Autoencoder Network for the Identification of Emotional Psychophysiological Responses in VR Urban Exploration

Chiara Maninetti[a,*], Gabriele Stancato[d], Barbara E.A. Piga[d], Marco Boffi[c], Carlotta Colonna[c], Nicola Rainisio[c], Lorenzo Mussone[1], Giandomenico Caruso[1], Luca Mainardi[a], Riccardo Barbieri[a]

[a]*Department of Electronics, Information and Bioengineering, Polytechnic of Milan, Giuseppe Ponzio Street, 34, 20133, Milan, Italy*
[b]*Department of Architecture and Urban Studies, Polytechnic of Milan, Edoardo Bonardi Street, 3, 20133, Milan, Italy*
[c]*Department of Cultural and Environmental Heritage, University of Milan, Noto Street, 6, 20141, Milan, Italy*
[d]*Department of Architecture and Urban Studies, Polytechnic of Milan, Giuseppe Ponzio Street, 34, 20133, Milan, Italy*

Abstract

Population-level analyses of autonomic responses to emotional stimuli implicitly assume homogeneous physiological rules, obscuring meaningful inter-individual heterogeneity. We propose an unsupervised autoencoder trained on HRV and EDA signals from 222 subjects to identify latent autonomic response patterns without predefined labels. Spectral clustering on reconstruction correlations identified two distinct subgroups, validated on an independent virtual reality urban walk dataset (69 participants). Cluster-stratified analyses revealed systematic between-cluster differences and subgroup-specific emotional sensitivity that were invisible at the aggregate level, demonstrat-

*Corresponding author
Email addresses: chiara.maninetti@polimi.com (Chiara Maninetti), gabriele.stancato@polimi.it (Gabriele Stancato), barbara.piga@polimi.it (Barbara E.A. Piga), marco.boffi@unimi.it (Marco Boffi), carlotta.colonna@unimi.it (Carlotta Colonna), nicola.rainisio@unimi.it (Nicola Rainisio), lorenzo.mussone@polimi.it (Lorenzo Mussone), giandomenico.caruso@polimi.it (Giandomenico Caruso), luca.mainardi@polimi.it (Luca Mainardi), riccardo.barbieri@polimi.it (Riccardo Barbieri)

ing that latent physiological heterogeneity can both mask true effects and generate spurious ones. The framework supports subgroup-specific characterization of autonomic responses with applications in environmental psychophysiology and affective computing.

Keywords: Autonomic nervous system, heart rate variability (HRV), electrodermal activity (EDA), deep learning, autoencoder, unsupervised learning, clustering, environmental psychophysiology, virtual reality, affective computing.

---

1. Introduction

A substantial body of literature investigating the autonomic responses to emotional stimuli focuses on the identification of direct associations between patterns of physiological activation and emotional states, often in search of generalizable "rules" linking measures such as heart rate variability (HRV) and galvanic skin response (GSR; electrodermal activity, EDA) to affective conditions. In this context, vagally mediated HRV has frequently been treated as an index of regulatory capacity and flexible adaptation, consistent with neurovisceral integration accounts [1, 2], and meta-analytic evidence supports a reliable but small association between higher HRV and better top-down self-regulation (including emotion regulation) [3]. Work on positive affect similarly suggests that elevated positive affect is often associated with higher vagally mediated HRV (e.g., RMSSD, HF power), but these ssociations vary substantially by context (rest vs. stress reactivity/recovery vs. daily life), operationalization (state vs. trait), and analytic choices [4]. In contrast, EDA/GSR is strongly sensitive to sympathetic-linked arousal and emotional intensity, and recent evidence in EDA-based emotion recognition indicates that arousal prediction typically outperforms valence prediction, implying that valence inference from EDA alone requires additional information or multivariate pattern analysis [5, 6].

The relationship between autonomic responses and emotional states, however, is far from being straightforward. Comprehensive reviews of autonomic activity in emotion emphasize that evidence for invariant, emotion-specific ANS "signatures" is limited across studies, and that the same emotion category can be realized through different appraisals, action tendencies, and task demands, yielding substantial situational variability in autonomic output [7, 8, 9]. Moreover, both HRV and GSR/EDA are multi-determined

measures: HRV indices depend not only on affective processes but also on respiration, posture, movement, circadian influences, and analysis choices, while EDA is sensitive to recording method and peripheral factors such as electrode placement and ambient conditions [10, 4, 11]. In line with this physiology, EDA-based emotion-recognition research consistently finds stronger links to arousal/intensity than to emotional valence, indicating that "one-to-one" mappings from EDA to discrete affective states may be especially fragile without contextual constraints [6, 9].

Individual differences further amplify this variability. Trait-like autonomic characteristics (e.g., resting vagal tone, electrodermal response lability) and stable differences in emotion regulation and coping styles modulate both baseline autonomic state and reactivity to emotional stimuli [12, 13, 14]. These moderators interact with biological and demographic factors (e.g., sex- and stress-phase-dependent HRV differences) as well as clinical status, for which umbrella-level evidence indicates reliably reduced HRV in several mental disorders [15, 16]. Importantly, induction method itself can shape observed autonomic patterns, such that the measured physiology reflects not only "what emotion" is elicited but also "how" it is elicited in the laboratory [17].

Consequently, averaged statistical analyses may obscure meaningful subgroup- or person-specific patterns of autonomic activation. Recent work on group-to-individual generalizability cautions that group-level associations do not necessarily describe within-person dynamics when processes are heterogeneous and time-varying, motivating analytic strategies that explicitly model inter-individual heterogeneity (e.g., moderation and latent profiles) and intra-individual dynamics rather than assuming a single, universal physiological rule set [18, 19].

In environmental psychophysiology, in particular, variations in emotional ratings of the stimuli are relatively minor, resulting in attenuated physiological responses which are hard to appreciate and categorize. Immersive environmental manipulations can produce measurable changes in autonomic activity, but not all effects are reflected across common physiological parameters and detectable differences are often subtle (e.g., perceptible on some measures but not others) [20]. Additionally, reviews of VR and natural environment exposure studies highlight heterogeneous and modest psychophysiological outcomes, with many designs lacking adequate power to distinguish small effects robustly [21]. Field studies using wearable sensors also show that only strong environmental contrasts (e.g., stressors) yield clear phys-

iological signal changes, with everyday environmental variations producing limited differentiation [22]. Finally, evidence suggests inconsistencies between psychological shifts and physiological metrics in response to natural stimuli, indicating that subtle autonomic changes often go undetected in standard analyses [23].

In this work, we present the case of a VR-based urban walk protocol in which participants's psychophysiological responses to various locations in the virtual environment were measured through both subjective emotional ratings and physiological monitoring. Specifically, we hypothesize that heterogeneous autonomic response patterns within the population may mask subpopulation-specific differences. Identifying groups of individuals with similar autonomic activation profiles may therefore uncover distinct patterns of response to environmental change and enable the derivation of subgroup-specific "rules". This approach has the potential to provide a more nuanced understanding of how urban design interventions influence psychophysiological wellbeing.

2. Methods

2.1. Autoencoder networks and anomaly detection

Autoencoder networks have recently emerged as a tool to analyze physiological signals, with the goals of extracting relevant features, denoising data, and detecting anomalies. The utility of autoencoder networks stems from their ability to learn compressed representations of complex data, encoding the information contained within the input into a low dimensional latent space that captures the most important features of the data. The network is then trained to decode this compressed representation to reconstruct the original data with minimal reconstruction error.

One of the primary applications of autoencoder networks is denoising: the encoder learns only the fundamental data patterns, discarding the superimpose noise artifacts which are thus absent in the decoder's reconstruction of the original input. Hence, dimensionality reduction and data compression go hand in hand with denoising, as demonstrated by Chiang et al. [24] who employed a fully convolutional denoising autoencoder to reduce the dimensionality of ECG signals while simultaneously improving signal quality.

One of the notable uses of autoencoder networks in the field of physiological time series analysis is anomaly detection. Anomalies can include, for instance, irregularities such as arrhythmia in ECG readings. In this approach,

the model is exclusively trained on normal datasets sourced from healthy individuals, so it can effectively learn the intrinsic characteristics of standard ECG sequences. Data to be analysed is passed through the encoder/decoder architecture, and the reconstruction error is computed. A minimal reconstruction error suggests that the new data align with the learned patterns that are typical of "standard" data (in this example, healthy ECG signals). In contrast, a substantial error indicates a deviation from these patterns, which indicates an anomaly. A discriminator network subsequently classifies this data by evaluating the autoencoder's loss function. This unsupervised method is advantageous as it operates without needing a labeled dataset.

