of time) in sequence prediction problems (Hochreiter & Schmidhuber, 1997). Standard RNNs tend to forget stored information in the presence of time lags, casting doubt on whether the network can exhibit significant practical advantages over time due to the vanishing gradient[1] problem (Hochreiter et al., 2001). In contrast to standard RNNs, the LSTM is not affected by the vanishing gradient problem, which makes them suitable for predicting time-series data that can have an unknown lag duration between events. For that reason, we therefore used an LSTM architecture to predict the HR and GSR from the EMAP data set as the EMAP data set presents moment-by-moment changes in the physiological measures, taking time into account.
   A detailed explanation of the LSTM architecture used and how the model was set up can be found in Supporting information.
2. Linear regression (LR): LR (Su et al., 2012) is a regression algorithm that can be used for modeling the relationship between a dependent and one or more independent variable (Weisberg, 2005). The algorithm finds a line that best fits the data points available on the plot, so it can be used to predict output (dependent variable) values for inputs that are not present in the data. LR was chosen to be used as it is simple and relatively fast.
3. Decision tree (DT): DT (Myles et al., 2004) is a non-parametric algorithm, which can be used for regression analysis. The algorithm has a hierarchical tree structure, which consists of a root node, branches, and leaf nodes (Song & Ying, 2015). DT was chosen to be used as it is easy to understand and faster in comparison to other tree-based algorithms such as random forests.

2.2.3 | Feature selection

Although the EMAP data set includes 260 features, it was anticipated that not all of the features play a role in predicting HR or GSR. Therefore, this work aims to determine the optimal number of features that contribute to predicting HR and GSR using the particle swarm optimization (PSO) algorithm. PSO was chosen due to its global search ability and quick convergence compared with other greedy search or evolutionary computation approaches (Coello et al., 2007; Shehu et al., 2021; Xue et al., 2014).

There are generally two versions of PSO: binary and continuous. This research utilized the continuous version as it has shown to achieve higher performance than the binary version (Hancer et al., 2018; Nguyen et al., 2021; Xue et al., 2012, 2015).

This study used PSO with LR and DT to investigate the optimal number of features selected due to their speed. LSTM was not utilized to perform feature selection as the algorithm is relatively slow.

A detailed explanation of PSO and the parameters used to design the algorithm can be found in Supporting information.

2.2.4 | Evaluation method

Equation 1 presents a commonly used evaluation metric for a regression model, that is, the root mean square error

[1]The vanishing gradient problem occurs when a network is unable to propagate useful gradient information from the input end of the model back to the output end of the model. As a result, a large change in the input will cause a small change in the output, which makes it hard to learn and tune the parameters of the earlier layers in the network.

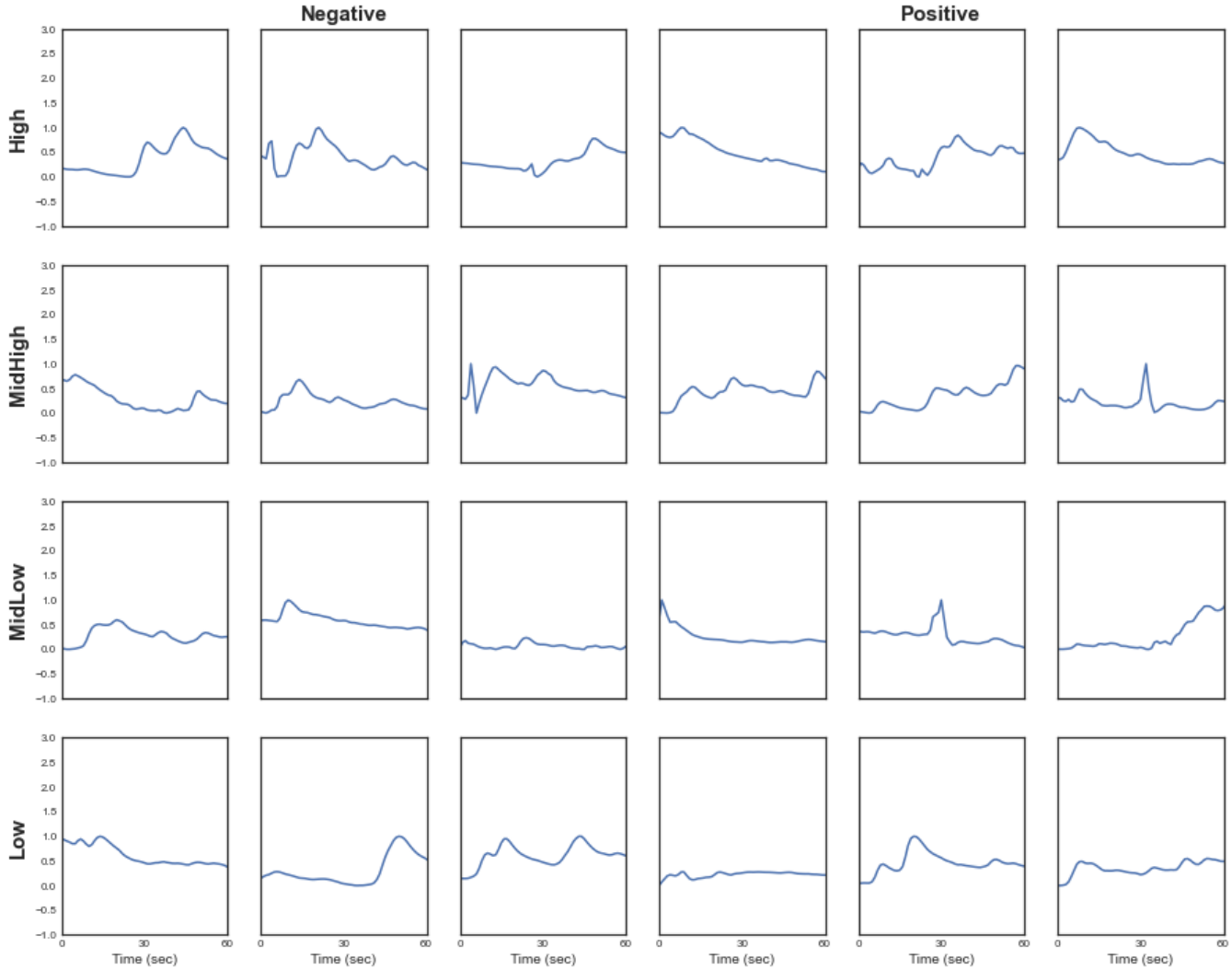

FIGURE 1 Change in GSR of an exemplar participant during each video. Videos are ordered according to a priori categories of valence (in columns; positive or negative) and arousal (in rows, low to high). The *y* axis represents ΔGSR (μS).

(RMSE). However, certain researchers prefer using a normalized RMSE over the standard RMSE to evaluate the performance of their regression model. For instance, when comparing model fit for different response variables or if the response variable is standardized, the RMSE does not perform well (Meese & Rogoff, 1983). A related study presenting the EMAP data set used data standardized by participants. For the sake of comparability, here, results are provided using the normalized RMSE (see Equation 2) as it is the metric used in the related work (Eisenbarth et al., in press).

$$\mathrm{RMSE} = \sqrt{\frac{\Sigma_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2}{N}} \tag{1}$$

where N represents the number of data points, y_i represents the ith measurement, and $\hat{y}_i$ is its corresponding prediction.

