EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Designing VR for Chronic Pain: Visualizing Autonomic States Through VR Biofeedback to Support Mindfulness

Sara Khalilipicha*, Diane Gromala, Armin Froozanfar, Chris Shaw, and Patti Derbyshire

School of Interactive Arts and Technology, Simon Fraser University, 250–13450 102 Avenue, Surrey, BC, V3T 0A3, Canada

Abstract

Chronic pain affects physical functioning and psychological well-being, reducing quality of life. This paper presents the *Virtual Meditative Walk*, a closed-loop VR biofeedback environment that supports mindfulness-based chronic pain management through indirect, in-world visualization of autonomic state. To improve adaptation, we integrate a machine learning-based physiological inference pipeline using Electrodermal Activity and Heart Rate Variability. Using the combined WESAD and StressID dataset, the final Extra Trees model improved over the original VMW feedback method in the cross-sample binary setting, reaching 91.77% binary cross-sample accuracy and 86.05% multiclass cross-sample accuracy. Leave-one-subject-out validation was lower, showing that subject-independent stress inference remains challenging. These findings support machine learning as a promising but limited method for adaptive VR biofeedback in chronic pain contexts.

Received on 31 January 2026; accepted on 26 August 2026; published on 11 September 2026

Keywords: Virtual reality biofeedback, chronic pain, Electrodermal Activity, Heart Rate Variability, stress classification, machine learning, mindfulness

doi:10.4108/eetpht.11.11776

1. Introduction

Chronic pain affects approximately 1 in 5 Canadians [1], with similar rates in industrialized countries. It is a significant public health issue with broad physical, psychological, social, and economic consequences [2, 3]. Despite its prevalence, chronic pain has no definitive biomarker for diagnosis and no known cure has been identified [4, 5]. Current management strategies often rely on symptomatic treatment, including opioids, which can provide short-term relief but also carry risks of dependency and misuse [4, 6, 7]. These limitations underscore the need for non-pharmacological and scalable health technologies for chronic pain support [8].

Virtual Reality (VR) has emerged as a promising nonpharmacological tool for chronic pain because immersive environments can mediate pain perception [9–11], support engagement with therapeutic practices [12], and address physical, emotional, intellectual, and psychological dimensions of pain [10, 13]. VR has also been explored for stress reduction, physical activity, and body-schema realignment [10, 12, 14]. By integrating biofeedback, VR systems can respond to patients' changing psychophysiological states and support self-regulation by making internal bodily processes more visible [15, 16].

Most biofeedback systems present physiological change through explicit displays such as graphs, numerical scores, or thresholds [17]. These representations can be useful, but they may also encourage users to evaluate or control their performance. This can conflict with mindfulness-based practice, where the aim is to observe bodily experience with less judgment and less reactivity [18]. For this reason, indirect biofeedback embedded in the virtual environment may be especially relevant for VR systems that support mindfulness-based chronic pain management.

Building on outcomes from several small and two longitudinal studies [19–21], we developed an

*Corresponding author. Email: khalilip@sfu.ca

approach that leverages biofeedback data to make patients' changing biopsychological states visible. This approach centers on using physiological signals, such as electrodermal activity, to infer patients' arousal states in real time. These signals serve as a bridge between patients' inner experiences and the external representation of those experiences.

This approach is integrated in the *Virtual Meditative Walk* (VMW) [21], an immersive VR system designed as a foggy virtual forest as depicted in Fig. 1, to assist CP patients. VMW directs patients' attention inward, guiding them to experientially learn and practice Mindfulness-Based Stress Reduction techniques. While meditation apps have become more accessible, VMW is unique in its emphasis on users' attention. For instance, streaming biofeedback alters the fog, visually reflecting shifts in arousal during mindfulness exercises. As stress decreases, the fog lifts (Fig. 3), helping patients stay attuned to interoceptive sensations without relying on charts. After sessions, a graph of their biofeedback data is provided, supporting the internalization of mindfulness techniques and encouraging long-term practice.

Although physiological stress classification using EDA and HRV has been widely studied [22, 23], its role in VMW is different from a standalone affect-recognition task. Here, the purpose of classification is to support stable, interpretable environmental adaptation during mindfulness practice. The original VMW feedback mechanism used a simple skin conductance slope-based rule [20], which was easy to implement but limited to a single signal trend. This made it difficult to capture more complex relationships between physiological signals and stress-related states. A machine learning approach may address this limitation by combining EDA and HRV features into a broader estimate of autonomic state, while still preserving the indirect and non-performance-based feedback style of VMW.

The goal of this study is to explain and evaluate VMW as a closed-loop VR biofeedback environment for Mindfulness-Based Stress Reduction (MBSR), with a specific focus on improving the physiological inference method that drives environmental adaptation. We focus on how VMW uses indirect biofeedback as a design strategy, mapping autonomic state onto subtle changes in the virtual environment, such as fog density, to support interoceptive awareness without charts, scores, or performance pressure. We then evaluate whether a machine learning-based physiological inference pipeline using EDA and HRV can provide a more reliable control signal than the original skin conductance slope-based method. This evaluation is intended as a proof-of-concept for adaptive VR biofeedback rather than as evidence of clinical effectiveness.

This paper makes three contributions. First, it positions VMW as a closed-loop VR biofeedback environment for chronic pain management, with a focus on indirect environmental feedback rather than explicit physiological displays. Second, it evaluates a machine learning-based physiological inference pipeline using EDA and HRV features from a combined WESAD and StressID dataset. Third, it compares this approach with the original skin conductance slope-based feedback rule and reports both model performance and real-time feasibility for adaptive VR use.

This paper is organized as follows: Section 2 reviews related work on VR biofeedback for chronic pain and machine learning-based physiological state inference. Section 3 describes the design rationale of VMW, including the environment, session structure, and the closed-loop mapping of physiology to in-world feedback. Section 4 details the technical pipeline used to support that design, including signal preprocessing, feature extraction, model development, and real-time integration. Section 5 reports the model evaluation results, compares the machine learning approach with VMW's original slope-based feedback mechanism, and examines model interpretation and inference latency. Section 6 discusses implications for VR biofeedback design. Section 7 presents limitations. Finally, Section 8 concludes with key contributions and directions for future work.

2. Related Studies

2.1. VR and Biofeedback for Chronic Pain

A growing body of research supports the use of VR-based interventions in pain management. Earlier studies showed that immersive virtual environments can reduce pain perception through distraction, cognitive engagement, and neuromodulation-related mechanisms [27–29]. Other work extended this direction by using VR for rehabilitation, graded exposure, and physical activity support [30, 31]. More recent reviews describe VR as a promising approach for acute and chronic pain, but also emphasize that existing interventions remain heterogeneous in design, outcome measures, and theoretical grounding [24]. This makes it important to clarify not only whether VR is used, but how it supports pain-related self-regulation.

Biofeedback extends VR by allowing the environment to respond to physiological state. Chittaro et al. showed that immersive VR combined with multisensory biofeedback may improve engagement and personalization for patients with fibromyalgia [32], while broader reviews of biofeedback for chronic pain show its relevance for autonomic regulation and emotional distress [15]. However, many biofeedback systems present physiological change through explicit displays such as graphs, numerical values, or thresholds.

These displays can be useful, but they may also encourage performance monitoring. For mindfulness-based chronic pain support, this is a design concern because the goal is not to make users compete with their physiology, but to help them observe bodily change with less judgment and reactivity.

2.2. Machine Learning for Physiological State Inference

Machine learning has increasingly been used to estimate stress and affective state from physiological signals such as Electrodermal Activity (EDA), Heart Rate Variability (HRV), respiration, and EEG [22, 23]. In VR and biofeedback contexts, these models can support adaptive systems by translating physiological signals into state estimates that guide environmental change. Recent work has also explored AI-enhanced VR, neurofeedback, and motion-adaptive biofeedback systems for pain and rehabilitation, suggesting that adaptive feedback may support more personalized therapeutic experiences [33–36].

At the same time, wearable stress detection remains difficult. Recent reviews show that EDA and HRV are widely used for physiological stress inference, but model generalization, labeling quality, preprocessing choices, class imbalance, and validation on unseen users remain major challenges [22, 23]. These issues are especially important for VMW because the classifier is not used as a standalone affect-recognition task. It is used as part of a closed-loop biofeedback environment, where prediction errors can directly affect the user experience.