In our research, we applied a similar approach, implementing an autoencoder network with the goal of idenfitying groups of subjects displaying different patterns of responses to emotional stimuli. Rather than using the reconstruction error to detect anomalies, we analyzed the network's performance in reconstructing three specific inputs corresponding to the time series of skin conductance, heart rate, and sympathovagal balance recorded in each experimental subject. By evaluating how well the autoencoder could reconstruct these physiological signals, we aimed to identify distinct patterns of autonomic activation across subjects, which could then be clustered into meaningful subgroups. This approach allowed us to uncover latent structures in the data that may correspond to different emotional response profiles, providing insights into the heterogeneity of autonomic responses to environmental stimuli.

Network Architecture

The proposed model follows an encode-fuse-decode structure that operates independently on each physiological channel before integrating cross-signal information through attention-based fusion. The network processes $F = 3$ input signals (HR, GSR, LF/HF), which are z-scored and windowed to $T = 150$ samples, and passed to the network in batches of size $B = 1024$ windows.

Per-channel encoder. Each physiological signal is encoded independently by its own encoder module, so that no cross-channel information is shared at this stage. Two alternative encoder architectures were implemented and compared: a) the *linear* encoder and b) the *hybrid* encoder.

The *linear* encoder is a four-layer fully connected network that processes each time step independently: a scalar input is projected through succes-

sive layers of widths $16n$, $8n$, $4n$, and n, where n denotes the per-channel latent dimension (number of neurons in the layer), with ReLU activations and dropout d after each hidden layer and a Tanh activation at the output. Because the linear encoder operates point-wise, it has no explicit mechanism for capturing local temporal structure, which is instead delegated entirely to the downstream attention layer. This network architecture is shown in Fig. 1.

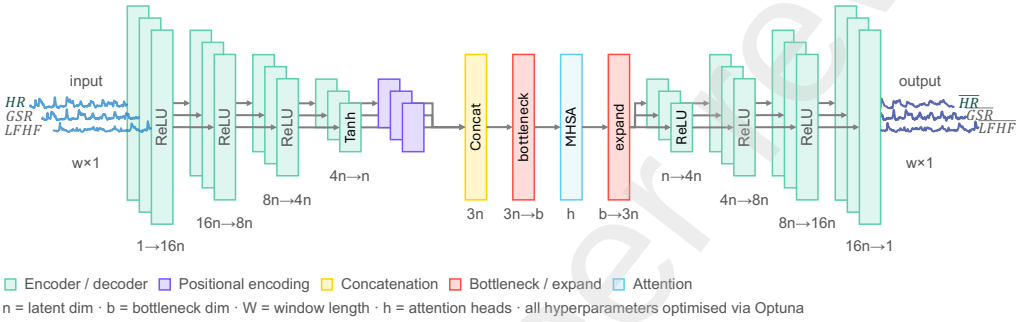

Figure 1: Architecture of the linear autoencoder network.

The *hybrid* encoder, shown in Fig. 2, augments this design with a convolutional front-end intended to extract local temporal features prior to compression. Two depthwise `Conv1D` layers with kernel size k and padding $\lfloor k/2 \rfloor$, preserving the temporal dimension, are applied in succession; depthwise convolution ensures that each channel remains independent at this stage. The resulting feature maps are projected to a higher-dimensional space via a grouped pointwise convolution, after which a three-layer MLP with widths $16n$, $8n$, $4n$, and output dimension n compresses the representation with a Tanh activation. Compared to the linear encoder, the hybrid variant can capture short-range temporal patterns within a receptive field of size k prior to global attention, at the cost of additional parameters and the hyperparameter k. In both cases the encoder output has shape $[B, T, n]$, preserving the temporal axis for downstream processing.

Channel fusion. After independent encoding, a cross-signal fusion step allows the model to capture inter-channel dependencies at each time step. The F per-channel latent vectors are first augmented with a learned channel identity embedding to break permutation symmetry across signals. Multi-head self-attention is then applied over the channel dimension at each time step,

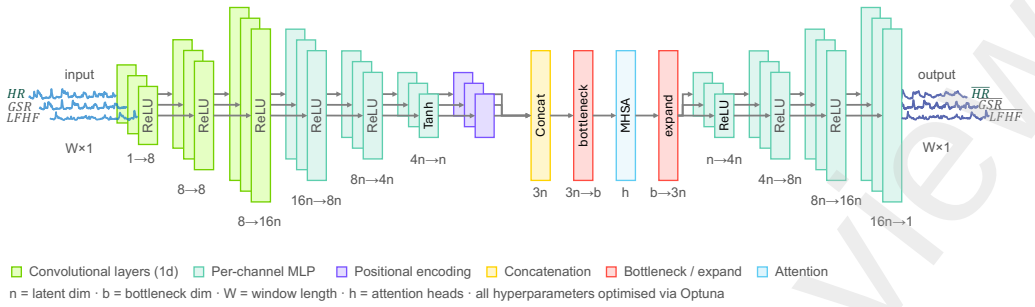

Figure 2: Architecture of the hybrid autoencoder network.

treating the F latents as a short sequence of length F; the attention output is added to the input via a residual connection. This operation produces a fused representation of shape $[B, T, F, n]$ in which each channel's latent has been contextualised by the others.

Joint bottleneck. A bottleneck module compresses the concatenated multi-channel latent into a low-dimensional shared space. The F per-channel latents are concatenated along the feature axis to form a tensor of dimension $F \cdot n$, which is projected by a linear layer followed by a Tanh activation to a bottleneck vector of dimension b. Prior to decoding, $\mathbf{z}_b$ is expanded back to dimension $F \cdot n$ by a second linear layer, and the result is partitioned to recover the F per-channel representations.

Temporal attention and positional encoding. Following channel fusion and optional bottleneck compression, temporal self-attention is applied independently to each per-channel latent sequence of shape $[B, T, n]$ to model long-range dependencies across the window. Three positional encoding strategies were implemented and evaluated. The *learned positional embedding (pos(t))* encoding consists of a learned absolute position embedding: a lookup table maps each integer time step $t \in \{0, \ldots, T-1\}$ to an n-dimensional vector that is added to the latent prior to attention. Each position is encoded independently, with no structural constraint on the relationship between adjacent steps. The *earned continuous time embedding (timeMLP)* encoding replaces the discrete lookup with a small two-layer MLP that maps a scalar time value (either an explicit timestamp or a uniform normalisation over $[0, 1]$) to an n-dimensional embedding summed with the latent. Because the mapping is continuous and smooth, neighbouring time steps receive similar embeddings, making this approach suited to variable-length or irregularly sampled

sequences. Finally, the *relative position bias (relPos)* strategy departs from the additive embedding paradigm entirely: rather than modifying the latent representation, positional information is injected directly into the attention score matrix. For each pair of positions (i, j), the signed relative distance $i-j$ is clamped to a symmetric range $[-r, r]$ and used to index a learned scalar bias; this bias is added to the query-key dot product before the softmax.

Per-channel decoder.. Each channel is reconstructed independently by a symmetric four-layer MLP that mirrors the linear encoder: the per-channel latent of dimension n is expanded through layers of widths n, $4n$, $8n$, $16n$, and finally projected to a scalar output, with ReLU activations and dropout at each hidden layer. The outputs of the F decoders are concatenated to yield the reconstructed signal tensor of shape $[B, T, F]$.

2.2. Training Data

Network training was performed on data from 6 publicly available datasets, to obtain a sufficient amount of data representative of a general population. All selected datasets included multimodal physiological recordings (including ECG and GSR) collected during psychophysiological protocols involving emotion stimulation and recognition, and/or stress detection. In particular, the following datasets were sourced from:

1. *Stress Recognition in Automobile Drivers*, in which 17 individuals drove a designated path in Boston designed to provoke varying stress levels. Collected metrics were ECG, EMG, foot and hand GSR, and respiration[25].
2. *Electrocardiogram, skin conductance and respiration from spider-fearful individuals watching spider video clips* included 57 participants suffering from arachnophobia who watched spider-related videos while measurements such ECG, GSR, and respiration were collected [26].
3. *Continuously Annotated Signals of Emotion (CASE)* involved 30 subjects who viewed 8 videos eliciting different arousal and valence levels, and collected physiological signals comprising ECG, GSR, PPG, respiration, EMG, and skin temperature[27].
4. *TRINA-EmotionDrive* is an open dataset of emotional responses collected during a simulated driving protocol via six biosignals (ECG, EDA, abdominal and chest respiration, skin temperature, and PPG) and self-report questionnaires[28].