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\mathrm{Max}(y) - \mathrm{Min}(y)} \tag{2}$$

where $\mathrm{Max}(y)$ represents the maximum value of the test data point whereas $\mathrm{Min}(y)$ represents the minimum value of the test data point.

2.3 | Analysis

A regression analysis was conducted on the 260 extracted features from the EMAP data set. When predicting HR, the analysis was performed with 256 EEG frequencies, three peripheral physiology (i.e., GSR, blood pressure, and respiration), as well as the arousal ratings. Therefore, the analysis included HR as the dependent variable.

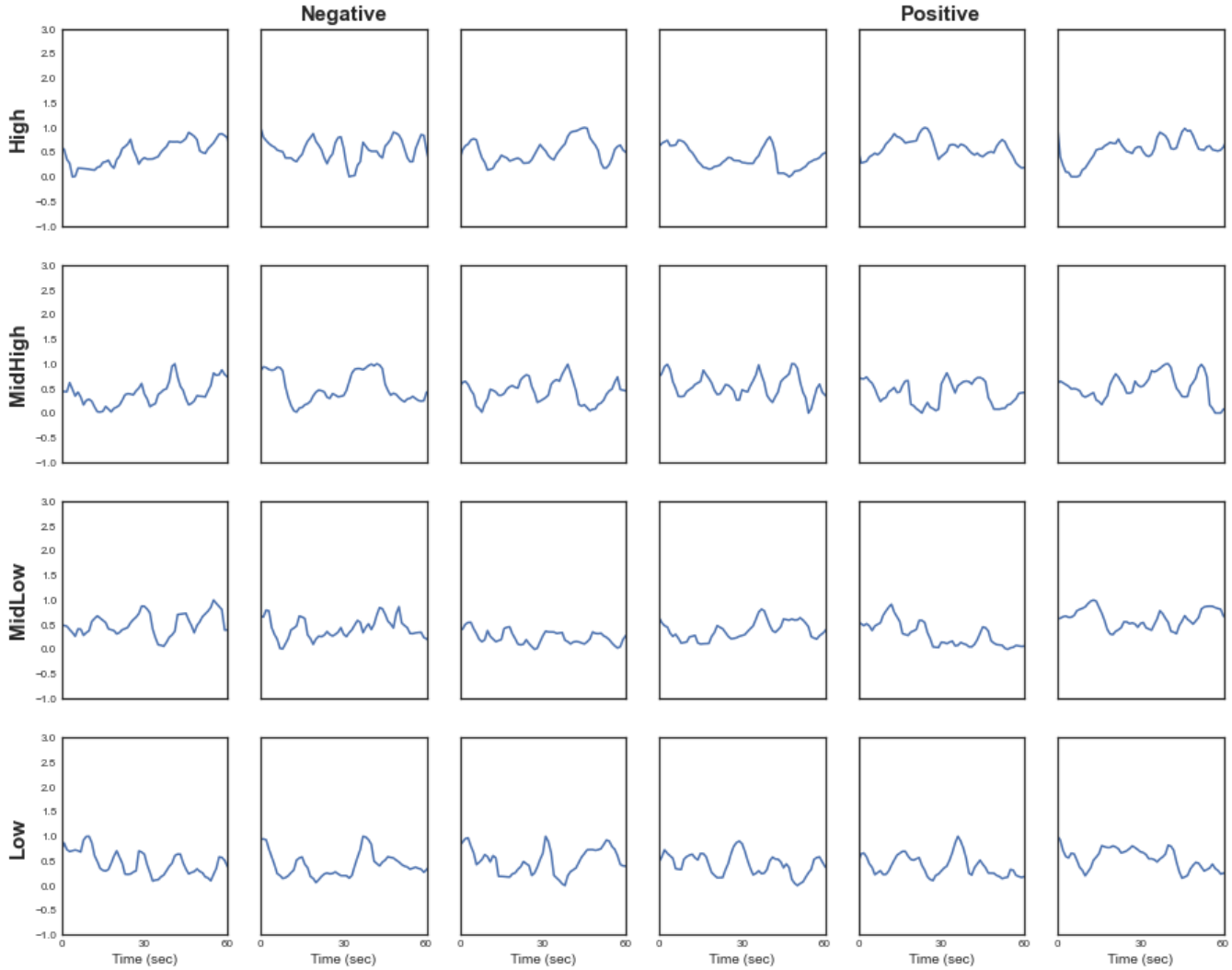

FIGURE 2 Change in HR of an exemplar participant during each video. Note that videos are ordered according to a priori categories of valence (in columns; positive or negative) and arousal (in rows, low to high). The *y* axis represents ΔBPM.

Similarly, when predicting GSR, the analysis was performed with 260 features, which include 256 EEG frequencies, three peripheral physiology (i.e., HR, blood pressure, and respiration), as well as the arousal ratings. Therefore, the analysis included GSR as the dependent variable.

Initially, we normalized the data by participants such that participants' variation was conserved (Chen et al., 2021). To test for a potential physiological delay, we predicted both the GSR and HR using temporal shifts. We shifted the GSR with 0, 5, 10, 15, 20, and 25 time bins, that is, 0, 2.5, 5, 7.5, 10, and 12.5 s, respectively, to analyze prediction error changes based on time shifts.

Moreover, as research has been confined to examining the motivational predictors in single behaviors rather than comparing effects across multiple behaviors (Brown et al., 2020), the present study addressed this gap by applying a data selection approach in the analysis to examine individual differences as a moderator of the effects of emotion on ANS measures through the following four analyses:

1. Group-by-group (GG): This analysis divides the EMAP data set into five different folds (i.e., 29 participants/fold). The training was performed with four and testing was performed on the remaining onefold. The procedure was repeated (five times) until all folds were tested.
2. Group-by-participant (GP): This analysis is the same as the leave-one-out cross-validation where the EMAP data set was divided into 145 different folds (i.e., one participant/fold). The training was performed with 144 and testing was performed on the remaining onefold. The procedure was repeated (145 times) until all folds were tested.
3. Participant-by-video (PV): Here, the analysis was performed based on a single participant data at a time. At any given time, one video from among the

videos watched by a participant was used as the test set whereas the remaining videos were used for training. The same procedure was performed until all videos watched by all participants were tested.

4. Stratified-by-video (SV): For this analysis, we performed a stratified[2] cross-validation across the videos watched by the participants such that each fold has one video from each of the following eight categories of videos watched (see also Section 2.1.3) by each participant: low arousal positive, low arousal negative, low-mid arousal positive, low-mid arousal negative, mid-high arousal positive, mid-high arousal negative, high arousal positive, and high arousal negative, which results in a total of three folds. The training was performed on twofold, and the test was performed on the remaining onefold. The procedure was repeated (three times) until all folds were tested. To ensure an adequate representation of every stratum (Taherdoost, 2016), in contrast to the other analysis performed above that utilized the complete data from the EMAP data set, this analysis only utilized the data of 139 participants that have no missing data as a result of eliminating noise and artifacts due to pre-processing.