Preliminary studies of the *Virtual Meditative Walk* showed that VR biofeedback can support chronic pain management by combining guided mindfulness practice with visual and sonic environmental feedback [19–21]. However, the original VMW feedback mechanism relied on a simple skin conductance slope-based rule. This rule was easy to implement and interpret, but it used only one signal trend. The present work extends VMW by evaluating whether a machine learning-based inference pipeline using EDA and HRV can provide a richer and more stable control signal for indirect environmental feedback. This positions VMW at the intersection of two incomplete literatures: VR-based mindfulness systems need more rigorous adaptive designs, and physiological stress classifiers need to be evaluated as situated components of real biofeedback systems rather than as standalone prediction models.

3. Virtual Meditative Walk

The *Virtual Meditative Walk* is a VR system built for chronic pain management using MBSR. It is not intended primarily as a distraction tool. It is a biofeedback environment that changes in real time based on the user's physiology. The system is designed to help users notice stress-related shifts and practice self-regulation by making autonomic nervous system activity visible through the surrounding environment.

VMW uses indirect biofeedback. It does not show numbers, charts, or explicit instructions like "you are stressed." Feedback is embedded in the environment to reduce cognitive load and avoid performance pressure. As users notice gradual changes tied to attention and bodily state, they can start to recognize their own regulation patterns. Over repeated sessions, the goal is skill learning and internalization, not a single-session solution. VMW is not designed to erase pain or stress in a single session. Rather, it is designed to support repeated practice and gradual skill development.

VMW runs as a closed-loop system. Physiological signals are measured continuously, processed on an external pipeline, and mapped onto the virtual environment. The system avoids explicit success criteria to reduce maladaptive control strategies and keep the experience aligned with MBSR. Instead of showing numbers or asking users to "win" a task, VMW externalizes internal regulation so users can observe how attention and bodily state move together. The full closed-loop workflow is provided in Appendix Fig. A.1.

3.1. Environment Design

VMW places the user in a quiet forest and moves them forward slowly along a fixed path. The environment stays visually consistent and intentionally simple. There are no obstacles, choices, or task interactions. This design reduces decision-making demands and limits attentional shifts away from interoception.

The environment includes fog, ambient lighting, and subtle audio. These are not fixed. They shift with physiological biofeedback. A natural setting is used because it can support relaxation without demanding active engagement. VMW avoids performance framing entirely. No scores, no progress bars, no rewards. The experience is built around continuity and gradual change, which fits MBSR's focus on non-judgmental awareness instead of outcomes.

Fog was selected as the primary feedback channel because it can change gradually without interrupting the user's movement through the environment. Unlike a score, alert, or graph, fog can make physiological change visible while remaining part of the virtual world. Lighting and ambient sound were used as secondary feedback channels because they are less obstructive than fog and can reinforce sustained relaxation without requiring explicit attention. These design choices were guided by the goal of keeping feedback ambient, continuous, and non-evaluative.

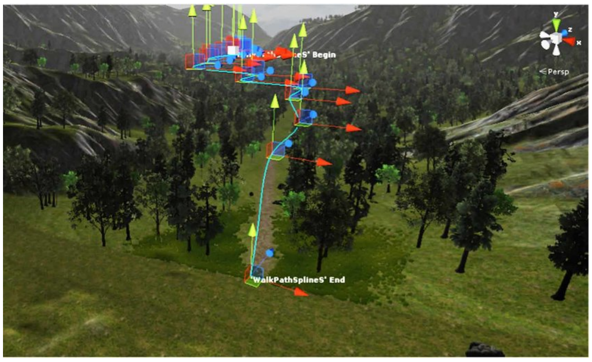


Figure 1. Trail design in the *Virtual Meditative Walk* VR environment [19]

3.2. Session Structure

Each session has three stages: baseline assessment, adaptive meditation, and termination. The total time is about 21 minutes. This length is long enough to stabilize signals but short enough to stay realistic for repeated use. At session start, user metadata is recorded (participant ID, height, narration day). Camera height is adjusted to match the user's height so the first-person perspective stays consistent across sessions. Guided meditation runs through the session. Narration varies across days to reduce habituation while keeping pacing and instruction style consistent.

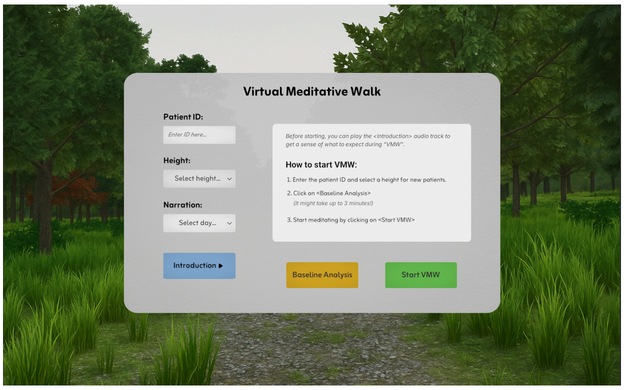


Figure 2. *Virtual Meditative Walk* session start interface

3.3. Baseline Physiological Assessment

The session begins with a two-minute baseline. During baseline, physiology is recorded while the environment stays visually static. Electrodermal Activity and heart rate are collected via wearable sensors and sent to an external processing pipeline for real-time analysis. This phase functions as the model initialization stage. Baseline also serves a functional role in user adaptation. The initial minutes allow users to adjust to VR setup, reducing novelty-related arousal that could otherwise affect early physiological responses.

3.4. Adaptive State Inference

During the adaptive meditation phase, the system continuously estimates the user's autonomic arousal state. At each update interval, the inference model assigns the current physiological window to one of three discrete classes: stress, relaxation/meditation, or other. The other class represents all physiological patterns that do not fall into either the stress or relaxation states. Therefore, it is not interpreted as a negative or positive condition, but as an intermediate or non-specific state that reflects normal physiological variability.

The inference output is used directly by the environment control policy. In the current implementation, the inferred class is mapped to three fog density levels: high fog density for stress, medium fog density for other, and low fog density for relaxation/meditation. This three-level mapping was chosen to keep the feedback legible while avoiding overly fine-grained changes that could be difficult for users to interpret during meditation. Fog updates are bounded and applied at the same fixed interval as the inference updates to preserve temporal consistency between physiological estimation and perceptual feedback.

To emphasize stability rather than momentary fluctuations, the system applies additional rules based on consecutive classifications. These rules are design safeguards rather than clinically validated thresholds. If the user remains in the relaxation/meditation state for three consecutive updates, the system further reduces fog density until the environment reaches the clearest allowed level. If the user remains in the relaxation/meditation state for five consecutive updates, the system introduces additional environmental reinforcement by increasing ambient lighting and adding natural ambient sounds. These changes are only triggered by sustained relaxation and are not activated by single-window improvements.

To prevent extended negative feedback, the inference-driven control policy limits prolonged high-fog conditions. Specifically, the system does not allow more than five consecutive stress-classified updates to maintain high fog density. After this limit is reached, fog density is reduced toward medium levels even if subsequent updates remain classified as stress. This rule was included to prevent persistent visual obstruction and reduce the risk that users interpret the feedback as punishment or become stuck in a perceived failure state. The exact number of consecutive updates should therefore be understood as an implementation choice that requires further empirical tuning with chronic pain users.

The inference process emphasizes temporal consistency and robustness rather than moment-to-moment reactivity. This reduces sensitivity to transient artifacts and supports environmental modulation that users can meaningfully interpret.

Figure 3. As patients approach an inferred meditative state, the fog begins to dissipate (left to right), and sounds become more audible and spatial

Figure 4. *Virtual Meditative Walk* interface and post-session biofeedback visualization

3.5. Biofeedback Visualization

Following completion of the VR meditation session, users are presented with a biofeedback visualization scene that summarizes the session retrospectively. This scene is designed for review and reflection rather than real-time regulation. It separates experiential practice from analytical feedback to avoid introducing evaluation during the meditation itself.

The upper portion of the scene presents a temporal structure of the meditation session, aligned with the guided narration. Key phases of the practice, such as introduction, body sensation, attention to breath, dealing with distraction, and pain-related prompts, are displayed along a timeline. Representative screenshots captured during the VR session are shown beneath this timeline to provide visual reference points corresponding to different moments in the experience.

The lower portion of the scene displays a time-series plot of physiological signals, including EDA and heart rate, plotted over the duration of the session. The EDA signal is shown relative to the individualized baseline, which is indicated as a reference line. Color coding is used to distinguish periods classified a s less stressed than baseline, more stressed than baseline, or unchanged. Heart rate is overlaid to provide additional physiological context but is not used as the primary feedback variable.