5. *Cognitive Load, Affect and Stress (CLAS)* includes synchronized recordings of physiological signals (ECG, PPG, EDA/EDA, and accelerometer) from 62 healthy volunteers engaged in interactive and perceptive tasks [29].
6. *Mental workload assessment using N-back task with wearable sensors (MAUS)* provides synchronized ECG, fingertip PPG, wrist-band PPG, and GSR signals from 22 subjects performing N-back tasks and includes subjective ratings (NASA-TLX) for workload [30].

Initial exploratory analysis of the dataset extracted ECG and EDA recordings from the multimodal datasets. Low quality signals were excluded after identification based on signal to noise ratio and physiological plausibility of automatically detected heart rate [31]. Then, the signals were preprocessed to obtain an uniform set of data from 222 experimental subjects. Specifically, the ECG signals were resampled at a constant sampling rate of 256Hz, R peaks were automatically annotated, and manual review corrected for artefacts and ectopic beats. The RR interval and the Heart Rate time series were then extracted. Point Process modeling of RR time series was performed to obtain a continuous time estimation of spectral band power; continuous time sympathovagal balance (LF/HF) was computed as the ratio between low frequency and high frequency band powers. GSR time series were filtered with a $[0.01 - 4Hz]$ bandpass filter, downsampled to 8Hz, and cleaned of motion artefacts via the wavelet method described by Weixuan Chen et al. [32]. A schematic representation of the preprocessing pipeline is provided in Figure 3.

The processed data was further downsampled to a sampling rate of 1Hz and smoothed with a 2s rolling-average window. A logarithmic transform was applied to the sympathovagal balance signals prior to normalization to reduce skewness. Data was then normalized via Z-score standardization using a scaler fitted on the training set. Finally, the data was segmented into fixed-length windows of 150 seconds with 30% overlap (stride = 70% of window length).

2.3. Hyperparameter Search and Network Training

Hyperparameters were optimized with Optuna using a Tree-structured Parzen Estimator (TPE) sampler, minimizing a model selection objective. The search space included:

- Encoder architecture type (linear / hybrid);

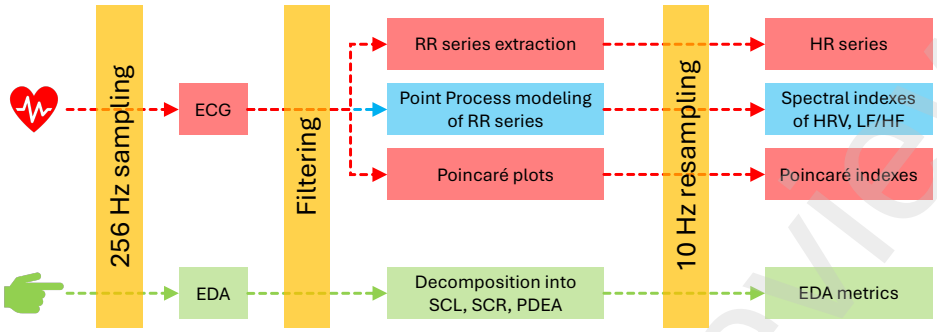

Figure 3: Processing pipeline for the extraction of autonomic indices from raw physiological signals.

- Positional encoding type (post / time MLP / relative bias);
- Latent dimension n (neurons per channel) in $\{4, 6, 8, 10\}$, dropout rate in $\{0.07, 0.01, 0.001, 0\}$, and learning rate in $\{10^{-3}, 10^{-4}, 10^{-5}\}$;
- Number of attention heads and shared bottleneck dimension;
- Weights for all six loss components (see below), normalized to a unit sum;
- Per-signal feature weights.

All models were trained with the Adam optimizer with early stopping. The top-scoring trials by objective value were retrained for 300 epochs and saved as checkpoints.

Instead of relying solely on the mean squared error, model optimization minimized a multi-term composite loss function designed to preserve not only amplitude fidelity but also the temporal and spectral structure of physiological dynamics. The loss was computed in two stages. First, for each output channel $c \in \{\text{SC, HR, LF/HF}\}$, the six loss components were combined into a single per-channel loss $\mathcal{L}_c$ as a weighted inner product:

$$\mathcal{L}_c = \mathbf{w}^\top \boldsymbol{\ell}^{(c)}, \qquad \sum_i w_i = 1, \quad w_i \geq 0, \tag{1}$$

where $\mathbf{w} = [w_{\text{rec}}, -w_{\text{corr}}, w_{\text{trend}}, w_{\text{freq}}, w_{\text{smooth}}, w_{\text{var}}]^\top$ is the vector of component weights, and $\boldsymbol{\ell}^{(c)} = [\mathcal{L}^{(c)}_{\text{rec}}, \mathcal{L}^{(c)}_{\text{corr}}, \mathcal{L}^{(c)}_{\text{trend}}, \mathcal{L}^{(c)}_{\text{freq}}, \mathcal{L}^{(c)}_{\text{smooth}}, \mathcal{L}^{(c)}_{\text{var}}]^\top$ is the vector of the following per-channel loss components:

- $\mathcal{L}_{\text{rec}}$: pointwise reconstruction error (MAE, MSE, or Huber), ensuring sample-level amplitude accuracy;
- $\mathcal{L}_{\text{corr}}$: negative Pearson correlation loss, maximizing temporal shape similarity between prediction and ground truth independently of amplitude;
- $\mathcal{L}_{\text{trend}}$: derivative-based loss on first-order differences, penalizing mismatched temporal dynamics;
- $\mathcal{L}_{\text{freq}}$: spectral loss on FFT magnitude, enforcing fidelity in oscillatory components with emphasis on high-frequency content relevant to autonomic signals;
- $\mathcal{L}_{\text{smooth}}$: second-derivative penalty (Huber or L1), regularizing against spurious oscillations in the reconstruction;
- $\mathcal{L}_{\text{var}}$: relative variance-matching penalty, preventing degenerate constant reconstructions.

The per-channel losses were then aggregated into the overall training objective via a weighted average across channels:

$$\mathcal{L}_{\text{total}} = \mathbf{v}^{\top}\boldsymbol{\lambda}, \qquad \sum_{c} v_c = 1, \quad v_c \geq 0, \tag{2}$$

where $\mathbf{v} = [v_{\text{SC}}, v_{\text{HR}}, v_{\text{LF/HF}}]^{\top}$ is the vector of per-channel weights and $\boldsymbol{\lambda} = [\mathcal{L}_{\text{SC}}, \mathcal{L}_{\text{HR}}, \mathcal{L}_{\text{LF/HF}}]^{\top}$ is the vector of per-channel losses.

The per-channel weights w_c allow the contribution of each output signal to be adjusted during the hyperparameter search, for instance to up-weight skin conductance if it is considered the primary reconstruction target. Both the component weights w_i and the channel weights w_c are included in the hyperparameter search space.

2.4. Clustering and Model Selection

The network configuration achieving the highest Optuna objective score was selected and fully retrained for 300 epochs. This model is hereby referred to as M_b (*best model*). The selected configuration used a linear MLP encoder with Time MLP positional encoding, $n = 6$ neurons per channel, 1 attention head, a bottleneck dimension of 2, a dropout rate of 0.01, a learning rate of $\eta = 10^{-3}$, and a Huber reconstruction loss.

Each subject's physiological recordings were then passed through M_b, and the per-channel reconstruction losses (one value per physiological signal) were computed for each subject, yielding a three-dimensional feature vector encoding how well the model reconstructed each signal. Spectral clustering was applied to these reconstruction loss vectors to partition subjects into groups with distinct reconstruction profiles. The number of clusters was fixed at 2, and the resulting partition is referred to as C_b *(best clustering)*. This binary partition reflects a deliberate modeling assumption: one cluster captures subjects whose autonomic responses conform to the dominant pattern learned by the network (typical responders), while the other aggregates those whose signals deviate from it. In principle, a purely data-driven approach with no a priori constraint on the number of clusters could reveal a richer subgroup structure. However, reliably identifying more than two latent groups would require a substantially larger dataset; two clusters represent the minimal partition that distinguishes typical vs. atypical autonomic profiles while remaining statistically tractable given the available sample size.