Acknowledging that there is a difference in the number of data points evaluated for each of these analyses. However, these approaches are still on par with recent studies addressing similar questions (Compton et al., 2022; Pyke et al., 2021; Sun et al., 2022). In addition, conducting these experiments in different ways is important to understand the variability of these ANS responses across different participants/groups.

Finally, we performed feature selection and reduced the prediction error by creating a relevant subset of features. Moreover, we also produced the feature maps, which highlighted the relevant EEG features that play a role in predicting GSR and HR.

In order to avoid a feature selection bias and to provide generalized results across the whole data set, the shift and the feature selection analysis also used the GG analysis (the group-based fivefold cross-validation).

2.4 | Statistical analysis

Due to the small number of resulting NRMSEs in some of the analyses (e.g., five samples from GG), we performed a nonparametric test, that is, Mann–Whitney U-test to test the significance of the results obtained between the model of temporal shift, data selection approach, and the feature selection analysis.

The test within analysis compared the three different methods used (i.e., LSTM, DT, and LR), whereas the test between analysis compared the prediction error rate of the LSTM method across different analyses. For all multiple tests, the alpha level was adjusted using *Bonferroni corrections*.

[2]In stratified cross-validation, data are split into folds such that each fold has the same proportion of observations with a given categorical value.

3 | RESULTS

Initially, we test the NRMSE differences between the models of temporal shift using different shifts, followed by the data selection approach analysis. Finally, we performed feature selection to select the relevant features that contribute to predicting HR and GSR.

3.1 | Results from shift analysis

Table 1 presents the error obtained from predicting the GSR and HR of participants with 0, 10, 15, 20, and 25 temporal shifts. As can be seen, the lowest error was obtained from predicting GSR with no temporal shift (i.e., 0 time bins). Conversely, the lowest error for predicting HR was obtained when the GSR was shifted with 20 and 25 time-bins, that is, 10 and 12.5 s, respectively, in the best case.

However, a Mann–Whitney U-test with an adjusted alpha level of 0.0083 using *Bonferroni correction* showed no significant difference in the results obtained for both GSR and HR (all $p > .33$).

Moreover, we re-performed the analysis by randomly shuffling the continuous GSR and HR labels to investigate whether the results obtained by the model are different from chance. A Mann–Whitney U-test showed that the results obtained by the model for predicting both actual GSR and HR were significantly (p's $< .02$) better than the results obtained with the random label at alpha = .05. These findings suggest that the model is finding certain patterns in the data as the prediction of the model has a significantly much lower error rate compared with random guessing. Hence, it is safe to conclude that potential latencies in autonomic responses are covered by the LSTM algorithm.

3.2 | Results from data selection approach analysis

Table 2 presents the results obtained from predicting GSR and HR using the data selection approach (see Section 2.3).

TABLE 1 Analysis of temporal shift from predicting GSR and HR using LSTM.

Outcome	GSR						HR					
GSR shift (time bins)	0	5	10	15	20	25	0	5	10	15	20	25
M-NRMSE	0.1515	0.1592	0.1648	0.1636	0.1554	0.1582	0.1381	0.1354	0.1360	0.1357	0.1340	0.1340
Best-NRMSE	0.1194	0.1206	0.1192	0.1185	0.0992	0.1208	0.0961	0.0938	0.0852	0.0932	0.0961	0.0964
SD-NRMSE	0.0340	0.0368	0.0482	0.0479	0.0421	0.0358	0.0312	0.0308	0.0341	0.0315	0.0343	0.0325

Abbreviations: *M*, mean; *SD*, standard deviation.

In reporting the statistical test on the data in Table 2, the letters a, b, and c are used to indicate whether the results of the LSTM method are statistically significant compared with the other methods. The same letter infers no significant difference whereas different letters show that there is a significant difference.

As can be seen from Table 2, LSTM outperformed DT and LR in all cases when the analysis was performed for all the different scenarios. Moreover, a Mann–Whitney U-test showed a significant difference for most cases at $\alpha = .0167$, which means that LSTM has a better chance of predicting the data with a much lower error rate compared with DT and then followed by LR algorithms.

Figure 4 shows a visualization of a moment-by-moment prediction of GSR, whereas Figure 3 shows a visualization of a moment-by-moment prediction of HR obtained from the prediction of the 24 video trials watched by a randomly selected participant using the LSTM approach for the PV analysis. As can be seen, the predictions obtained by the LSTM model have shown to fit the data well by following the rising and falling trends of the ANS responses. A step-by-step prediction obtained from the initial to the final video watched by an exemplar participant can also be found in Supporting information.

Moreover, we compared the results obtained with LSTM from predicting GSR across the different data selection approaches. We found a significant difference in all results except for when GG was compared with SV ($p = .117$). Similarly, we found a significant difference in all the results obtained from predicting HR except for when GG was compared with GP ($p = .097$) and SV ($p = .068$) at $\alpha = .0167$, which means that while the LSTM outperformed other machine learning techniques used, its performance may also vary depending on the data selection approach.

The error rate obtained from predicting HR is lower than the error rate obtained from predicting GSR in all four different analyses. A comparison of the results obtained from predicting HR and GSR showed a significant difference ($p < .001$), which means HR can be predicted more accurately than GSR in the EMAP data set.

Comparing the results obtained from predicting GSR and HR by participants (PV) and by group (GG), we found that training the algorithm within participants resulted in a significantly (all $p < .001$) higher error than training across participants.

3.3 | Feature selection

In feature selection, the search space grows exponentially with the number of features (2^n). For that reason, the analysis in this part used only DT and LR as we found in

TABLE 2 Comparison of *data grouping* for predicting GSR and HR.

	LSTM		DT		LR		
Analysis type	**Mean**	***SD***	**Mean**	***SD***	**Mean**	***SD***	**Outcome**
GG	0.1420[a]	0.0482	0.2118[a]	0.0457	0.8964[a]	0.5055	GSR
GP	0.2240[a]	0.0240	0.4037[b]	0.1066	1.9719[c]	0.2234	
PV	0.2535[a]	0.0667	0.3669[b]	0.1124	0.5235[c]	0.1423	
SV	0.1020[a]	0.0017	0.1732[a]	0.0379	0.7953[a]	0.4843	
GG	0.1259[a]	0.0282	0.2012[b]	0.0582	0.8726[a]	0.3525	HR
GP	0.1418[a]	0.0078	0.2386[b]	0.0126	1.8545[c]	1.2552	
PV	0.1970[a]	0.0735	0.3292[b]	0.1260	1.0431[c]	1.2294	
SV	0.0932[a]	0.0153	0.1401[a]	0.0119	0.6122[a]	0.2056	

Note: Note that the best-performing model is denoted using the letter a and so forth. See Section 2.3 for a detailed description of the analysis.
Abbreviations: GG, group-by-group; GP, group-by-participant; PV, participant-by-video; SV, stratified-by-video.