This visualization is intended to support post-session interpretation rather than in-session control. By aligning physiological trends with the structure of the meditation and visual snapshots of the environment, the scene allows users and researchers to examine how autonomic responses evolved over time without framing these changes as success or failure.

4. Methodology

This section presents the physiological state inference pipeline used to support real-time biofeedback in the *Virtual Meditative Walk* system. The pipeline combines physiological signal preprocessing, feature extraction, supervised machine learning, model evaluation, and real-time integration. The goal is to estimate stress-related physiological states from Electrodermal Activity (EDA) and Heart Rate Variability (HRV), and to use these estimates to guide adaptive environmental feedback in VMW. Fig. 5 presents the block diagram of the stress inference and assessment framework.

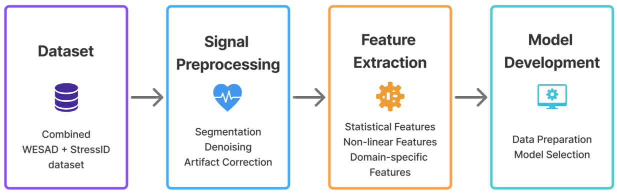


Figure 5. Block diagram of the stress inference and assessment framework

4.1. Dataset

The pipeline was evaluated using a combined dataset from WESAD and StressID. WESAD (Wearable Stress and Affect Detection) is a publicly available multimodal dataset designed to support research in affective computing and stress detection [37]. The dataset was collected in a controlled laboratory setting and includes physiological and motion data recorded from 15 subjects. It is labeled according to evoked emotional state, including baseline, stress, amusement, and meditation/relaxation conditions.

StressID is a multimodal dataset designed specifically for stress identification [38]. I t i ncludes synchronized facial video, audio, and physiological recordings collected from 65 participants across 11 tasks. The physiological signals include electrocardiography (ECG), electrodermal activity (EDA), and respiration, while the tasks include emotional stimuli, cognitive exercises, and public speaking scenarios. StressID also includes participant ratings of relaxation, stress, arousal, and valence, making it useful for both stress classification and broader affective state analysis.

In this study, only the physiological signals relevant to the VMW pipeline were used. Specifically, EDA and ECG-derived HRV features were extracted so the model inputs matched the biosignals available for real-time VR biofeedback. Combining WESAD and StressID increased the number of subjects and physiological windows available for training and evaluation, allowing the model to be tested across a broader range of stress responses than WESAD alone.

After preprocessing and windowing, the combined dataset included 2,223 feature rows: 1,406 from WESAD and 817 from StressID. In total, the dataset included 79 subject groups, with 15 WESAD subjects and 64 StressID subjects. Although StressID includes 65 participants, 64 remained after preprocessing because one participant had no valid 60-second window.

Two target structures were used. The binary target classified e ach w indow a s s tress o r n on-stress. The multiclass target classified e ach w indow a s stress, relaxation, or other. The other class was used for physiological windows that did not clearly represent either stress or relaxation and was treated as an intermediate or non-specific autonomic state.

The three-class structure also aligns with the VMW biofeedback design. In the adaptive environment, stress triggers negative feedback by increasing fog density for the current window. Relaxation triggers positive feedback by decreasing fog density to a low level. The other class triggers neutral feedback by keeping fog density at a medium level. This mapping reflects the goal of VMW as an MBSR-based system: to support awareness of stress and relaxation shifts without treating all other physiological variation as meaningful for the meditation task.

Table 1. Composition of the combined WESAD and StressID dataset after preprocessing

Dataset	Subjects	Feature rows	Percentage
WESAD	15	1,406	63.25%
StressID	64	817	36.75%
Total	79	2,223	100%

4.2. Signal Preprocessing

Signal preprocessing ensures fidelity and reliability of physiological data. It involves three main phases: segmentation, denoising, and artifact correction, each applied to both EDA and HRV signals to enhance interpretability and model performance. Because WESAD and StressID were collected with different recording settings, signals were first harmonized before feature extraction. ECG was used to derive RR intervals for HRV analysis, while EDA was used to extract tonic and phasic features related to sympathetic arousal.

Table 2. Target label distributions in the combined dataset

Task	Class	Samples	Percentage
Binary	Non-stress	1,720	77.37%
Binary	Stress	503	22.63%
Multiclass	Other	925	41.61%
Multiclass	Relaxation	916	41.21%
Multiclass	Stress	382	17.18%

Segmentation. Research indicates that windows ranging from 30 seconds to 5 minutes effectively capture physiological markers of stress while enabling timely interventions and accurate classification [39–42]. The raw physiological signals were divided into fixed 60-second overlapping segments with 30 seconds of overlap. This segmentation strategy enables a balance between capturing sufficient signal variation and maintaining temporal responsiveness for adaptive VR feedback.

Denoising. The majority of relevant spectral energy in EDA lies below 1 Hz, particularly within the 0.045–0.25 Hz range [43]. To remove high-frequency noise while preserving tonic and phasic components, EDA signals were low-pass filtered at approximately 1 Hz [43–45]. ECG signals were bandpass filtered between approximately 0.5 Hz and 40 Hz to reduce baseline wander, muscle noise, and electrical interference while preserving the frequency range needed for R-peak detection and HRV extraction [46].

Artifact Correction. Artifact correction was performed to improve signal reliability. EDA artifacts were reduced using moving average smoothing to limit abrupt irregularities [47]. For HRV, RR intervals outside the physiologically plausible range of 0.3–2.0 seconds were replaced with the mean of valid intervals [48]. Corrected RR intervals were then interpolated and resampled to 16 Hz for time-aligned and frequency-domain analyses [49].

4.3. Feature Extraction

Features were extracted to quantify stress-related physiological variation using statistical, nonlinear, and domain-specific measures from both EDA and HRV signals. EDA features were included because they reflect sympathetic arousal, while HRV features were included because they provide complementary information about autonomic regulation [43, 56]. Together, these feature groups allow the model to estimate stress-related physiological state from both arousal and cardiac variability rather than relying on a single signal trend. The full list of model input features is provided in Appendix Table B.1.

Statistical features captured central tendency, variability, distributional spread, and signal shape, including measures such as mean, median, standard deviation, percentiles, skewness, and kurtosis [50]. Nonlinear features captured complexity and irregularity in physiological dynamics, including approximate entropy, correlation dimension, and Poincare plot measures such as SD1 and SD2 [51–53]. Domain-specific HRV features included time-domain and frequency-domain measures such as RMSSD, SDNN, AVNN, VLF, LF, HF, and LF/HF ratio [54–57]. EDA-specific features included tonic skin conductance level and phasic skin conductance response measures, including SCR amplitude, frequency, rise time, and area under the curve [58].

4.4. Machine Learning Model Development

Data Preparation. To address potential class imbalance in the combined WESAD and StressID dataset, balancing strategies were applied during model training. These included class weighting and sample weighting so that minority classes, particularly stress, were not ignored by the classifier. Missing feature values were handled using median imputation. For models that are sensitive to feature scale, including Logistic Regression and Support Vector Machine (SVM), standardization was also applied within the training pipeline.

Model performance was evaluated using two validation strategies. First, cross-sample validation was conducted using stratified 5-fold cross-validation with shuffled splits. In this setting, the class distribution was preserved across folds, but samples from the same subject could appear in both training and testing sets. This validation was used as an initial estimate of performance across the combined WESAD and StressID feature space.

Second, leave-one-subject-out cross-subject validation was conducted using subject ID as the grouping variable. In each fold, all windows from one subject were held out for testing, and the model was trained on the remaining subjects. This strategy provides a stricter evaluation because the model must classify physiological data from a subject it has not seen during training. Folds were excluded only when the training split did not contain all target classes.

Evaluation Metrics. Model performance was assessed using accuracy, balanced accuracy, precision, recall, F1-score, ROC-AUC, confusion matrices, precision-recall curves, and calibration plots. Balanced accuracy was included because both the binary and multiclass tasks were imbalanced. For binary classification, stress was treated as the positive class. For multiclass classification, ROC-AUC was computed using a one-vs-rest strategy with macro averaging. All metrics were computed within each validation fold and then averaged across folds. Cross-sample metrics were averaged across five stratified folds, while leave-one-subject-out metrics were averaged across valid subject-held-out folds.

Model Selection and Optimization. Multiple models were trained and evaluated: Logistic Regression, Support Vector Machine (SVM), Random Forest, Extra Trees, and XGBoost. These models were selected to compare linear, kernel-based, bagging-based, and boosting-based approaches using the same feature set and validation procedures. Class balancing was applied through class weights or balanced sample weights, depending on the model.