2.5. Validation

2.5.1. The Validation Dataset

The validation dataset consisted in physiological data (including ECG and Skin Conductance), psychological metrics, and sociodemographic information collected on 69 experimental subjects during a Virtual Reality urban walk protocol. The virtual environment was developed in Unity [33] as a three-dimensional simulation of a Milanese street (Via Celoria, part of the Politecnico and Statale campus area in the Città Studi neighborhood), reproducing its spatial configuration and visual characteristics, together with its urban soundscape. The model was implemented in a highly photorealistic manner and incorporates dynamic elements such as moving vehicles and pedestrians. Participants experienced the scenario in immersive virtual reality and navigated the environment using a point and teleport locomotion technique [34] along a predefined path. The virtual environment was presented to each subject in two scenarios: one depicting the area as it currently appears in real life (PS, Present State), and the other depicting a simulation incorporating the proposed viability changes (VC, Viability Changes). Subjects walked the same route in both scenarios, which included 6 specific Points of Interest (POI). At each POI, the virtual environment prompted subject to rate their emotional response to surroundings via a brief questionnaire. The order of presentation of the scenarios was randomized for each subject.

Between the two sections of the experiment, subjects rested comfortably on a chair for 5 minutes to rest after the first half of the protocol and to acquire baseline physiological signals.

During the entire VR walk and baseline rest period, physiological signals were continuously recorded via wearable devices. The electrocardiographic signal was measured via a Shimmer3 ECG unit, with electrodes positioned on the subject's shoulder and left arm, while skin conductance was recorded via a Shimmer3 GSR unit, with electrodes placed on the index and middle fingers of the non dominant hand.

Psychological responses were collected at each POI through a digital questionnaire based on the exp-EIA™ framework [35], which asks subjects to rate their emotional responses along the affective dimensions of Valence and Arousal. The digital survey included four additional psychological scales adapted from the model proposed by Perugini et al. [36] for the emotional rating of environmental contexts, which were implemented as semantic differential scales in the VR environment, as described by Betella and Verschure [37]. In particular, the scales correspond to four bipolar dimensions of the emotional experience: *Boring/Stimulating*, *Depressing/Exciting*, *Relaxing/Stressful*, and *Unpleasant/Pleasant*. Each scale's slider ranged from -4 to +4, with negative values corresponding to the emotion at the lower end of the scale and positive values to the upper end.

2.5.2. Network Validation Procedure

Data preparation for the validation dataset followed the same preprocessing steps that were applied to the training dataset (see Section 2.2), to ensure consistency in the datasets. Data was normalized applying the same standard scaler that was initially fitted on the training dataset, and segmented into windows. For each subject, the three-channel composite loss function was computed as the average of the losses over all windows. Finally, C_b was applied to assign each validation subject to a cluster.

Sociodemographic influences on cluster assignment. To assess whether cluster assignment was associated with sociodemographic characteristics, we fitted a linear mixed-effects model with cluster membership (0 vs. 1) as the dependent variable, age range and gender as fixed-effect predictors, and a random intercept per subject to account for repeated observations. Neither the age range factor ($F(3, 64) = 2.13$, $p = 0.105$) nor gender ($F(1, 64) = 2.83$, $p = 0.097$) reached significance at the omnibus level. At the level of individual

coefficients, however, subjects in the 50–64 age group showed a significantly higher probability of being assigned to Cluster 1 compared to the 18–33 reference group ($\hat{\beta} = 0.333$, $SE = 0.159$, $t(64) = 2.09$, $p = 0.040$); all other age ranges and gender did not differ significantly from their respective references (all $p > 0.09$). Distribution of the gender, age range, and cluster labels is reported in Table 1.

Table 1: Distribution of Clusters by sociodemographic variables

Age Range	**Cluster**		**Total**
	0	**1**	
18–33	9	9	18
34–49	7	8	15
50–64	3	15	18
65–80	9	9	18

	Gender	**Cluster**		**Total**
		0	**1**	
	F	17	17	34
	M	11	24	35

Age Range	**Gender**	**Cluster**		**Total**
		0	**1**	
18–33	F	5	4	9
	M	4	5	9
34–49	F	6	3	9
	M	1	5	6
50–64	F	2	7	9
	M	1	8	9
65–80	F	4	3	7
	M	5	6	11
Total		**28**	**41**	**69**

2.6. Evaluation of Psychophysiological Responses on Full and Clustered Validation Dataset

2.6.1. Emotional Labeling of Protocol Segments

The subjective psychological ratings assigned by each subject at each POI were categorized into three levels (low, medium, high) based on the following thresholds:

- "Low": between -4 and -2.67;
- "Medium": between -2.66 and 2.66;
- "High": between 2.67 and 4,

Each segment of the recording, corresponding to the time window recorded while one subject was standing at one POI during one presented scenario, was thus assigned to one of the three classes for each of the four psychological scales. In total, 978 segments were considered. Their distribution into the psychological metric classes, as well as the discrete labels and their respective thresholds are shown in table 2.

Table 2: Emotional labels and corresponding thresholds. Numbers in parentheses indicate the number of segments in each category.

Scale	**"Low" (-4 to -2.67)**	**"Medium" (-2.66 to 2.66)**	**"High" (2.67 to 4)**
Boring*/*Stimulating	*Boring* (119)	*Neutral* (554)	*Stimulating* (349)
Depressing*/*Exciting	*Depressing* (87)	*Neutral* (542)	*Exciting.* (469)
Relaxing*/*Stressful	*Stressful* (372)	*Neutral* (492)	*Relaxing.* (114)
Unpleasant*/*Pleasant	*Unpleasant* (63)	*Neutral* (506)	*Pleasant.* (409)

2.6.2. Metrics of Autonomic Activation

Signal processing and metrics extraction were performed as described in [38] and represented in Fig. 3. The following metrics were computed for each segment:

- Time-domain metrics of heart rate variability: SD1, SD2, SD1SD2, SD1/SD2, AVNN, logRMSSD;
- Low- and High-frequency power (LF and HF), and sympathovagal balance (LF/HF), computed through the Point Process approach[39];
- Indices of electrodermal activity: the number of skin conductance responses per second (NPS), and their average amplitude (A_P) and rising time (τ_P).

A full description of the computed metrics is provided in Table 3.

Indexes were computed twice on each segment: with normalization with respect to baseline, and without normalizing (in raw form). In all cases, data were analyzed across conditions defined by one within-subject factor (the emotional rating of the segment, acording to a specific psychological scale) and one between-subject factor (the subjects' cluster). For each condition, normality was tested using the Lilliefors test, noting that some dependent variables violated the assumption of normality $p < 0.05$.

Therefore, non-parametric methods were used instead of ANOVA for all variables. Within-subject effects were assessed using the Friedman test, the non-parametric equivalent of repeated-measures ANOVA. Between-subject effects were evaluated using the Kruskal-Wallis test. In case of a main effect reaching statistical significance $p < 0.05$, post-hoc multiple comparisons with Dunn's correction were performed to identify specific pairs of conditions showing significant differences.

Table 3: Computed indexes of autonomic activation.

Physiological Signal	**Derived Metrics**	**Description**
RR series	LF	**Low-Frequency Power**. Reflects both sympathetic and parasympathetic modulation of heart rate variability (HRV).
	HF	**High-Frequency Power**. Primarily reflects parasympathetic (vagal) modulation of HRV.
	LF/HF	**Sympathovagal Balance**. Ratio of sympathetic to vagal modulation.
	SD1	**Standard Deviation in Minor Axis**. Short-term HRV, related to vagal modulation.
	SD2	**Standard Deviation in Major Axis**. Long-term autonomic fluctuations in HRV.
	SD1SD2	**Axes Product**. Proportional to the area of the Poincaré ellipse.
	SD1/SD2	**Axes Ratio**. Ratio between short- and long-term HRV.
	AVNN	**Average Normal-to-Normal Interval**. Average duration of a normal-to-normal RR interval.
	logRMSSD	**Logarithmic Root Mean Square of Successive Differences**. RMS of successive RR intervals, in logarithmic form. Related to short-term HRV.
EDA signal	NPS	**Number of Peaks per Second**. Number of skin conductance responses per second
	A_P.	**Peak Amplitude**. Average amplitude of skin conductance responses.
	τ_P	**Peak Rising Time**. Average time taken for skin conductance responses to reach their peak amplitude.

2.6.3. Principal Component Analysis

To characterize the multivariate structure of autonomic responses and compare feature relevance across clusters, we performed an ALS-based Principal Component Analysis (PCA) on three datasets: the full cohort, the subset of Cluster 0 subjects, and the subset of Cluster 1 subjects. The analysis included the metrics of autonomic activation described in Table 3. Prior to PCA, each variable was independently scaled to the $[0, 1]$ range using min-max normalization to ensure comparable numerical ranges and to prevent dominance by features with larger absolute magnitudes.