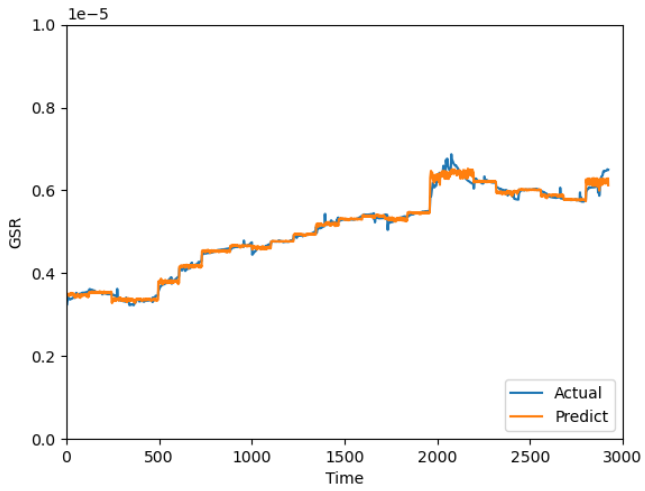

FIGURE 3 Moment-by-moment prediction of skin conductance response (in μS) obtained using the LSTM model (with PV analysis) for the 24 video trials watched by an exemplary participant.

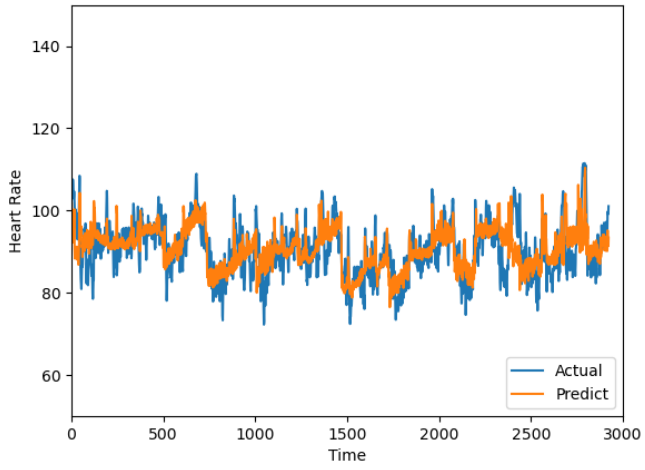

FIGURE 4 Moment-by-moment prediction of heart rate (in Δ BMP) obtained using the LSTM model (with PV analysis) for the 24 video trials watched by an exemplary participant.

empirical findings that these algorithms are (at least 30 times) faster compared with the LSTM.

Although the analysis in this part also used the GG selection approach, all analyses were run independently and participants were randomly selected and placed into a group. Therefore, it is important to note that the results in Table 2 are not related to that in Table 3.

Table 3 presents the results obtained before and after feature selection was performed to predict GSR and HR from the EMAP data set. As can be seen, feature selection not only helped reduce the number of features from the EMAP data set, but also reduced the prediction error for both algorithms (DT and LR).

A Mann–Whitney U-test with an adjusted alpha level of 0.025 using *Bonferroni correction* showed a significant difference between the LR compared with the CPSO-LR-based results $U = 2.0, n_1 = n_2 = 5, p = .018$ (HR)) when predicting HR. However, there was no significant difference between LR compared with the CPSO-LR-based results ($U = 3.0, n_1 = n_2 = 5, p = .03$ (GSR)) when predicting GSR and DT compared with the CPSO-DT-based results $U = 11.5, n_1 = n_2 = 5, p = .458$ (GSR), $U = 10.0, n_1 = n_2 = 5, p = .338$ (HR)) when predicting both the GSR and HR independently.

We also compared the results obtained by DT to the DT-CPSO-based results and LR to the LR-CPSO-based results (i.e., before and after feature selection) in predicting both GSR and HR using Mann–Whitney U-tests. We found a significant difference between the results $(U = 17, n_1 = n_2 = 10, p = .007)$-(DT vs. DT-CPSO), $(U = 18, n_1 = n_2 = 10, p = .009)$-(LR vs. LR-CPSO) at $\alpha = .025$. This suggests that feature selection can considerably reduce the number of features while significantly reducing or maintaining the error rate.

The feature maps based on the features selected by the feature selection approach are depicted in Figures 5 and 6. As each run might have a different number of features

Method	LR	CPSO-LR	DT	CPSO-DT	Outcome
M-NRMSE	0.8964	0.3509	0.1453	0.1429	GSR
Best-NRMSE	0.5699	0.1279	0.1015	0.1006	
SD-NRMSE	0.5126	0.4386	0.0358	0.0355	
Ave-#FS	260	118	260	151	
M-NRMSE	0.8492	0.2366	0.1255	0.1208	HR
Best-NRMSE	0.4449	0.1083	0.0930	0.0867	
STD-NRMSE	0.3520	0.1869	0.0230	0.0243	
Ave-#FS	260	124	260	142	

Abbreviations: Ave-#FS, average number of features selected; *M*, mean; *SD*, standard deviation.

TABLE 3 Feature selection analysis for predicting GSR and HR.

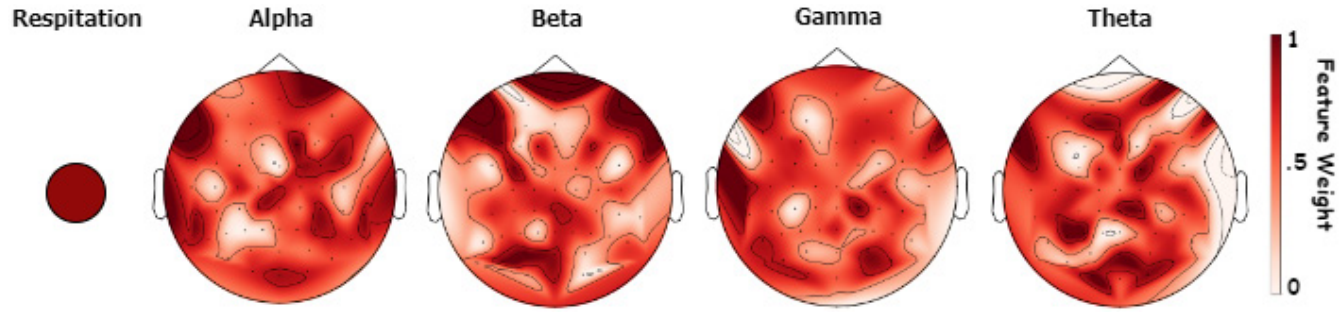

FIGURE 5 Maps of selected features obtained for predicting GSR. Note that all regions are highlighted based on their respective feature weights.