After the initial benchmark, Extra Trees was selected for final evaluation because it provided the strongest overall performance while remaining computationally practical for real-time inference. A targeted grid search was then conducted for Extra Trees. The final binary and multiclass configurations are summarized in Table 3.

4.5. Real-Time Integration into VMW System

The final Extra Trees model was prepared for integration into the *Virtual Meditative Walk* system as the physiological inference component for adaptive biofeedback. Incoming EDA and ECG-derived HRV data follow the same preprocessing and feature extraction steps used during model development. The resulting feature vector is passed to the trained classifier, which estimates the current physiological state as stress, relaxation, or other.

The predicted state is then mapped to the environmental feedback policy described in Section 3. Stress predictions increase fog density, relaxation predictions reduce fog density, and other predictions maintain a medium fog level. Sustained relaxation predictions can trigger additional environmental reinforcement, such as brighter ambient lighting and natural sounds. Prolonged stress predictions are bounded to prevent persistent high-fog feedback. To examine real-time feasibility, model-level latency was measured using the final selected multiclass Extra Trees pipeline. The latency results are reported in Section 5.

Table 3. Model configurations and class-balancing strategies

Model	Configuration	Class-balancing strategy
Logistic Regression	Maximum iterations = 3000	`class_weight = balanced`
SVM	RBF kernel; $C = 10$; γ = scale; probability estimation enabled	`class_weight = balanced`
Random Forest	500 estimators	`class_weight = balanced`
XGBoost	300 estimators; learning rate = 0.05; maximum depth = 6; subsample = 0.8; column subsample = 0.8; histogram-based tree method	Balanced sample weights
Extra Trees, binary	300 estimators; Gini criterion; maximum depth = 30; maximum features = 0.5; minimum samples per leaf = 3; minimum samples split = 8	`class_weight = balanced`
Extra Trees, multiclass	300 estimators; entropy criterion; maximum depth = 30; maximum features = 0.5; minimum samples per leaf = 1; minimum samples split = 2	`class_weight = balanced`

5. Results

This section reports the evaluation of the physiological state inference pipeline using the combined WESAD and StressID dataset described in Section 4. The analysis focuses on four questions. First, we examine model selection across candidate classifiers. Second, we evaluate the final Extra Trees model under cross-sample and leave-one-subject-out validation. Third, we compare the machine learning approach with the original skin conductance slope-based feedback rule used in VMW. Finally, we examine model interpretation and real-time inference latency.

5.1. Model Selection

Initial benchmarking compared Logistic Regression, Support Vector Machine, Random Forest, Extra Trees, and XGBoost using the same EDA and HRV feature set. Table 4 summarizes the cross-sample benchmark results for the binary and multiclass tasks.

Table 4. Initial cross-sample benchmark across candidate classifiers. Values are averaged across five stratified folds

Task	Model	Acc.	Bal. Acc.	F1	ROC-AUC
Binary	Logistic Regression	0.6356	0.7463	0.5411	0.8596
Binary	SVM	0.7958	0.8209	0.6578	0.9111
Binary	Random Forest	0.9001	0.8003	0.7358	0.9490
Binary	XGBoost	0.9082	0.8520	0.7864	0.9485
Binary	Extra Trees	0.9222	0.8385	0.7990	0.9675
Multiclass	Logistic Regression	0.6230	0.6753	0.6159	0.8289
Multiclass	SVM	0.7202	0.7465	0.7200	0.8822
Multiclass	Random Forest	0.8201	0.8062	0.8169	0.9404
Multiclass	XGBoost	0.8273	0.8237	0.8281	0.9430
Multiclass	Extra Trees	0.8538	0.8441	0.8547	0.9544

Extra Trees showed the strongest overall performance in the initial benchmark. For the binary task, it achieved the highest accuracy, F1-score, and ROC-AUC, although XGBoost produced slightly higher balanced accuracy. For the multiclass task, Extra Trees achieved the highest value across all reported benchmark metrics, including accuracy, balanced accuracy, F1-score, and ROC-AUC.

The leave-one-subject-out benchmark showed a more difficult generalization problem. In the binary task, Extra Trees achieved the strongest initial cross-subject performance among the candidate models, with 78.61% accuracy and 71.13% balanced accuracy. In the multiclass task, Extra Trees remained competitive, with 61.55% accuracy and 55.85% balanced accuracy, although performance differences among the tree-based models were smaller. Based on this overall pattern, Extra Trees was selected for final tuning and evaluation.

5.2. Binary Stress Classification

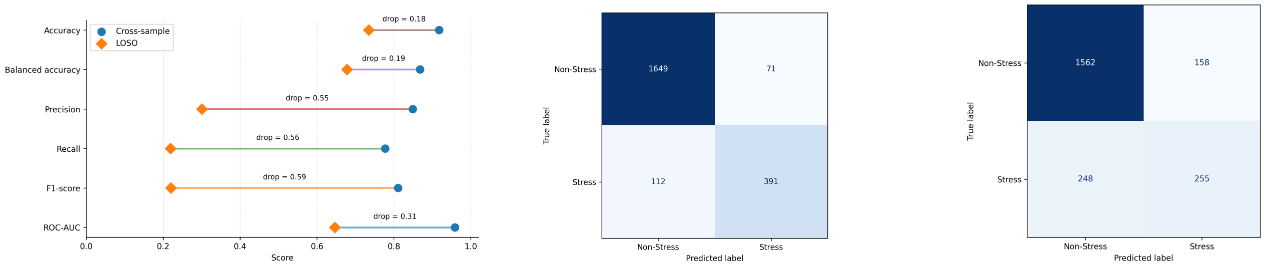
Table 5 summarizes the final binary classification results. Under cross-sample validation, the Extra Trees model achieved 91.77% accuracy, 86.80% balanced accuracy, 84.89% precision, 77.73% recall, 81.10% F1-score, and 95.90% ROC-AUC. These results indicate strong discrimination between stress and non-stress windows when training and test windows are drawn from the same combined data distribution.

Leave-one-subject-out validation produced lower fold-averaged performance, with 73.50% accuracy, 67.84% balanced accuracy, 30.11% precision, 21.93% recall, 22.04% F1-score, and 64.65% ROC-AUC. These results show a substantial reduction in performance under the unseen-subject validation setting.

Table 5. Final Extra Trees performance for binary stress classification. Values are averaged across validation folds

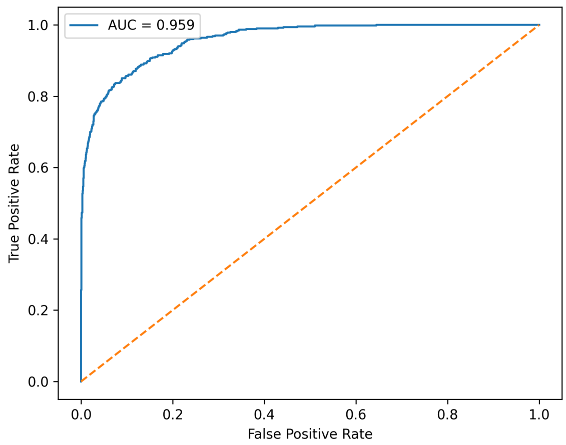
Validation	Acc.	Bal. Acc.	Precision	Recall	F1	ROC-AUC
Cross-sample	0.9177	0.8680	0.8489	0.7773	0.8110	0.9590
LOSO	0.7350	0.6784	0.3011	0.2193	0.2204	0.6465

The pooled cross-sample confusion matrix showed 1,649 correctly classified non-stress windows and 391 correctly classified stress windows. It also showed 71 false stress predictions and 112 missed stress windows. In the pooled leave-one-subject-out confusion matrix, the model correctly identified 255 stress windows but missed 248 stress windows. The pooled confusion matrices reflect the same performance pattern shown in Table 5. The binary ROC curves in Fig. 7 are plotted from pooled fold predictions.