For each PCA model, variable importance was quantified by combining loading magnitude and component explanatory power. Let l_{jk} denote the loading of variable j on component k, and let λ_k be the k-th eigenvalue of the covariance matrix, so that the explained variance ratio of component k is

$$\mathrm{EV}_k = \frac{\lambda_k}{\sum_{k'} \lambda_{k'}}.$$

For each component, absolute loadings were normalized to $[0, 1]$:

$$\tilde{l}_{jk} = \frac{|l_{jk}|}{\max_j |l_{jk}|}.$$

A global importance score was then computed for each variable as

$$I_j = \sum_k \mathrm{EV}_k \, \tilde{l}_{jk}.$$

Finally, importance profiles were compared across the three PCA models (overall dataset, Cluster 0, Cluster 1) to identify shared and cluster-specific physiological drivers.

3. Results

3.1. Network Training

The Optuna TPE search completed 1,528 valid trials out of 1,612 total. The search converged clearly: all top-performing configurations shared the same encoder type (linear MLP), positional encoding (Time MLP), reconstruction loss (Huber), number of hidden neurons ($n = 6$), and learning rate ($\eta = 10^{-3}$), with minor variation in the number of attention heads (1 or 4) and dropout rate (0.07–0.10).

The best model (Trial 1382) achieved an aggregated loss score of $loss_{best} = -0.912$ with a mean reconstruction correlation of $\bar{r} = 0.977$ across all three physiological channels, confirming high-fidelity reconstruction and strong cluster separation. Notably, configurations using the hybrid Conv1D encoder or non-Huber loss functions consistently underperformed relative to the top cluster.

3.2. Evaluation of Psychophysiological Responses on Full and Clustered Validation Dataset

Spectral clustering applied to the autoencoder reconstructions on the validation dataset partitioned the 69 subjects into two groups: 28 subjects in Cluster 0 and 41 in Cluster 1 (see Table I). Inspection of the per-channel reconstruction error distributions (Fig. 4) reveals a clear qualitative distinction between the two groups. Cluster 0 displays compact, low-dispersion error distributions across all three channels, and can therefore be interpreted as representing a more *uniform* autonomic response pattern captured by the model. Cluster 1, by contrast, exhibits a more variegated response profile: while its HR reconstruction errors are comparably distributed to those of Cluster 0, the errors on LF/HF and SC are substantially higher and considerably more sparse, indicating that the model struggles to coherently reconstruct these channels for this subgroup. This differential reconstruction behaviour suggests that the two clusters capture subjects whose autonomic responses are dominated by different physiological modalities.

3.2.1. Inter-cluster Evaluation

When considering physiological responses only, almost all metrics exhibit significant between-cluster differences in at least one comparison (Table 4). In raw form, HF, NPS, P_A, SD1, SD2, SD1SD2, and logRMSSD all reach significance; after normalization, only SD2 and SD1SD2 retain significant between-cluster differences.

When stratifying by levels of subjective emotional ratings, between-cluster differences were widespread and dominated by raw P_A, which reached significance at virtually every level of every psychological scale (Table 7). Additional HRV metrics—SD1, SD2, SD1SD2, and logRMSSD—showed raw between-cluster differences at selected levels, most notably at the *Boring* end of the *Boring/Stimulating* scale, at the *Depressing* and *Neutral* levels of *Depressing/Exciting*, at the high end of *Relaxing/Stressful*, and at the

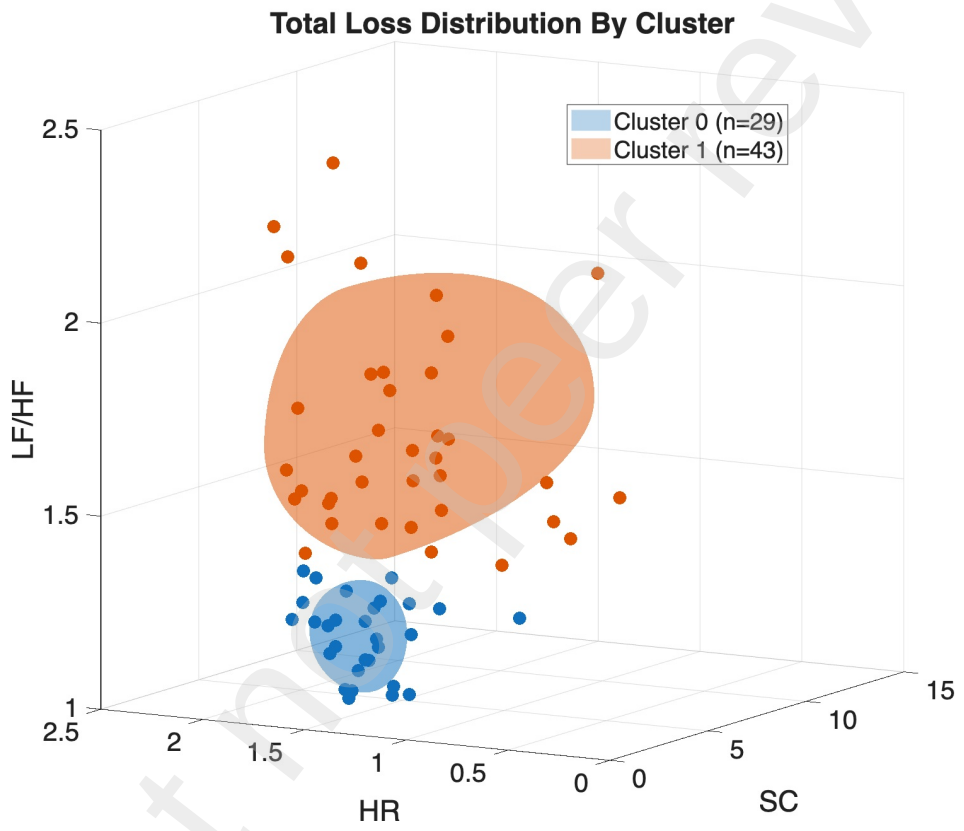

Figure 4: The distribution of total reconstruction loss over the three channels for the two clusters.

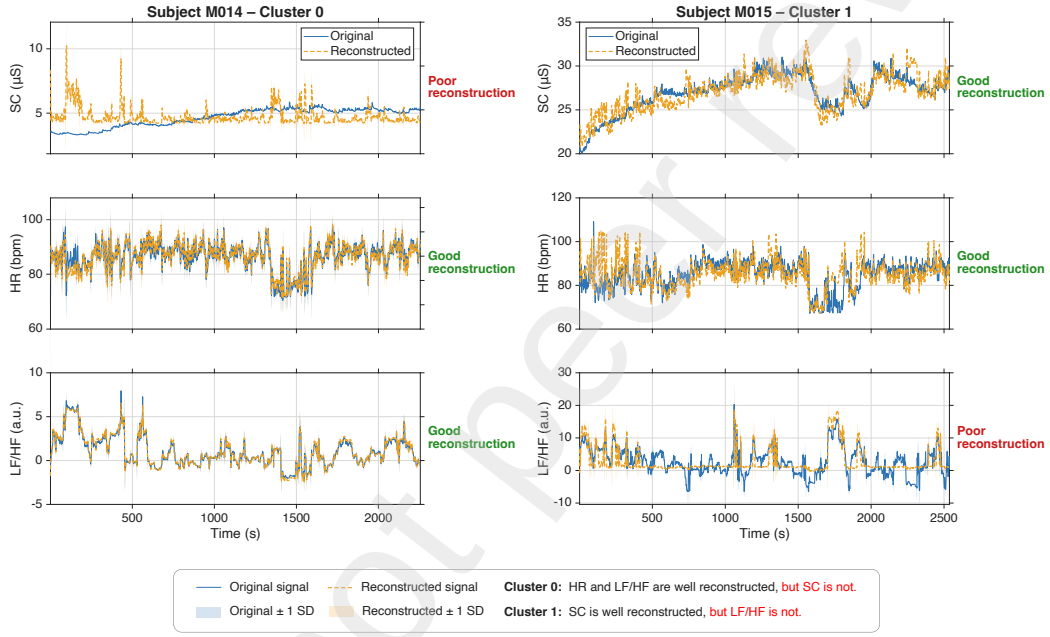

Figure 5: Example of the original and reconstructed physiological signals for a Cluster 0 (left side plots) and a Cluster 1 (right side plots) subject from the validation dataset. Notice how, in the Cluster 0 case, the network faithfully reconstructed the HR and LF/HF time series for the Cluster 0 subject, but struggled with the SC signal. Viceversa, SC was better reconstructed in the Cluster 1 subject, while the reconstruction of LF/HF was much less accurate.

Unpleasant and *Neutral* levels of *Unpleasant/Pleasant*. AVNN reached significance only in normalized form and exclusively at the *Unpleasant* level.