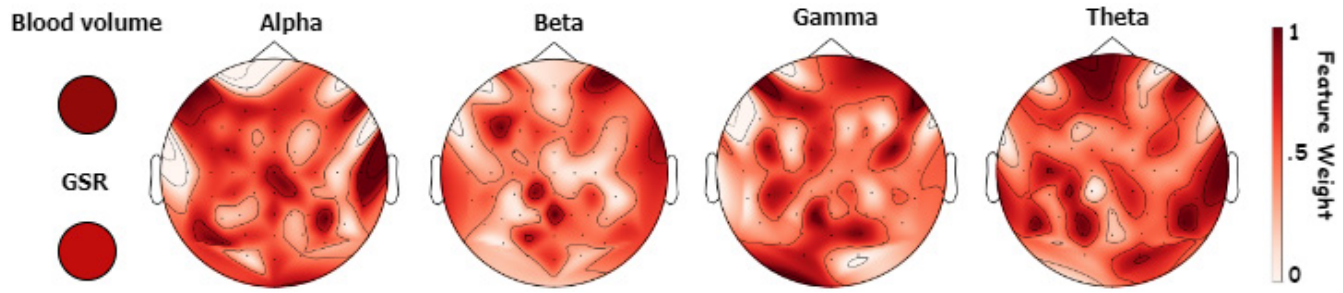

FIGURE 6 Maps of selected features obtained for predicting HR. Note that all regions are highlighted based on their respective feature weights.

selected, 154 and 118 selected features from the best run using *CPSO–DT* obtained from predicting GSR and HR respectively were used to plot the topographical plots.

Out of the 154 features selected for predicting the GSR, one of the features included is respiration. Similarly, of the 118 features selected for predicting the HR, two of the features included are the GSR and the blood volume. For that reason, the diagram representing the selected features for the GSR included not only the topographical plots but also the selected peripheral measure, that is, respiration whereas the diagram representing the selected features for the HR also included the GSR and the blood volume.

Figure 5 shows a representation of the features selected to predict the GSR, whereas Figure 6 shows a representation of the features selected to predict HR. As can be seen, the features selected by the same method (with the lowest prediction error, i.e., CPSO–DT) to predict GSR and HR are largely distinct. Although frontal *Beta*, temporal *Alpha*, and *Gamma* were mainly selected to predict the GSR, the majority of parietal *Alpha* and *Theta* were selected to predict the HR, while *Beta* had a lower weight. *Theta* and *Gamma* frequency have an almost equal distribution of features selected across the frontal, temporal, and parietal lobe for both GSR and HR.

Moreover, we plotted the most frequently selected features, that is, the features selected in all of the five independent runs (in the GG analysis). Figure 7 shows a representation of the most frequently selected features for predicting GSR whereas Figure 8 shows a representation of the most frequently selected features for predicting HR.

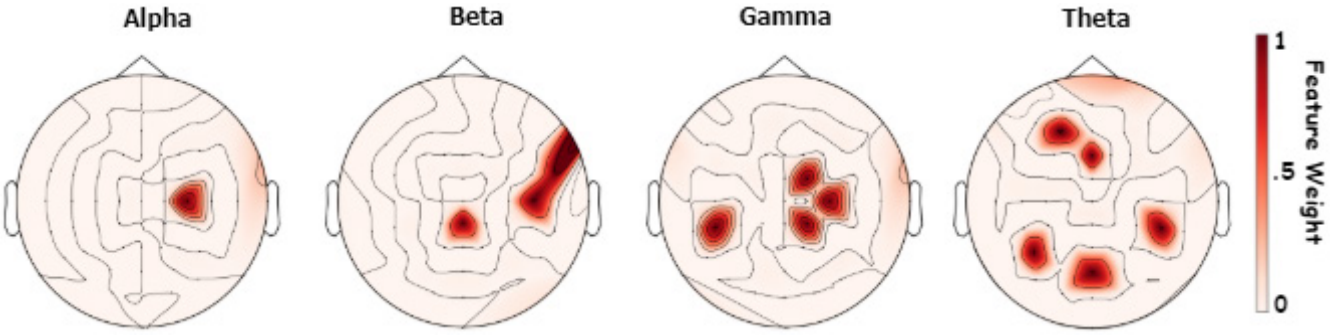

FIGURE 7 Maps of most frequently selected features for predicting GSR. Note that the highlighted regions are only the regions selected in all five independent runs.

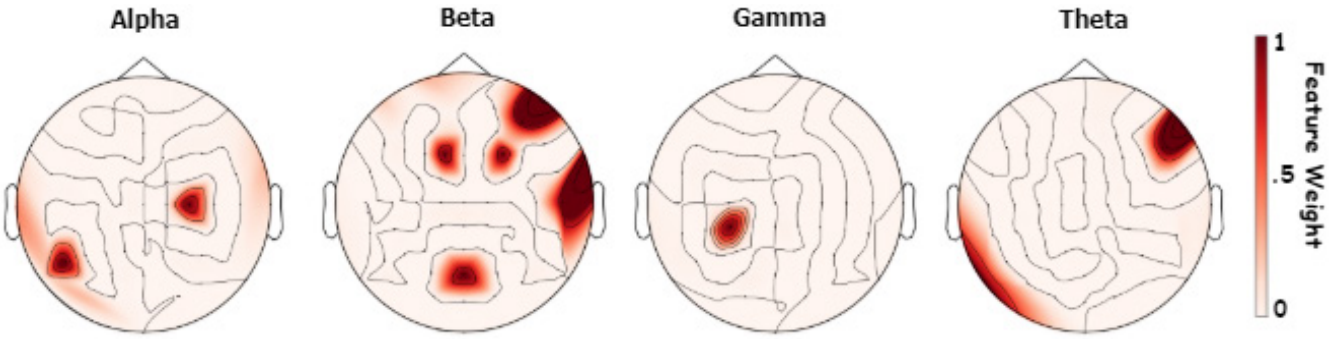

FIGURE 8 Maps of most frequently selected features for predicting HR. Note that the highlighted regions are only the regions selected in all five independent runs.

As can be seen, parietal *Beta*, *Gamma*, and *Theta* were most relevant for GSR prediction, but *Alpha* spectra seem to have the least weight. Conversely, *Gamma* seems to have the least weight in predicting HR compared with *Alpha*, *Beta*, and *Theta* frequency bands.

4 | DISCUSSION

This study investigated the moment-by-moment brain activity to predict emotional arousal-related ANS responses of participants as they watched emotion-provoking videos. Initially, we analyzed the physiological delay of the data by performing an analysis using temporal shifts before performing a data selection approach analysis to examine the effect of individual differences on the ANS responses. Lastly, we performed feature selection to determine the minimum number of features, as well as investigated the brain patterns that play a role in predicting these measures.

Considering that GSR is observed with some latency after an emotional cue (Horn et al., 2020), it was unexpected that various temporal shifts of the GSR data did not have a significant effect on the prediction results. This suggests that potential delays might not deteriorate the prediction by the LSTM model, which includes temporal mechanism (see Section 3.1). Although unexpected, this finding is consistent with the analysis predicting arousal ratings with the combined EEG and ANS EMAP data set, which also found that shifting the GSR data did not have a significant effect on the predictive relevance (Eisenbarth et al., in press). However, it is worth noting that the finding is only valid for the type of analysis and algorithms used in this study as different algorithms might lead to a different result. Especially those that unlike LSTM do not include a temporal mechanism.