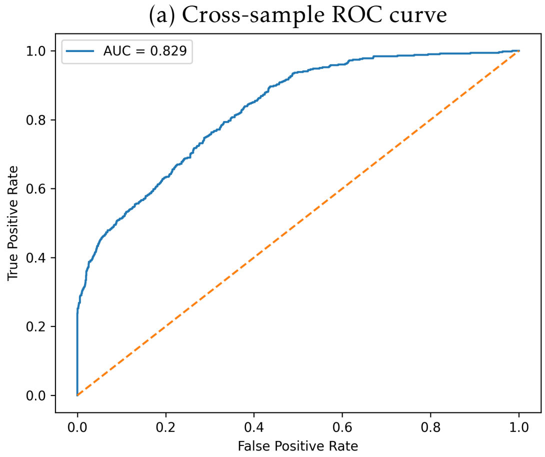


(a) Cross-sample vs. LOSO performance. **(b)** Cross-sample confusion matrix. **(c)** LOSO confusion matrix.

Figure 6. Binary stress classification results. The left panel compares cross-sample and leave-one-subject-out performance, while the middle and right panels show the corresponding confusion matrices



(a) Cross-sample ROC curve



(b) LOSO ROC curve

Figure 7. ROC curves for binary stress classification under cross-sample and leave-one-subject-out evaluation

5.3. Comparison with the Original VMW Biofeedback Mechanism

The original *Virtual Meditative Walk* feedback mechanism used a simple skin conductance trend to infer whether the user was moving toward or away from relaxation. In this slope-based rule, an increase in SCL was interpreted as stress, while a decrease or no increase was interpreted as non-stress. To compare the machine learning pipeline with this baseline, the SCL slope-based rule was evaluated on the same combined binary target. The rule-based model achieved 62.35% accuracy, 57.87% balanced accuracy, 29.98% precision, 49.70% recall, and 37.40% F1-score for stress detection. The confusion matrix showed 1,136 correctly classified non-stress windows, 584 false stress predictions, 250 correctly classified stress windows, and 253 missed stress windows.

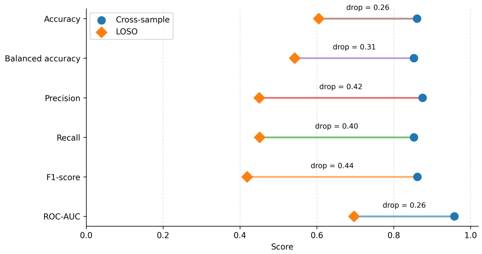
Compared with this baseline, the final Extra Trees model provided a clear improvement in the cross-sample setting, achieving 91.77% accuracy and 81.10% F1-score. The comparison shows that the Extra Trees model improved over the SCL slope-based rule in the cross-sample setting, while the leave-one-subject-out results remained lower.

Table 6. Comparison between the final Extra Trees binary classifier and the original SCL slope-based rule

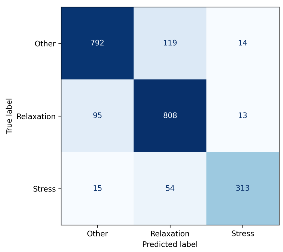
Model	Acc.	Bal. Acc.	Precision	Recall	F1
SCL slope rule	0.6235	0.5787	0.2998	0.4970	0.3740
Extra Trees, cross-sample	0.9177	0.8680	0.8489	0.7773	0.8110
Extra Trees, LOSO	0.7350	0.6784	0.3011	0.2193	0.2204

5.4. Multiclass Physiological State Classification

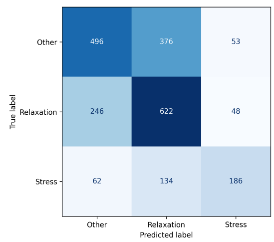
Table 7 summarizes the final multiclass results. Under cross-sample validation, the Extra Trees model achieved 86.05% accuracy, 85.26% balanced accuracy, 87.52% precision, 85.26% recall, 86.16% F1-score, and 95.79% ROC-AUC. These results suggest that the model can distinguish among other, relaxation, and stress states when evaluated on held-out windows from the same combined feature distribution. Leave-one-subject-out validation again produced lower fold-averaged performance, with 60.50% accuracy, 54.26% balanced accuracy, 45.03% precision, 45.10% recall, 41.85% F1-score, and 69.68% ROC-AUC. As in the binary task, leave-one-subject-out validation produced lower performance than cross-sample validation.



(a) Cross-sample vs. LOSO performance



(b) Cross-sample confusion matrix



(c) LOSO confusion matrix

Figure 8. Multiclass physiological state classification results. The left panel compares cross-sample and leave-one-subject-out performance, while the middle and right panels show the corresponding confusion matrices

The pooled multiclass cross-sample confusion matrix showed strong classification across all three classes. The model correctly classified 792 other windows, 808 relaxation windows, and 313 stress windows. The most common errors were confusion between other and relaxation, and between stress and relaxation. In the pooled leave-one-subject-out confusion matrix, the model correctly classified 496 other windows, 622 relaxation windows, and 186 stress windows. In the leave-one-subject-out setting, stress windows were more often confused with relaxation than in the cross-sample setting. The multiclass ROC curves in Fig. 9 are plotted from pooled fold predictions.

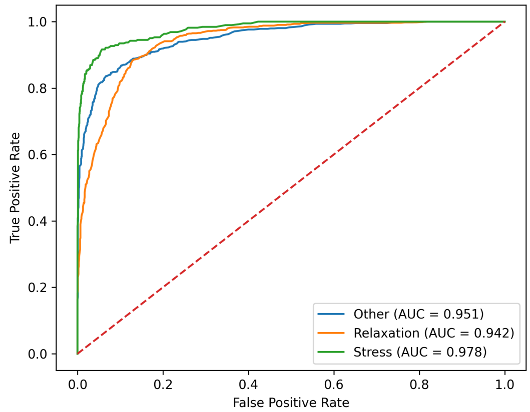
Table 7. Final Extra Trees performance for multiclass physiological state classification. Values are averaged across validation folds

Validation	Acc.	Bal. Acc.	Precision	Recall	F1	ROC-AUC
Cross-sample	0.8605	0.8526	0.8752	0.8526	0.8616	0.9579
LOSO	0.6050	0.5426	0.4503	0.4510	0.4185	0.6968

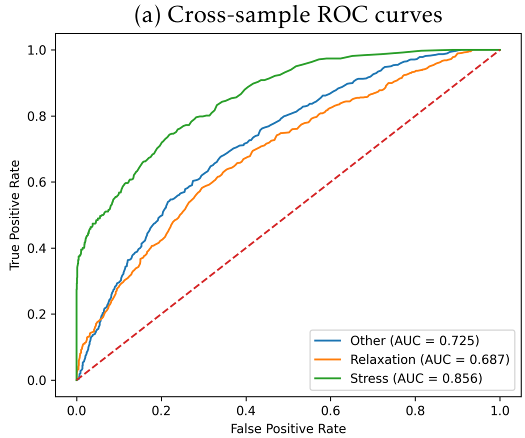
5.5. Model Interpretation

Model interpretation was used to examine which physiological features contributed most strongly to the final Extra Trees predictions. The feature importance results show that both EDA and HRV features were relevant. In the binary task, SCR frequency was the strongest feature, followed by AVNN, HRV mean, HRV harmonic mean, and HRV percentile features. In the multiclass task, SCR frequency again appeared as an important feature, alongside HRV median, HRV percentile features, AVNN, SCL, and EDA central tendency measures.

SHAP feature importance analysis further supported this pattern. The binary model relied strongly on EDA-derived sympathetic activity and HRV-derived cardiac variability features, while the multiclass model showed a broader distribution of feature contributions across both signal groups. Together, the impurity-based and



(a) Cross-sample ROC curves



(b) LOSO ROC curves

Figure 9. ROC curves for multiclass physiological state classification under cross-sample and leave-one-subject-out evaluation

SHAP-based analyses indicate that the final models used information from both EDA and HRV feature groups. The SHAP feature importance plots for the binary and multiclass models are provided in Appendix Fig. G.12.

Table 8. Most important features from the final Extra Trees models. Importance values are based on the fitted model's impurity-based feature importance

Task	Feature	Importance
Binary	SCR Frequency	0.1042
Binary	AVNN	0.0488
Binary	HRV mean	0.0467
Binary	HRV harmonic mean	0.0456
Binary	HRV 75th percentile	0.0427
Multiclass	SCR Frequency	0.0537
Multiclass	HRV median	0.0383
Multiclass	HRV 25th percentile	0.0364
Multiclass	HRV mean	0.0359
Multiclass	AVNN	0.0356

5.6. Real-Time Inference Latency

Latency was evaluated using the final selected multiclass Extra Trees model and the same 59-feature input representation. Batch prediction required 0.104 ms per sample, while batch probability prediction required 0.093 ms per sample. Single-sample prediction had a mean latency of 103.42 ms with a standard deviation of 55.89 ms. Single-sample probability prediction had a mean latency of 95.36 ms with a standard deviation of 24.81 ms. These results suggest that model inference is computationally feasible for real-time closed-loop VR adaptation, particularly because the physiological windowing process operates on a longer time scale than the model prediction step.