Table 4: Significant differences in physiological metrics between Cluster 0 and Cluster 1 subjects. R: raw indexes, N: normalized with respect to baseline conditions.
*: $p < 0.05$, **: $p < 0.01$

	HF	**NPS**	P_A	**SD1**	**SD2**	**SD1SD2**	**logRMSSD**
R	**	***	***	***	***	***	***
N					**	**	

Table 5: Significant differences between low, medium, and high levels of psychological ratings in the overall dataset. R: raw indexes, N: normalized with respect to baseline conditions.
*: $p < 0.05$, **: $p < 0.01$

Comparison	**Level 1**	**Level 2**	**LF**
Arousal	Low	Medium	* (R/N)
Boring/Stimulating	High	Medium	* (R/N)
Depressing/Exciting	High	Low	* (N)
	Low	Medium	* (N)

3.2.2. Intra-cluster Evaluation

Significant differences between levels of physiological ratings over the whole dataset were limited. LF power was the only metric showing significant effects across multiple comparisons: between Low and Medium levels of Arousal, between Medium and High levels of *Boring/Stimulating*, and between multiple level pairings of *Depressing/Exciting* (both raw and normalized for the first two, normalized only for the latter; see Table 5).

When considering the two clusters separately, more differentiated patterns emerged (Table 6). Cluster 1 showed a single significant intra-cluster effect: raw HF power differed between *Unpleasant* and *Neutral* levels of *Unpleasant/Pleasant*. Cluster 0 displayed a richer intra-cluster structure. Along the *Boring/Stimulating* scale, AVNN was the main discriminator, reaching

significance between *Boring* and *Neutral* levels (both raw and normalized) and between *Neutral* and *Stimulating* levels (normalized only), with LF/HF also reaching significance for the latter contrast. Within *Depressing/Exciting*, normalized HF was the most consistent discriminator, significant across all three level pairings; SD1 and SD1SD2 (both normalized) also reached significance for the *Depressing* vs. *Neutral* comparison. Finally, LF (normalized) and LF/HF (raw) showed significant effects between *Neutral* and *Pleasant* levels of *Unpleasant/Pleasant* in Cluster 0. The full list of significant differences between levels of psychological ratings in the two clusters is reported in Table 6.

Table 6: Significant differences between low, medium, and high levels of psychological ratings over the two cluster subsets of the dataset.
Abbreviations: Ar. = Arousal, Pleas. = Pleasure, B/S = Boring/Stimulating, D/E = Depressing/Exciting, R/S = Relaxing/Stressful, U/P = Unpleasant/Pleasant.
R: raw indexes, N: normalized with respect to baseline conditions.
*****: $p < 0.05$, ******: $p < 0.01$, *******: $p < 0.001$

	Scale	**Level 1**	**Level 2**	**AVNN**	**HF**	**LF**	**LF/HF**	**SD1**	**SD1SD2**
Cl. 1	U/P	Low	Medium		* (R)				
Cl. 0	B/S	Low	Medium	** (N/R)					
		Medium	High	* (N)			* (N)		
	D/E	Depres.	Exciting		* (N)				
			Neutral		*** (N)			* (N)	* (N)
		Neutral	Exciting		* (N)	** (N)			
	Pleas.	Medium	High			* (N)			
	U/P	Neutral	Pleas.				** (R)		

Finally, baseline-normalized and non-normalized indices did not yield identical inferential patterns. In several cases, normalization increased sensitivity to emotional-level effects after cluster stratification: such is the case of AVNN in Cluster 0 along the *Boring/Stimulating* scale, where significance between *Neutral* and *Stimulating* levels was observed only after normalization, and of HF power in Cluster 0 for *Depressing/Exciting*, which reached significance exclusively in the normalized version.

Table 7: Significant differences in physiological metrics between Cluster 0 and Cluster 1 subjects at different levels of the psychological scales.
Abbreviations: Ar. = Arousal, Pleas. = Pleasure, B/S = Boring/Stimulating, D/E = Depressing/Exciting, R/S = Relaxing/Stressful, U/P = Unpleasant/Pleasant.
R: raw indexes, N: normalized with respect to baseline conditions.
*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

	Level	AVNN	HF	NPS	P_A	SD1	SD2	SD1SD2	SD1/SD2	logRMSSD
Ar.	Medium			* (R)	*** (R)					
	High				* (R)					
B/S	Boring				** (R)	* (R)	*** (R)	* (R)		** (R)
	Neutral				** (R)		* (R)			
	Stim.				* (R)					
D/E	Depres.		* (N)	* (R)	* (R)		* (R)			
	Neutral				*** (R)	* (R)	* (R)			* (R)
Pleas.	Medium				** (R)	* (R)	* (R)		* (R)	** (R)
	High				* (R)					
R/S	Relaxing				** (R)					
	Neutral				** (R)					
	Stress.					* (R)	* (R)	* (R)		** (R)
U/P	Unpl.	** (N)			* (R)		* (R)	* (R)		
	Neutral				** (R)		* (R)			* (R)
	Pleas.				* (R)					

3.2.3. PCA Results

PCA-based importance profiling indicated distinct multivariate organizations across clusters. In Cluster 0, the representation was relatively distributed across both HRV and EDA metrics: SD1/SD2 was the leading contributor (16.5%), followed by P_A (13.8%), AVNN (12.4%), and τ_P (11.4%), with remaining metrics contributing more modestly. In Cluster 1, the leading variable was τ_P (19.8%), followed by AVNN (17.1%), SD1/SD2 (13.1%), and P_A (12.5%), suggesting a more balanced interplay between electrodermal and cardiovascular features compared to Cluster 0.

When PCA was applied to the full dataset without cluster separation, the four most influential variables were SD1/SD2 (20.5%), P_A (18.2%), AVNN (15.4%), and τ_P (13.3%), together accounting for approximately 67% of the total explained variance, while the remaining metrics contributed more mod-

Table 8: PCA variable importance scores (normalized by explained variance) for Cluster 0, Cluster 1, and the full dataset. Values are percentage contributions; rows are sorted by total importance.

Variable	**Cluster 0**	**Cluster 1**	**Total**
SD1/SD2	16.49	13.13	20.51
P_A	13.78	12.50	18.23
AVNN	12.41	17.12	15.39
τ_P	11.42	19.78	13.26
SD1	8.04	5.04	8.44
SD2	8.38	8.22	5.92
logRMSSD	7.15	5.80	6.35
SD1SD2	6.68	3.30	4.25
HF	4.83	3.94	1.63
LF	4.17	3.06	1.83
LF/HF	3.53	3.59	2.99
NPS	3.13	4.52	1.21

estly. This pattern matches the feature importance profile of Cluster 0, which represents the "typical" autonomic response cohort.

4. Discussion

The physiological response profiles uncovered by the autoencoder reveal distinct autonomic organization patterns across clusters: Cluster 0 is primarily characterized by HRV variability indices (SD1/SD2) alongside electrodermal amplitude (P_A), while Cluster 1 exhibits a distinct pattern in which electrodermal dynamics (τ_P) and mean heart rate (AVNN) are the dominant drivers, with sympathovagal balance metrics playing only a secondary role.

Beyond the structural differences identified by the network, the inferential analyses of between-cluster and within-cluster effects further clarify the functional relevance of the discovered subgroups. Between-cluster comparisons revealed systematic divergences in both absolute physiological levels and baseline-normalized reactivity. In particular, Cluster 0 generally exhibited lower raw HRV and electrodermal values but, for SD2 and P_A, a diver-

gent pattern of normalized reactivity, suggesting that cluster differentiation reflects complex reorganization rather than simple scaling effects. Complementarily, within-cluster analyses showed that emotional-level discrimination was not evenly distributed across subgroups. Cluster 1 was largely insensitive to within-subject emotional variation, with only a single significant effect (raw HF for *Unpleasant/Pleasant*), whereas Cluster 0 showed selective sensitivity across multiple scales: AVNN along *Boring/Stimulating*, normalized HF power throughout *Depressing/Exciting*, and LF/HF for *Unpleasant/Pleasant*.

These findings have important methodological and practical implications. On one hand, pooling heterogeneous distributions can mask true effects if a physiological variable is emotionally responsive in only one cluster; the unresponsive subgroup dilutes the overall signal, reducing statistical power. On the other hand, combining two non-responsive distributions can create spurious significance through mixture effects, as demonstrated by the LF results under the *Low Arousal vs Medium Arousal* comparison. Consequently, accurate characterization of autonomic responses to emotional stimuli requires cluster-stratified analysis, enabling the identification of subgroup-specific associations that would otherwise remain obscured.