The findings that the error rate obtained from predicting HR is significantly lower than the error rate for predicting GSR for all the algorithms suggest that GSR is much harder to predict in the EMAP data set than HR. As the stimuli set shown to the participants consists of both positive and negative stimuli (see Section 2.1.3), it is anticipated that HR was mainly sensitive to changes due to negative stimuli (mainly in one direction) compared with GSR, which is anticipated to change due to both positive and negative stimuli (both directions). As a result, a possible explanation of why HR was much easier to predict than GSR could be because HR was more stable than GSR, which reflected a bi-directional change.

Due to individual differences between participants, we would have expected to see a much lower error rate when the prediction is made by participants, that is, train and test with a single participant's data (e.g., PV) than when it is made by groups, that is, train and test with multiple participants data (e.g., GG). However, the error rate obtained from predicting both the GSR and HR by participant (PV) was significantly higher than the error rate obtained when the prediction was made by groups (GG) (all $p < .001$); it seems that individual differences ANS reactions seem to have such a high impact. The finding is inconsistent with those of previous studies, which found that the prediction error from analysis by participant is lower than the error obtained when the prediction was made across different participants (Andreassi, 2010; Fitch et al., 2020).

It is evident from Figures 3 and 4 that the LSTM model reacts to smooth changes and follows the upward and downward trends in predicting these autonomic responses. These findings serve as evidence, as well as providing a better understanding of what a prediction with a lower error rate looks like. However, it is important to keep in mind that the finding that the LSTM model has a higher chance of predicting the ANS responses (in comparison with other techniques) with a lower error rate is only relevant for the three algorithms (i.e., LSTM, DT, and LR) used in this research as different machine learning algorithms might lead to obtaining a different performance result.

Although feature selection did not significantly decrease the error rate for DT when predicting both HR and GSR or for LR when predicting the GSR, the results suggest that feature selection can obtain a comparable performance using only a small subset of relevant features.

As can be seen in Figures 5 and 6, there is a clear difference between the pattern for predicting GSR and HR. While only respiration, out of the three peripheral measures (i.e., blood volume, heart rate, and respiration) used to predict the GSR was selected to be an important predictor, blood volume and GSR were the selected peripheral measures, out of the three (i.e., blood volume, GSR, and respiration) used for predicting the HR. Based on these results, it is, therefore, reasonable to conclude that respiration is a strong predictor of GSR whereas blood volume and GSR were strong predictors of HR. The finding that the blood volume and GSR correlated with HR is consistent with previous studies that also found a correlation between blood volume, GSR, and HR (Filipovskỳ et al., 1992; Lazarus et al., 1963; Lin et al., 2011; Loggia et al., 2011).

Moreover, Figures 7 and 8 suggest that of the selected EEG signals, while the Alpha frequency band contributed least to predicting GSR, Gamma frequency had the lowest weight in predicting HR. This combination of findings is in line with previous findings pointing toward different brain areas to be related to HR and GSR respectively (Eisenbarth et al., 2016), which is reflected here in the different electrode sites and frequency bands.

It is also worth noting that the EEG features included here were only the frequency spectra of the brain activity and do not allow inference about specific brain regions involved, despite providing the relevant information about timely processing of such affective stimuli (Schubring & Schupp, 2021). To uncover factors that might contribute to our findings, combining EEG with fMRI could provide a better spatial resolution and thereby elucidate the differences between participant and group-based analyses.

One limitation of the feature selection method is that it could only be utilized for DT and LR due to their processing time requirements. The LSTM algorithm was not used for the analysis due to its sequential computation in the LSTM layer, which makes the algorithm very slow. The algorithm takes an average of 33 minutes to complete a single epoch, and there is a tendency for the algorithm to reach up to 100 epochs if results continuously improve (see Supplementary Material, Section 4). In addition, the features selected by either DT or LR might not lead to achieving high accuracy with the LSTM model as the wrapper-based feature selection method selects the features based on the specific algorithm used within the wrapper method. As a result, we suggest that the LSTM algorithm is not suited practically for the feature selection task.

To check whether the relationship between the variables in the data is linear, we used linear regression to perform the analysis. However, the results suggest that there is no linear relationship between the variables as linear regression achieved the worst result among the three algorithms used.

In general, the findings in this study provide consistent evidence that the LSTM method, which learns dependence in sequential problems achieved higher performance results compared with traditional machine learning algorithms in predicting ANS responses with little evidence of moderation by individual differences. The findings have implications for generalizability that the LSTM method improved the performance of relevant regression methods. Moreover, the produced feature maps using the feature selection technique increased our understanding of the relevant brain parts and peripheral measures that play a role in predicting these ANS responses.

However, the study has some limitations. First, the data set was collected by presenting videos of different categories (i.e., low, mid, and high arousal, positive and negative valence) to participants. It is noted that the videos watched might not have induced the intended or even any emotion in the participants, which raises concern that the research may be limited by its reliance on a small number of potentially idiosyncratic stimuli. Future studies should include a variety of videos to choose from or rather ask the

participants to recall an event that they believe will induce the intended emotion.

Second, this study investigates the ANS by considering the intensity of emotion (i.e., emotional arousal) across experiences that vary in emotional valence. Future work could also investigate, if the prediction models vary depending on the emotional valence that was underlying the physiological response, which might be informative to understand whether and which variation would be driven by emotional valence.

Third, the current finding refers to the relationship between EEG signals and peripheral measures of the ANS specifically HR and GSR. We focused on testing models that include all other physiological features as predictors, which may influence the results obtained to investigate the relationship between the ANS and brain activity. Similarly, including the EEG features may have an influence on the prediction of GSR and HR by the remaining ANS features. Future studies should therefore investigate the relationship between EEG-based brain activity and peripheral measures separately using sole modalities of the EEG and peripheral measures, respectively.

5 | CONCLUSION

Together, the results obtained from this study suggest that the LSTM algorithm covered the potential latencies in the autonomic responses, while providing little evidence for the moderation of the prediction error by individual differences. The results also suggest that specific brain regions and peripheral measures support the differential processing of HR and GSR. This is evidenced by the prediction error was significantly reduced using only a small number of feature subsets than using all features from the EMAP data set. We suggest that these regions provide a substrate for translating video stimuli into differential cardiac and galvanic responses, and thereby represent an interface for the selective generation of visceral reactions that contribute to the embodied component of emotional arousal.

AUTHOR CONTRIBUTIONS

Harisu Abdullahi Shehu: Conceptualization; data curation; investigation; methodology; project administration; validation; visualization; writing – original draft. **Matt Oxner:** Supervision; writing – review and editing. **Will N. Browne:** Resources; supervision; writing – review and editing. **Hedwig Eisenbarth:** Resources; supervision; writing – review and editing.

ACKNOWLEDGMENTS

The authors would like to acknowledge Amy Walsh and Tim Gastrell for their help in gathering the data. Open access publishing facilitated by Victoria University of Wellington, as part of the Wiley - Victoria University of Wellington agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

This research was not supported by any organization.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Data used in this study can be downloaded from: https://www.wgtn.ac.nz/emap/.