6. Discussion

This study evaluated a machine learning-based physiological inference pipeline for the *Virtual Meditative Walk*, a VR biofeedback system designed to support mindfulness-based chronic pain management. The purpose was not to introduce a standalone stress classifier, but to examine whether EDA and HRV features could provide a stronger and more flexible control signal for adaptive environmental feedback than the original skin conductance slope-based rule.

The results support this direction, but they also require caution. In the cross-sample setting, the final Extra Trees model reached 91.77% accuracy and 81.10% F1-score for binary stress classification, and 86.05% accuracy and 86.16% F1-score for multiclass physiological state classification. The comparison with the original SCL slope-based rule shows why the machine learning pipeline is useful: the slope-based rule reached 62.35% accuracy and 37.40% F1-score in the binary task, indicating that a single skin conductance trend is too limited to support reliable adaptive feedback on its own.

The leave-one-subject-out results provide the more important boundary on this claim. Binary LOSO performance dropped to 73.50% accuracy, 67.84% balanced accuracy, and 22.04% F1-score. Multiclass LOSO performance dropped to 60.50% accuracy, 54.26% balanced accuracy, and 41.85% F1-score. This does not invalidate the model, but it changes how it should be used. The model can learn meaningful physiological structure from the combined dataset, but subject-independent inference remains difficult. For VMW, the classifier should therefore be treated as a probabilistic guide for environmental adaptation rather than as ground truth about the user.

6.1. Implications for VMW Biofeedback Design

The main design implication is that machine learning can make the VMW feedback loop more physiologically expressive, but the environmental mapping must remain conservative. The Extra Trees model combines EDA and HRV features, allowing the system to consider sympathetic arousal and cardiac variability together instead of relying only on SCL slope. The feature-importance and SHAP analyses support this interpretation: SCR frequency was consistently important, but HRV features such as AVNN, HRV mean, HRV median, and HRV percentile features also contributed.

This matters because the classifier is not the final product. Its value depends on how its output is translated into the VR experience. VMW uses fog, lighting, and sound as indirect feedback materials rather than presenting users with scores or labels. This is important for mindfulness-based chronic pain support, where feedback should encourage awareness without creating pressure to perform. The model provides a richer signal, but the design determines how that signal is felt.

The multiclass model is especially relevant because VMW does not only need to distinguish stress from non-stress. It also needs a neutral state for physiological windows that are ambiguous or not clearly related to stress or relaxation. The "other" class supports this design need by allowing medium fog density rather than forcing every window into a positive or negative feedback category. This helps the system avoid over-interpreting normal physiological variation and better matches the uncertainty present in mindfulness practice.

6.2. Real-Time Feasibility and Interaction Timing

The latency results suggest that the final Extra Trees model is computationally feasible for the type of real-time adaptation used in VMW. Batch prediction and probability prediction required less than one millisecond per sample, while single-sample prediction took approximately 103 ms on average. This is acceptable because VMW feedback is intentionally gradual. Fog, lighting, and ambient sound do not need to update at the speed of a game input; in fact, slower adaptation better supports the calm, non-evaluative pacing of the experience.

Overall, the findings suggest that machine learning can improve the adaptive capacity of VMW, but the contribution is not only technical. The stronger contribution lies in how physiological inference is integrated into an ambient feedback system for chronic pain support. The model improves what the system can estimate, while the VR design shapes how that estimate becomes meaningful to the user.

7. Limitations

Several limitations should be acknowledged. First, the model was trained and evaluated using WESAD and StressID rather than physiological data collected from chronic pain users during actual VMW sessions. Although these datasets are useful for developing and testing the inference pipeline, they do not fully represent the physiological, emotional, and contextual variability of chronic pain. Stress responses in chronic pain may be shaped by pain intensity, fatigue, medication use, sleep quality, movement restrictions, prior trauma, and individual coping strategies. In addition, the stress labels come from controlled experimental tasks rather than naturally occurring stress during mindfulness practice. Therefore, the model outputs should be interpreted as estimates of stress-related physiological state, not as direct measures of chronic pain experience or therapeutic progress.

Second, combining WESAD and StressID increased the number of subjects and physiological windows, but it also introduced heterogeneity in recording conditions, task design, sensor setup, sampling procedures, and labeling structure. The dataset also remained imbalanced, especially for the stress class. Class weighting, sample weighting, and balanced metrics were used to reduce this problem, but they do not fully solve it. The weaker leave-one-subject-out results show that subject-independent generalization remains limited. A future VMW system may need individual calibration, adaptive baselines, personalization, or periodic model updating before the inference pipeline can be used reliably with new users.

Third, the current evaluation focuses on model-level performance and latency rather than full closed-loop system performance. A deployed version of VMW must also account for sensor streaming, signal preprocessing, feature extraction, classification, communication with the VR environment, and rendering updates. The current pipeline was also evaluated using processed physiological windows rather than live wearable data collected during active VR use. Real-world use may introduce motion artifacts, sensor displacement, missing data, changes in skin contact, and variations in posture or breathing.

Finally, the environmental feedback rules used in VMW are design safeguards rather than clinically validated thresholds. They were included to make feedback more stable and less punitive, but they require empirical tuning with chronic pain users. This study also does not test whether the updated inference pipeline improves pain management, mindfulness skill development, emotional regulation, self-efficacy, or long-term engagement. The present work should therefore be interpreted as a proof-of-concept evaluation of the physiological inference and feedback pipeline, not as clinical validation of VMW as an intervention.

8. Conclusion

This paper presented an updated physiological inference pipeline for the *Virtual Meditative Walk*, a closed-loop VR biofeedback system designed to support mindfulness-based chronic pain management. The goal was to improve the reliability of the environmental feedback mechanism by moving beyond the original skin conductance slope-based rule and evaluating a machine learning approach using EDA and HRV features.

The final Extra Trees model showed strong cross-sample performance, reaching 91.77% accuracy for binary stress classification and 86.05% accuracy for multiclass physiological state classification. It also outperformed the original SCL slope-based rule in the cross-sample binary comparison, suggesting that multivariate physiological inference can provide a more expressive control signal for adaptive VR biofeedback. At the same time, leave-one-subject-out validation produced lower performance, showing that subject-independent stress inference remains a central challenge.

These findings support machine learning as a promising direction for improving the adaptive capacity of VMW, but not as a complete solution. The contribution of this work lies in connecting physiological state inference with ambient VR biofeedback design: the model improves what the system can estimate, while the VR environment determines how that estimate is translated into feedback that remains gradual, indirect, and non-evaluative. Future work should evaluate the updated VMW system with chronic pain users during live VR sessions, including end-to-end latency, wearable sensor robustness, individual calibration, and whether adaptive feedback improves mindfulness engagement, interoceptive awareness, self-regulation, or pain-related outcomes over repeated use.

9. Declarations

Funding

This research was supported by NSERC grant 371783-2013(16262): VR systems for body image and body schema.

Competing Interests

Diane Gromala declares a financial interest in Easa Therapeutics. The authors state that this interest has no direct connection to the study reported in this manuscript. The remaining authors declare no competing interests.

Ethics and Data Use

No new human-subject physiological data were collected for the machine learning evaluation reported in this paper. The analysis used the WESAD and StressID datasets. WESAD is publicly available from its original providers. StressID was used under an academic research license from Inria and EURECOM.

Data Availability

The combined processed dataset used in this study is not publicly available from the authors because it includes StressID data, which are subject to license restrictions and cannot be redistributed. Researchers interested in using StressID should obtain access directly from the original dataset providers and comply with their respective terms of use.

Code Availability

The code used for the analyses reported in this paper is available on GitHub: Visualizing Autonomic States Through VR Biofeedback to Support Mindfulness.

Data Privacy

For future clinical deployment of VMW, we emphasize the importance of privacy-preserving biosignal processing, including on-device processing where feasible rather than cloud-based streaming, in line with GDPR-informed data protection principles.

Disclaimer

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, the publisher, the editors, or the reviewers.