From an applied perspective, the identification of physiologically coherent subgroups has direct relevance for personalized intervention design in domains such psychophysiology-driven urban planning [40], virtual reality therapy, and affective computing. Rather than applying population-averaged design principles, which may be optimal for neither subgroup, practitioners can tailor environmental features, therapeutic protocols, or adaptive interfaces according to individuals' predicted cluster membership. For instance, urban spaces intended to reduce stress might emphasize different sensory modalities depending on whether users exhibit electrodermal-dominant or integrative autonomic profiles. Similarly, emotion-aware systems could adjust stimuli dynamically based on real-time physiological patterns, improving both engagement and wellbeing outcomes. The autoencoder-derived clustering therefore provides a scalable, data-driven foundation for precision approaches in human-environment interaction research.

The simplicity of the architecture, and in particular the small size of the bottleneck, forces the network to encode only the most salient and recurrent features of the physiological data. This constraint prevents the model from memorizing idiosyncratic fluctuations and instead promotes the discovery of stable cross-signal dependencies. Additionally, the inclusion of a multi-head

self-attention mechanism enables the network to learn temporal dynamics that extend beyond pointwise similarities. The attention layer emphasizes segments characterized by strong or rapidly changing autonomic responses, weighting them more heavily during reconstruction and therefore integrating information across time to form a temporally coherent latent representation. As a result, the network learns not only how the individual signals behave in isolation, but also how they jointly evolve in response to emotional or situational changes.

Traditional HRV or GSR analyses often assume linear relationships or require segmentation choices that may discard important temporal information. Likewise, PCA-based methods lack the capacity to incorporate nonlinear interactions between signals. In contrast, the autoencoder learns nonlinear cross-modal relationships directly from the data and preserves both temporal and spectral characteristics through its multi-term loss function. The inclusion of attention further differentiates this model from state-of-the-art deep-learning approaches commonly applied to physiological signals, as it improves sensitivity to transient arousal events and better captures long-range dependencies. Together, these design choices lead to representations that are more robust, interpretable, and suitable for downstream tasks such as clustering or anomaly detection.

Despite these advantages, several limitations must be acknowledged. First, the interpretability of the model is inherently constrained by the unsupervised nature of the approach; while the model learns stable patterns, these patterns do not map directly onto specific emotional states or physiological mechanisms without complementary analysis. Second, the limited sample size of the validation dataset reduces statistical power, making it difficult to detect between-group differences at individual points of the route, even when the overall population trends suggest meaningful changes. Third, while the attention mechanism enhances temporal modeling, it does not explicitly differentiate between cause and correlation, and further validation on controlled datasets would help clarify the specific temporal dependencies learned by the network. Finally, the model remains relatively simple compared to more complex architectures such as transformers or recurrent-convolutional hybrids, which might extract additional layers of temporal abstraction at the cost of interpretability.

Overall, the proposed autoencoder demonstrates that even a compact architecture, when equipped with attention and a physiologically informed loss function, can uncover meaningful patterns of autonomic activity and reveal

structure within heterogeneous populations. These findings highlight the potential of unsupervised deep learning not only for physiological analysis but also for the broader characterization of human responses to environmental and emotional stimuli. Future work should explore larger validation cohorts, integrate behavioral and environmental data more explicitly, and investigate how the learned latent representations can support real-time emotional state estimation and adaptive intervention strategies.

5. Conclusion

This study introduced an unsupervised autoencoder framework for identifying latent autonomic response patterns to emotionally relevant environmental stimuli. By training the network on heterogeneous multimodal physiological datasets and selecting the optimal configuration through clustering performance rather than reconstruction error alone, we derived a compact representation that captured stable cross-signal dependencies and temporal dynamics of autonomic regulation.

Validation on an independent VR-based urban walk dataset demonstrated that the learned representation generalized beyond the training data and revealed two physiologically coherent subgroups. These clusters were not explained by basic sociodemographic variables and exhibited distinct multivariate organizations, different baseline levels, and different patterns of baseline-normalized reactivity. Importantly, several emotional-level effects emerged only after cluster stratification, while some population-level significances disappeared when subgroups were analyzed separately, highlighting the impact of latent heterogeneity on statistical inference.

Methodologically, the results show that reconstruction-based deep learning combined with cluster-oriented model selection can uncover meaningful autonomic structures that remain hidden under conventional aggregate analyses. The findings support the view that emotional-autonomic mappings in environmental contexts are probabilistic and subgroup-dependent rather than universal.

Future work should validate the stability of the identified clusters in larger and more diverse cohorts, investigate longitudinal consistency of cluster membership, and explore the integration of contextual, behavioral, and environmental features within the latent representation. Such extensions may enable real-time subgroup identification and support adaptive, physiology-informed

interventions in domains such as urban design, virtual reality, and affective computing.

Acknowledgments

Funding

This research is funded by MED4PED: Mobility Experience Decision support system for pre-assessing PEDedestrian walkability through on-site and off-site simulation project.

D.D. 104 of 02/02/2022 (call PRIN 2022), funded by the European Union – NextGenerationEU. i.Drive, an interdepartmental laboratory of Politecnico di Milano, supported this research (http://www.idrive.polimi.it/).

Author Contributions

The overall effort on the article and the writing is equally distributed among the authors, whose contribution according to CRediT (Contributor Roles Taxonomy) is described as follows.

Conceptualization: Barbara E.A. Piga, Marco Boffi, Giandomenico Caruso, Lorenzo Mussone;

Data Curation: Chiara Maninetti;

Formal Analysis: Chiara Maninetti, Carlotta Colonna;

Funding Acquisition: Barbara E.A. Piga, Marco Boffi, Lorenzo Mussone, Giandomenico Caruso;

Investigation: Chiara Maninetti, Gabriele Stancato, Nicola Rainisio, Marco Boffi;

Methodology: Chiara Maninetti, Gabriele Stancato, Barbara E.A. Piga, Marco Boffi, Carlotta Colonna, Nicola Rainisio, Lorenzo Mussone, Giandomenico Caruso, Luca Mainardi, Riccardo Barbieri;

Project Administration: Marco Boffi, Barbara E.A. Piga, Lorenzo Mussone, Giandomenico Caruso;

Resources: Lorenzo Mussone, Giandomenico Caruso;

Software: Chiara Maninetti;

Supervision: Marco Boffi, Barbara E.A. Piga, Lorenzo Mussone, Giandomenico Caruso, Luca Mainardi, Riccardo Barbieri;

Visualization: Chiara Maninetti;

Writing — Original Draft:, Chiara Maninetti;

Writing — Review & Editing:, Chiara Maninetti, Barbara E.A. Piga, Marco Boffi, Luca Mainardi, Riccardo Barbieri.

All authors have read and agreed to the published version of the manuscript.

Intellectual property

exp-EIA© (experiential Environmental Impact Assessment) is a patented method nationally and internationally [PATENT 1: Italian Publication Number 102021000017168 — 30/06/2021; International Publication Number WO 2023/275679 A1 — 5.1.2023; PATENT 2: Italian Patent Publication Number 102024000011161 — 16.5.2024
COPYRIGHTS:N. D000020125 — 16.06.2023; N. 130516 — 25.02.2021; N. 123453 — 06.05.2020]

References

[1] J. F. Thayer, R. D. Lane, A model of neurovisceral integration in emotion regulation and dysregulation, Journal of Affective Disorders 61 (2000) 201–216. doi:10.1016/S0165-0327(00)00338-4.

[2] J. F. Thayer, A. L. Hansen, E. Saus-Rose, B. H. Johnsen, Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health, Annals of Behavioral Medicine 37 (2009) 141–153. doi:10.1007/s12160-009-9101-z.

[3] J. B. Holzman, D. J. Bridgett, Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: A meta-analytic review, Neuroscience & Biobehavioral Reviews 74 (2017) 233–255. doi:10.1016/j.neubiorev.2016.12.032.

[4] M. Schneider, C. Rominger, A. R. Schwerdtfeger, Associations between positive affect and heart rate variability: A systematic review, Current Cardiology Reports 27 (2025) 148. doi:10.1007/s11886-025-02299-4.

[5] W. Boucsein, Electrodermal Activity, 2 ed., Springer US, Boston, MA, 2012. doi:10.1007/978-1-4614-1126-0.

[6] T. A. D'Amelio, L. A. Galán, E. A. Maldonado, A. A. Díaz Barquinero, J. Rodríguez Cuello, N. M. Bruno, E. Tagliazucchi, D.-A. Engemann, Emotion recognition systems with electrodermal activity: From affective science to affective computing, Neurocomputing 651 (2025) 130831. doi:10.1016/j.neucom.2025.130831.