CODE AVAILABILITY

Code implemented in this study can be downloaded from OSF (https://osf.io/cyf86/) or accessed via Code Ocean (https://doi.org/10.24433/CO.7379558.v1).

ORCID

Harisu Abdullahi Shehu https://orcid.org/0000-0002-9689-3290
Hedwig Eisenbarth https://orcid.org/0000-0002-0521-2630

REFERENCES

Andreassi, J. L. (2010). *Psychophysiology: Human behavior and physiological response*. Psychology Press. https://doi.org/10.4324/9780203880340

Antonelli, L., Guarracino, M. R., Maddalena, L., & Sangiovanni, M. (2019). Integrating imaging and omics data: A review. *Biomedical Signal Processing and Control*, *52*, 264–280. https://doi.org/10.1016/j.bspc.2019.04.032

Azari, B., Westlin, C., Satpute, A. B., Hutchinson, J. B., Kragel, P. A., Hoemann, K., Khan, Z., Wormwood, J. B., Quigley, K. S., Erdogmus, D., Dy, J., Brooks, D. H., & Barrett, L. F. (2020). Comparing supervised and unsupervised approaches to emotion categorization in the human brain, body, and subjective experience. *Scientific Reports*, *10*(1), 20284. https://doi.org/10.1038/s41598-020-77117-8

Ballester, P. J., & Mitchell, J. B. O. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, *26*(9), 1169–1175. https://doi.org/10.1093/bioinformatics/btq112

Beauchaine, T. P., Hong, J., & Marsh, P. (2008). Sex differences in autonomic correlates of conduct problems and aggression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(7), 788–796. https://doi.org/10.1097/CHI.Ob013e318172ef4b

Brown, C. L., Van Doren, N., Ford, B. Q., Mauss, I. B., Sze, J. W., & Levenson, R. W. (2020). Coherence between subjective experience and physiology in emotion: Individual differences and implications for well-being. *Emotion*, *20*(5), 818–829. https://doi.org/10.1037/emo0000579

Brown, H. (2011). The role of emotion in decision-making. *The Journal of Adult Protection*, *13*(4), 194–202. https://doi.org/10.1108/14668201111177932

Candia, J., & Tsang, J. S. (2019). Enetxplorer: An R package for the quantitative exploration of elastic net families for generalized linear models. *BMC Bioinformatics*, *20*, 1–11. https://doi.org/10.1186/s12859-019-2778-5

Chen, H., Sun, S., Li, J., Ruilan, Y., Li, N., Li, X., & Bin, H. (2021). Personal-Zscore: Eliminating individual difference for eeg-based cross-subject emotion recognition. *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2021.3137857

Coello, C. A., Coello, G. B., Lamont, D. A., & Veldhuizen, V. (2007). *Evolutionary algorithms for solving multi-objective problems* (Vol. 5). Springer. https://doi.org/10.1007/978-0-387-36797-2

Compton, R. J., Jaskir, M., & Jianing, M. (2022). Effects of post-response arousal on cognitive control: Adaptive or maladaptive? *Psychophysiology*, *59*(4), e13988. https://doi.org/10.1111/psyp.13988

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38), E7900–E7909. https://doi.org/10.1073/pnas.1702247114

Craske, M. G., & Rachman, S. J. (1987). Return of fear: Perceived skill and heart-rate responsivity. *British Journal of Clinical Psychology*, *26*(3), 187–199. https://doi.org/10.1111/j.2044-8260.1987.tb01346.x

Critchley, H. D., Mathias, C. J., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B.-K., Cipolotti, L., Shallice, T., & Dolan, R. J. (2003). Human cingulate cortex and autonomic control: Converging neuroimaging and clinical evidence. *Brain*, *126*(10), 2139–2152. https://doi.org/10.1093/brain/awg216

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

Damasio, A. R., Tranel, D., & Damasio, H. C. (1991). Behavior: Theory and preliminary testing. In *Frontal lobe function and dysfunction* (pp. 217–229). Oxford University Press.

Eisenbarth, H., Chang, L. J., & Wager, T. D. (2016). Multivariate brain prediction of heart rate and skin conductance responses to social threat. *Journal of Neuroscience*, *36*(47), 11987–11998. https://doi.org/10.1523/JNEUROSCI.3672-15.2016

Eisenbarth, H., Oxner, M., Shehu, H. A., Gastrell, T., Walsh, A., Browne, W., & Xue, B. (in press). Emotional arousal pattern database (emap): A new database for psychophysiological responding to affective stimuli and its arousal pattern. *Psychophysiology*. https://www.wgtn.ac.nz/emap/

Filipovskỳ, J., Ducimetiere, P., & Safar, M. E. (1992). Prognostic significance of exercise blood pressure and heart rate in middle-aged men. *Hypertension*, *20*(3), 333–339. https://doi.org/10.1161/01.HYP.20.3.333

Fitch, D. T., Sharpnack, J., & Handy, S. L. (2020). Psychological stress of bicycling with traffic: Examining heart rate variability of bicyclists in natural urban environments. *Transportation Research Part F: Traffic Psychology and Behaviour*, *70*, 81–97. https://doi.org/10.1016/j.trf.2020.02.015

Gordis, E. B., Feres, N., Olezeski, C. L., Rabkin, A. N., & Trickett, P. K. (2010). Skin conductance reactivity and respiratory sinus arrhythmia among maltreated and comparison youth: Relations with aggressive behavior. *Journal of Pediatric Psychology*, *35*(5), 547–558. https://doi.org/10.1093/jpepsy/jsp113

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields i: A critical tutorial review. *Psychophysiology*, *48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Hancer, E., Xue, B., & Zhang, M. (2018). Differential evolution for filter feature selection based on information theory and feature ranking. *Knowledge-Based Systems*, *140*, 103–119. https://doi.org/10.1016/j.knosys.2017.10.028

Helwig, N. E. (2019). Statistical nonparametric mapping: Multivariate permutation tests for location, correlation, and regression problems in neuroimaging. *Wiley Interdisciplinary Reviews: Computational Statistics*, *11*(2), e1457. https://doi.org/10.1002/wics.1457

Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). Gradient flow in recurrent nets: The difficulty of learning long-term dependencies. In *A field guide to dynamical recurrent networks* (pp. 237–243). IEEE Press. https://doi.org/10.1109/9780470544037.ch14

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horn, M., Fovet, T., Vaiva, G., Thomas, P., Amad, A., & D'Hondt, F. (2020). Emotional response in depersonalization: A systematic review of electrodermal activity studies. *Journal of Affective Disorders*, *276*, 877–882. https://doi.org/10.1016/j.jad.2020.07.064

Keltner, D. (2019). Toward a consensual taxonomy of emotions. *Cognition and Emotion*, *33*(1), 14–19. https://doi.org/10.1080/02699931.2019.1574397

Kepler, J. (2008). *Scipy Simson's rule integration*. Retrieved June 23, 2022, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.simpson.html#scipy.integrate.simpson

Lazarus, R. S., Speisman, J. C., & Mordkoff, A. M. (1963). The relationship between autonomic indicators of psychological stress: Heart rate and skin conductance. *Psychosomatic Medicine*, *25*(1), 19–30.