References

[1] Schopflocher, D., Taenzer, P., Jovey, R.: The prevalence of chronic pain in Canada. Pain Research & Management **16**(6), 445–450 (2011). doi:10.1155/2011/876306

[2] Foley, H.E., Knight, J.C., Ploughman, M., Asghari, S., Audas, R.: Association of chronic pain with comorbidities and health care utilization: a retrospective cohort study using health administrative data. Pain **162**(11), 2737–2749 (2021). doi:10.1097/j.pain.0000000000002264

[3] Gatchel, R.J., Peng, Y.B., Peters, M.L., Fuchs, P.N., Turk, D.C.: The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychological Bulletin **133**(4), 581–624 (2007). doi:10.1037/0033-2909.133.4.581

[4] Nadeau, S.E., Wu, J.K., Lawhern, R.A.: Opioids and chronic pain: an analytic review of the clinical evidence. Frontiers in Pain Research **2**, 721357 (2021). doi:10.3389/fpain.2021.721357

[5] Mackey, S., Aghaeepour, N., Gaudilliere, B., Kao, M.-C., Kaptan, M., Lannon, E., Pfyffer, D., Weber, K.: Innovations in acute and chronic pain biomarkers: enhancing diagnosis and personalized therapy. Regional Anesthesia & Pain Medicine **50**(2), 110–120 (2025). doi:10.1136/rapm-2024-106030

[6] Jamison, R.N., Ross, E.L., Michna, E., Chen, L.Q., Holcomb, C., Wasan, A.D.: Substance misuse treatment for high-risk chronic pain patients on opioid therapy: a randomized trial. Pain **150**(3), 390–400 (2010). doi:10.1016/j.pain.2010.02.033

[7] Davis, K.D., Aghaeepour, N., Ahn, A.H., Angst, M.S., Borsook, D., Brenton, A., Burczynski, M.E., Crean, C., Edwards, R., Gaudilliere, B., Hergenroeder, G.W., Iadarola, M.J., Iyengar, S., Jiang, Y., Kong, J.-T., Mackey, S., Saab, C.Y., Sang, C.N., Scholz, J., Pelleymounter, M.A.: Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. Nature Reviews Neurology **16**(7), 381–400 (2020). doi:10.1038/s41582-020-0362-2

[8] Topol, E.: Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books, New York (2019). doi:10.5555/3350442

[9] Hoffman, H.G., Patterson, D.R., Carrougher, G.J., Sharar, S.R.: Effectiveness of virtual reality-based pain control with multiple treatments. The Clinical Journal of Pain **17**(3), 229–235 (2001). doi:10.1097/00002508-200109000-00007

[10] Moreau, S., Therond, A., Cerda, I.H., Studer, K., Pan, A., Tharpe, J., Crowther, J.E., Abd-Elsayed, A., Gilligan, C., Tolba, R., Ashina, S., Schatman, M.E., Kaye, A.D., Yong, R.J., Robinson, C.L.: Virtual reality in acute and chronic pain medicine: an updated review. Current Pain and Headache Reports **28**(9), 893–928 (2024). doi:10.1007/s11916-024-01246-2

[11] Gopalan, R., Pande, H., Narayanan, S., Chinnaswami, A.: Virtual reality as a nonpharmacological tool for acute pain management: a scoping review. Innovations in Clinical Neuroscience **22**(1–3), 28–50 (2025).

[12] Wong, K.P., Tse, M.M.Y., Qin, J.: Effectiveness of virtual reality-based interventions for managing chronic pain

on pain reduction, anxiety, depression and mood: a systematic review. Healthcare **10**(10), 2047 (2022). doi:10.3390/healthcare10102047

[13] Siu, K.-C., Hao, J.: Psychological aspects of virtual reality in chronic non-communicable diseases. In: Martin, C.R., Preedy, V.R., Patel, V.B., Rajendram, R. (eds.) Handbook of the Behavior and Psychology of Disease, pp. 1–20. Springer, Cham (2025). doi:10.1007/978-3-031-32046-0_128-1

[14] Garofano, M., Del Sorbo, R., Calabrese, M., Giordano, M., Di Palo, M.P., Bartolomeo, M., Ragusa, C.M., Ungaro, G., Fimiani, G., Di Spirito, F., Amato, M., Ciccarelli, M., Pascarelli, C., Scanniello, G., Bramanti, P., Bramanti, A.: Remote rehabilitation and virtual reality interventions using motion sensors for chronic low back pain: a systematic review of biomechanical, pain, quality of life, and adherence outcomes. Technologies **13**(5), 186 (2025). doi:10.3390/technologies13050186

[15] Calderone, A., Mazzurco Masi, V.M., De Luca, R., Gangemi, A., Bonanno, M., Floridia, D., Corallo, F., Morone, G., Quartarone, A., Maggio, M.G., Calabro, R.S.: The impact of biofeedback in enhancing chronic pain rehabilitation: a systematic review of mechanisms and outcomes. Heliyon **11**(2), e41917 (2025). doi:10.1016/j.heliyon.2025.e41917

[16] Nasir, A., Mahmood, S., Mahmood, S.: Digital pain interventions: a promising new paradigm for pain control. Anaesthesia, Pain & Intensive Care **27**, 256–265 (2023). doi:10.35975/apic.v27i2.2175

[17] Yu, B., Funk, M., Hu, J., Wang, Q., Feijs, L.: Biofeedback for everyday stress management: A systematic review. Frontiers in ICT **5**, Article 23 (2018). doi:10.3389/fict.2018.00023

[18] Kabat-Zinn, J.: *Mindfulness Meditation for Pain Relief: Practices to Reclaim Your Body and Your Life*. St. Martin's Essentials / Sounds True, New York (2023)

[19] Tong, X., Gromala, D., Choo, A., Amin, A., Shaw, C.: The virtual meditative walk: an immersive virtual environment for pain self-modulation through mindfulness-based stress reduction meditation. In: Shumaker, R., Lackey, S. (eds.) Virtual, Augmented and Mixed Reality, pp. 330–339. Springer, Cham (2015). doi:10.1007/978-3-319-21067-4_40

[20] Gromala, D., Tong, X., Choo, A., Karamnejad, M., Shaw, C.D.: The virtual meditative walk: virtual reality therapy for chronic pain management. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), Seoul, Republic of Korea, pp. 521–524. ACM, New York (2015). doi:10.1145/2702123.2702344

[21] Gromala, D., Barnes, S., Song, M., Fox, T.: Virtual Meditative Walk (2010).

[22] Vos, G., Trinh, K., Sarnyai, Z., Rahimi Azghadi, M.: Generalizable machine learning for stress monitoring from wearable devices: a systematic literature review. International Journal of Medical Informatics **173**, 105026 (2023). doi:10.1016/j.ijmedinf.2023.105026

[23] Pinge, A., Gad, V., Jaisighani, D., Ghosh, S., Sen, S.: Detection and monitoring of stress using wearables: a systematic review. Frontiers in Computer Science **6**, 1478851 (2024). doi:10.3389/fcomp.2024.1478851

[24] Ding, M.E., Traiba, H., Perez, H.R.: Virtual reality interventions and chronic pain: scoping review. Journal of Medical Internet Research **27**, e59922 (2025). doi:10.2196/59922

[25] Garcia, L.M., Birckhead, B.J., Krishnamurthy, P., Sackman, J., Mackey, I.G., Louis, R.G., Salmasi, V., Maddox, T., Darnall, B.D.: An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled trial conducted during COVID-19. Journal of Medical Internet Research **23**(2), e26292 (2021). doi:10.2196/26292

[26] O'Connor, S., Mayne, A., Hood, B.: Virtual reality based mindfulness for chronic pain management: a scoping review. Pain Management Nursing **23**(3), 359–369 (2022). doi:10.1016/j.pmn.2022.03.013

[27] Mahrer, N.E., Gold, J.I.: The use of virtual reality for pain control: a review. Current Pain and Headache Reports **13**(2), 100–109 (2009). doi:10.1007/s11916-009-0019-8

[28] Malloy, K.M., Milling, L.S.: The effectiveness of virtual reality distraction for pain reduction: a systematic review. Clinical Psychology Review **30**(8), 1011–1018 (2010). doi:10.1016/j.cpr.2010.07.001

[29] Ng, J., Lo, H., Tong, X., Gromala, D., Jin, W.: Farmooo, a virtual reality farm simulation game designed for pediatric cancer patients to distract their pain during chemotherapy treatment. Electronic Imaging **2018**(3), 4321–4324 (2018). doi:10.2352/ISSN.2470-1173.2018.03.ERVR-432

[30] Sveistrup, H.: Motor rehabilitation using virtual reality. Journal of NeuroEngineering and Rehabilitation **1**(1), 10 (2004). doi:10.1186/1743-0003-1-10