[7] K. S. Quigley, L. F. Barrett, Is there consistency and specificity of autonomic changes during emotional episodes? Guidance from the conceptual act theory and psychophysiology, Biological Psychology 98 (2014) 82–94. doi:10.1016/j.biopsycho.2013.12.013.

[8] S. D. Kreibig, Autonomic nervous system activity in emotion: A review, Biological Psychology 84 (2010) 394–421. doi:10.1016/j.biopsycho.2010.03.010.

[9] I. B. Mauss, M. D. Robinson, Measures of emotion: A review, Cognition & Emotion 23 (2009) 209–237. doi:10.1080/02699930802204677.

[10] Task Force Of The European Society Of Cardiology The North A Electrophysiology, Heart rate variability: Standards of measurement, physiological interpretation, and clinical use, Circulation 93 (1996) 1043–1065. doi:10.1161/01.CIR.93.5.1043.

[11] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures, Publication recommendations for electrodermal measurements, Psychophysiology 49 (2012) 1017–1034. doi:10.1111/j.1469-8986.2012.01384.x.

[12] B. M. Appelhans, L. J. Luecken, Heart rate variability as an index of regulated emotional responding, Review of General Psychology 10 (2006) 229–240. doi:10.1037/1089-2680.10.3.229.

[13] S. Laborde, F. Lautenbach, M. S. Allen, The contribution of coping-related variables and heart rate variability to visual search performance under pressure, Physiology & Behavior 139 (2015) 532–540. doi:10.1016/j.physbeh.2014.12.003.

[14] A. Crider, Personality and electrodermal response lability: An interpretation, Applied Psychophysiology and Biofeedback 33 (2008) 141–148. doi:10.1007/s10484-008-9057-y.

[15] A. Hamidovic, K. Van Hedger, S. H. Choi, S. Flowers, M. Wardle, E. Childs, Quantitative meta-analysis of heart rate variability finds reduced parasympathetic cardiac tone in women compared to men during laboratory-based social stress, Neuroscience & Biobehavioral Reviews 114 (2020) 194–200. doi:10.1016/j.neubiorev.2020.04.005.

[16] Z. Wang, Y. Zou, J. Liu, W. Peng, M. Li, Z. Zou, Heart rate variability in mental disorders: An umbrella review of meta-analyses, Translational Psychiatry 15 (2025) 104. doi:10.1038/s41398-025-03339-x.

[17] J. J. McGinley, B. H. Friedman, Autonomic specificity in emotion: The induction method matters, International Journal of Psychophysiology 118 (2017) 48–57. doi:10.1016/j.ijpsycho.2017.06.002.

[18] A. J. Fisher, J. D. Medaglia, B. F. Jeronimus, Lack of group-to-individual generalizability is a threat to human subjects research, Proceedings of the National Academy of Sciences 115 (2018). doi:10.1073/pnas.1711978115.

[19] J. K. Adolf, E. I. Fried, Ergodicity is sufficient but not necessary for group-to-individual generalizability, Proceedings of the National Academy of Sciences 116 (2019) 6541–6541. doi:10.1073/pnas.1818675116.

[20] M. L. Weijs, D. Jonauskaite, R. Reutimann, C. Mohr, B. Lenggenhager, Effects of environmental colours in virtual reality: Physiological arousal affected by lightness and hue, Royal Society Open Science 10 (????) 230432. doi:10.1098/rsos.230432.

[21] G. Spano, A. Theodorou, G. Reese, G. Carrus, G. Sanesi, A. Panno, Virtual nature, psychological and psychophysiological outcomes: A systematic review, Journal of Environmental Psychology 89 (2023) 102044. doi:10.1016/j.jenvp.2023.102044.

[22] A. Birenboim, M. Dijst, F. E. Scheepers, M. P. Poelman, M. Helbich, Wearables and location tracking technologies for mental-state sensing in outdoor environments, The Professional Geographer 71 (2019) 449–461. doi:10.1080/00330124.2018.1547978. arXiv:1910.06137.

[23] T. Mizumoto, H. Ikei, K. Hagiwara, T. Matsubara, F. Higuchi, M. Kobayashi, T. Yamashina, J. Sasaki, N. Yamada, N. Higuchi, K. Haraga, F. Kirihara, E. Okabe, K. Asai, M. Hirotsu, C. Chen, Y. Miyazaki, S. Nakagawa, Physiological adjustment effects of viewing natural environment images on heart rate variability in individuals with depressive and anxiety disorders, Scientific Reports 15 (2025) 16317. doi:10.1038/s41598-025-00681-4.

[24] H.-T. Chiang, Y.-Y. Hsieh, S.-W. Fu, K.-H. Hung, Y. Tsao, S.-Y. Chien, Noise reduction in ECG signals using fully convolutional denoising autoencoders, IEEE Access 7 (2019) 60806–60813. doi:10.1109/ACCESS.2019.2912036.

[25] J. A. Healey, R. W. Picard, Stress recognition in automobile drivers, 2008. doi:10.13026/C2SG6B.

[26] F. R. Ihmig, A. Gogeascoechea, S. Schäfer, J. Lass-Hennemann, T. Michael, Electrocardiogram, skin conductance and respiration from spider-fearful individuals watching spider video clips, ???? doi:10.13026/KC0E-6A32.

[27] K. Sharma, CASE dataset, ????

[28] Hossein Hamidi, "TRINA EmotionDrive: Dataset of emotional responses using six biosignals and questionnaires in driving simulation", ???? doi:10.21227/X51T-JR44.

[29] V. Markova, T. Ganchev, K. Kalinkov, CLAS: A database for cognitive load, affect and stress recognition, in: 2019 International Conference on Biomedical Innovations and Applications (BIA), IEEE, Varna, Bulgaria, 2019, pp. 1–4. doi:10.1109/BIA48344.2019.8967457.

[30] W.-K. Beh, Y.-H. Wu, A.-Y. A. Wu, MAUS: A dataset for mental workload assessment on N-back task using wearable sensor, 2021. doi:10.21227/Q4TD-YD35.

[31] N. Alexendre, F. Kuetche, N. E. Pascal, S. Thierry, EcgScorer: An open source MATLAB toolbox for ECG signal quality assessment, SoftwareX 28 (2024) 101900. doi:10.1016/j.softx.2024.101900.

[32] Weixuan Chen, N. Jaques, S. Taylor, A. Sano, S. Fedor, R. W. Picard, Wavelet-based motion artifact removal for electrodermal activity, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, Milan, 2015, pp. 6223–6226. doi:10.1109/EMBC.2015.7319814.

[33] Unity, Unity Technologies, 2025.

[34] E. Bozgeyikli, A. Raij, S. Katkoori, R. Dubey, Point & teleport locomotion technique for virtual reality, in: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, ACM, Austin Texas USA, 2016, pp. 205–216. doi:10.1145/2967934.2968105.

[35] B. E. A. Piga, N. Rainisio, G. Stancato, M. Boffi, Mapping the In-motion emotional urban experiences: An evidence-based method, Sustainability 15 (2023) 7963. doi:10.3390/su15107963.

[36] M. Perugini, M. Bonnes, A. Aiello, A. P. Ercolani, Il modello circonflesso delle qualità affettive dei luoghi: Sviluppo di uno strumento valutativo italiano. [the affective quality of places' circumplex model: Development of an italian valutative tool.], Testing Psicometria Metodologia 9 (2002) 131–152.

[37] A. Betella, P. F. M. J. Verschure, The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions, PLOS ONE 11 (2016) e0148037. doi:10.1371/journal.pone.0148037.

[38] C. Maninetti, R. Laureanti, B. E. A. Piga, N. Rainisio, M. Boffi, G. Stancato, L. Mainardi, R. Barbieri, An integrative psychophysiological approach for emotional assessment in outdoor urban settings, Frontiers in Psychology 16 (2025). doi:10.3389/fpsyg.2025.1690851.

[39] R. Barbieri, E. C. Matten, A. A. Alabi, E. N. Brown, A point-process model of human heartbeat intervals: New definitions of heart rate and heart rate variability, American Journal of Physiology-Heart and Circulatory Physiology 288 (2005) H424–H435. doi:10.1152/ajpheart.00482.2003.

[40] B. E. A. Piga, Experiential-walk: Experiencing and representing the city for urban design purposes, in: B. E. A. Piga, D. Siret, J.-P. Thibaud (Eds.), Experiential Walks for Urban Design, Springer International Publishing, Cham, 2021, pp. 187–206. doi:10.1007/978-3-030-76694-8_11.