Lemche, E., Anilkumar, A., Giampietro, V. P., Brammer, M. J., Surguladze, S. A., Lawrence, N. S., Gasston, D., Chitnis, X., Williams, S. C. R., Sierra, M., Joraschky, P., & Phillips, M. L. (2008). Cerebral and autonomic responses to emotional facial expressions in depersonalisation disorder. *The British Journal of Psychiatry*, *193*(3), 222–228. https://doi.org/10.1192/bjp.bp.107.044263

Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review*, *6*(2), 100–112. https://doi.org/10.1177/1754073913512003

Lin, H.-P., Lin, H.-Y., Lin, W.-L., & Huang, A. C.-W. (2011). Effects of stress, depression, and their interaction on heart rate, skin conductance, finger temperature, and respiratory rate: Sympathetic-parasympathetic hypothesis of stress and depression. *Journal of Clinical Psychology*, *67*(10), 1080–1091. https://doi.org/10.1002/jclp.20833

Loggia, M. L., Juneau, M., & Catherine Bushnell, M. (2011). Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *Pain*, *152*(3), 592–598. https://doi.org/10.1016/j.pain.2010.11.032

Mackersie, C. L., & Calderon-Moultrie, N. (2016). Autonomic nervous system reactivity during speech repetition tasks: Heart rate variability and skin conductance. *Ear and Hearing*, *37*, 118S–125S. https://doi.org/10.1097/AUD.0000000000000305

Madrigal, R. (2003). *Exploring subjective emotion intensity: Antecedents and postconsumption consequences*. ACR North American Advances.

Meese, R. A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, *14*(1–2), 3–24. https://doi.org/10.1016/0022-1996(83)90017-X

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(6), 275–285. https://doi.org/10.1002/cem.873

Nguyen, B. H., Xue, B., Andreae, P., & Zhang, M. (2021). A new binary particle swarm optimization approach: Momentum and dynamic balance between exploration and exploitation. *IEEE Transactions on Cybernetics*, *51*(2), 589–603. https://doi.org/10.1109/TCYB.2019.2944141

Norman, G. J., Berntson, G. G., & Cacioppo, J. T. (2014). Emotion, somatovisceral afference, and autonomic regulation. *Emotion Review*, *6*(2), 113–123. https://doi.org/10.1177/1754073913512006

Pat, N., Wang, Y., Bartonicek, A., Candia, J., & Stringaris, A. (2022). Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cerebral Cortex*, *33*, 2682–2703. https://doi.org/10.1093/cercor/bhac235

Patterson, J. C., II, Ungerleider, L. G., & Bandettini, P. A. (2002). Task-independent functional brain activity correlation with skin conductance changes: An fMRI study. *NeuroImage*, *17*(4), 1797–1806. https://doi.org/10.1006/nimg.2002.1306

Pyke, A., Rovira, E., Murray, S., Pritts, J., Carp, C. L., & Thomson, R. (2021). Predicting individual differences to cyber attacks: Knowledge, arousal, emotional and trust responses. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *15*(4). https://doi.org/10.5817/CP2021-4-9

Radford, M. (2004). Emotion and creativity. *Journal of Aesthetic Education*, *38*(1), 53–64. https://doi.org/10.2307/3527362

Roy, M. J., Costanzo, M. E., Jovanovic, T., Leaman, S., Taylor, P., Norrholm, S. D., & Rizzo, A. A. (2013). Heart rate response to fear conditioning and virtual reality in subthreshold ptsd. *Annual Review of Cybertherapy and Telemedicine*, *2013*, 115–119. https://doi.org/10.3233/978-1-61499-282-0-115

Schubring, D., & Schupp, H. T. (2021). Emotion and brain oscillations: High arousal is associated with decreases in alpha- and lower beta-band power. *Cerebral Cortex*, *31*(3), 1597–1608. https://doi.org/10.1093/cercor/bhaa312.

Shehu, H. A., Browne, W., & Eisenbarth, H. (2021). Particle swarm optimization for feature selection in emotion categorization. In *2021 IEEE congress on evolutionary computation (CEC)* (pp. 752–759). IEEE. https://doi.org/10.1109/CEC45853.2021.9504986

Sierra, M., Senior, C., Dalton, J., McDonough, M., Bond, A., Phillips, M. L., O'Dwyer, A. M., & David, A. S. (2002). Autonomic response in depersonalization disorder. *Archives of General Psychiatry*, *59*(9), 833–838. https://doi.org/10.1001/archpsyc.59.9.833

Song, Y.-Y., & Ying, L. U. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044

Su, X., Yan, X., & Tsai, C.-L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(3), 275–294. https://doi.org/10.1002/wics.1198

Sun, K. T., Lam, J. W. Y., Tai, W. C. S., Lau, B. W. M., & Yee, B. K. (2022). Within-subjects vs between-subjects co-variation of prepulse-elicited reaction and the diminution of startle to the succeeding pulse stimulus in the prepulse inhibition paradigm. *Behavioural Brain Research*, *430*, 113924. https://doi.org/10.1016/j.bbr.2022.113924

Svetnik, V., Liaw, A., Christopher Tong, J., Culberson, C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and qsar modeling. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1947–1958. https://doi.org/10.1021/ci034160g

Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research; How to choose a sampling technique for research. *International Journal of Academic Research in Management (IJARM)*, *5*(2), 18–27. https://doi.org/10.2139/ssrn.3205035

Tetereva, A., Li, J., Deng, J. D., Stringaris, A., & Pat, N. (2022). Capturing brain-cognition relationship: Integrating task-based fMRI across tasks markedly boosts prediction and test-retest reliability. *NeuroImage*, *263*, 119588. https://doi.org/10.1016/j.neuroimage.2022.119588

Venables, P. H., Gartshore, S. A., & O'riordan, P. W. (1980). The function of skin conductance response recovery and rise time. *Biological Psychology*, *10*(1), 1–6. https://doi.org/10.1016/0301-0511(80)90002-2

Weisberg, S. (2005). *Applied linear regression* (Vol. 528). John Wiley & Sons. https://doi.org/10.1002/0471704091

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, *15*(2), 70–73. https://doi.org/10.1109/TAU.1967.1161901

Xue, B., Zhang, M., & Browne, W. N. (2012). Multi-objective particle swarm optimisation (PSO) for feature selection. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation* (pp. 81–88). ACM. https://doi.org/10.1145/2330163.2330175

Xue, B., Zhang, M., & Browne, W. N. (2014). Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Applied Soft Computing*, *18*, 261–276. https://doi.org/10.1016/j.asoc.2013.09.018

Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2015). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, *20*(4), 606–626. https://doi.org/10.1109/TEVC.2015.2504420

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1: Supporting Information

How to cite this article: Shehu, H. A., Oxner, M., Browne, W. N., & Eisenbarth, H. (2023). Prediction of moment-by-moment heart rate and skin conductance changes in the context of varying emotional arousal. *Psychophysiology*, *60*, e14303. https://doi.org/10.1111/psyp.14303