[31] Tong, X., Gromala, D., Machuca, F.: LumaPath: an immersive virtual reality game for encouraging physical activity for senior arthritis patients. In: International Conference on Human-Computer Interaction (HCII 2019), pp. 384–397. Springer, Cham (2019). doi:10.1007/978-3-030-21565-1_26

[32] Chittaro, L., Longhino, S., Serafini, M., Cacioppo, S., Quartuccio, L.: Efficacy of immersive virtual reality combined with multisensor biofeedback on chronic pain in fibromyalgia: a pilot randomized controlled trial. ACR Open Rheumatology **7**(5), e70048 (2025). doi:10.1002/acr2.70048

[33] R., R., Moger, S., Umesh, S., Bhuvana, G.: Revolutionizing pain management: the future of AAT, BCI, exoskeleton, and soft robotics. In: Emerging Trends in Assistive Technologies for Health and Wellbeing, pp. 129–164. IGI Global, Hershey, PA (2025). doi:10.4018/979-8-3693-9501-1.ch005

[34] Diotaiuti, P., Corrado, S., Tosti, B., Spica, G., Di Libero, T., D'Oliveira, A., Zanon, A., Rodio, A., Andrade, A., Mancone, S.: Evaluating the effectiveness of neurofeedback in chronic pain management: a narrative review. Frontiers in Psychology **15**, 1369487 (2024). doi:10.3389/fpsyg.2024.1369487

[35] Reddy, K.J.: Motor rehabilitation and biofeedback. In: Reddy, K.J. (ed.) Innovations in Neurocognitive Rehabilitation: Harnessing Technology for Effective Therapy, pp. 231–266. Springer, Cham (2025). doi:10.1007/978-3-031-88117-6_11

[36] Gkintoni, E., Vassilopoulos, S.P., Nikolaou, G., Vantarakis, A.: Neurotechnological approaches to cognitive rehabilitation in mild cognitive impairment: a systematic review of neuromodulation, EEG, virtual reality, and emerging AI applications. Brain Sciences **15**(6), 582 (2025). doi:10.3390/brainsci15060582

[37] Schmidt, P., Reiss, A., Duerichen, R., Marberger, C., Van Laerhoven, K.: Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18), Boulder, CO, USA, pp. 400–408. ACM, New York (2018). doi:10.1145/3242969.3242985

[38] Chaptoukaev, H., Strizhkova, V., Panariello, M., Dalpaos, B., Reka, A., Manera, V., Thummler, S., Ismailova, E., Evans, N.W., Bremond, F., Todisco, M., Zuluaga, M.A., Ferrari, L.M.: StressID: a multimodal dataset for stress identification. In: Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track (2023). https://openreview.net/forum?id=qWsQi9DGJb

[39] Yin, Q., Shen, D., Ding, Q.: Influence of sliding time window size selection based on heart rate variability signal analysis on intelligent monitoring of noxious stimulation under anesthesia. Neural Plasticity **2021**, 6675052 (2021). doi:10.1155/2021/6675052

[40] Anusha, A.S., Joy, J., Preejith, S.P., Joseph, J., Mohanasankar, S.: Physiological signal based work stress detection using unobtrusive sensors. Biomedical Physics & Engineering Express **4**(6), 065001 (2018). doi:10.1088/2057-1976/aadbd4

[41] Mohammadpoor Faskhodi, M., Fernandez-Chimeno, M., Garcia-Gonzalez, M.A.: Arousal detection by using ultra-short-term heart rate variability (HRV) analysis. Frontiers in Medical Engineering **2**, 1209252 (2023). doi:10.3389/fmede.2023.1209252

[42] Baek, H.J., Cho, C.H., Cho, J., Woo, J.M.: Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. Telemedicine and e-Health **21**(5), 404–414 (2015). doi:10.1089/tmj.2014.0104

[43] Posada-Quintero, H.F., Florian, J.P., Orjuela-Canon, A.D., Aljama-Corrales, T., Charleston-Villalobos, S., Chon, K.H.: Power spectral density analysis of electrodermal activity for sympathetic function assessment. Annals of Biomedical Engineering **44**, 3124–3135 (2016). doi:10.1007/s10439-016-1606-6

[44] Amin, M.R., Faghih, R.T.: Identification of sympathetic nervous system activation from skin conductance: a sparse decomposition approach with physiological priors. IEEE Transactions on Biomedical Engineering **68**, 1726–1736 (2020). doi:10.1109/TBME.2020.3034632

[45] Aladag, S., Guven, A., Ozbek, H., Dolu, N.: A comparison of denoising methods for Electrodermal Activity signals. In: 2015 Medical Technologies National Conference (TIPTEKNO), pp. 1–4 (2015). doi:10.1109/TIPTEKNO.2015.7374563

[46] Gupta, V., Mittal, M., Mittal, V.: Performance evaluation of various pre-processing techniques for R-peak detection in ECG signal. IETE Journal of Research **68**(5), 3267–3282 (2022). doi:10.1080/03772063.2020.1756473

[47] Lee, G., Choi, B., Jebelli, H., Ahn, C.R., Lee, S.: Noise reference signal-based denoising method for EDA collected by multimodal biosensor wearable in the field. Journal of Computing in Civil Engineering **34**(6), 04020044 (2020). doi:10.1061/(ASCE)CP.1943-5487.0000927

[48] Goldberger, J.J., Johnson, N.P., Subacius, H., Ng, J., Greenland, P.: Comparison of the physiologic and prognostic implications of the heart rate versus the RR interval. Heart Rhythm **11**(11), 1925–1933 (2014). doi:10.1016/j.hrthm.2014.07.037

[49] Vest, A.N., Poian, G.D., Li, Q., Liu, C., Nemati, S., Shah, A., Clifford, G.D.: An open source benchmarked toolbox for cardiovascular waveform and interval analysis. Physiological Measurement **39**(10), 105004 (2018). doi:10.1088/1361-6579/aae021

[50] Ishaque, S., Khan, N., Krishnan, S.: Physiological signal analysis and stress classification from VR simulations using decision tree methods. Bioengineering **10**(7), 766 (2023). doi:10.3390/bioengineering10070766

[51] Moridani, M., Zahra, M., Nooshin, J.: Heart rate variability features for different stress classification. Bratislavske Lekarske Listy **121**, 619–627 (2020). doi:10.4149/BLL_2020_107

[52] Patlar Akbulut, F.: Evaluating the effects of the autonomic nervous system and sympathetic activity on emotional states. Istanbul Ticaret Universitesi Fen Bilimleri Dergisi **21**(41), 156–169 (2022). doi:10.55071/ticaretfbd.1125431

[53] Melillo, P., Bracale, M., Pecchia, L.: Nonlinear heart rate variability features for real-life stress detection. Case study: students under stress due to university examination. BioMedical Engineering OnLine **10**(1), 96 (2011). doi:10.1186/1475-925X-10-96

[54] Long, K., Zhang, X., Wang, N., Lei, H.: Heart rate variability during online video game playing in habitual gamers: effects of internet addiction scale, ranking score and gaming performance. Brain Sciences **14**(1), 29 (2023). doi:10.3390/brainsci14010029

[55] Goni, M., Mohino-Herranz, I., Llerena, C., Gil-Pita, R., Rosa-Zurera, M.: Feature selection in mental stress analysis using multiple biological signals. In: Proceedings of the 2013 IASTED International Conference on Biomedical Engineering (BioMed 2013) (2013). doi:10.2316/P.2013.798-090

[56] Dalmeida, K.M., Masala, G.L.: HRV features as viable physiological markers for stress detection using wearable devices. Sensors **21**(8), 2873 (2021). doi:10.3390/s21082873

[57] Wu, M., Cao, H., Nguyen, H.L., Surmacz, K., Hargrove, C.: Modeling perceived stress via HRV and accelerometer sensor streams. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2015), pp. 1625–1628. IEEE, Milan (2015). doi:10.1109/EMBC.2015.7318686

[58] Sanchez-Reolid, R., Lopez de la Rosa, F., Sanchez-Reolid, D., Lopez, M.T., Fernandez-Caballero, A.: Machine learning techniques for arousal classification from electrodermal activity: a systematic review. Sensors **22**(22), 8886 (2022). doi:10.3390/s22228886

Appendix A. VMW Closed-Loop System Workflow

The following figure shows how physiological signals are collected, processed, classified, and mapped into environmental feedback in the *Virtual Meditative Walk* system.

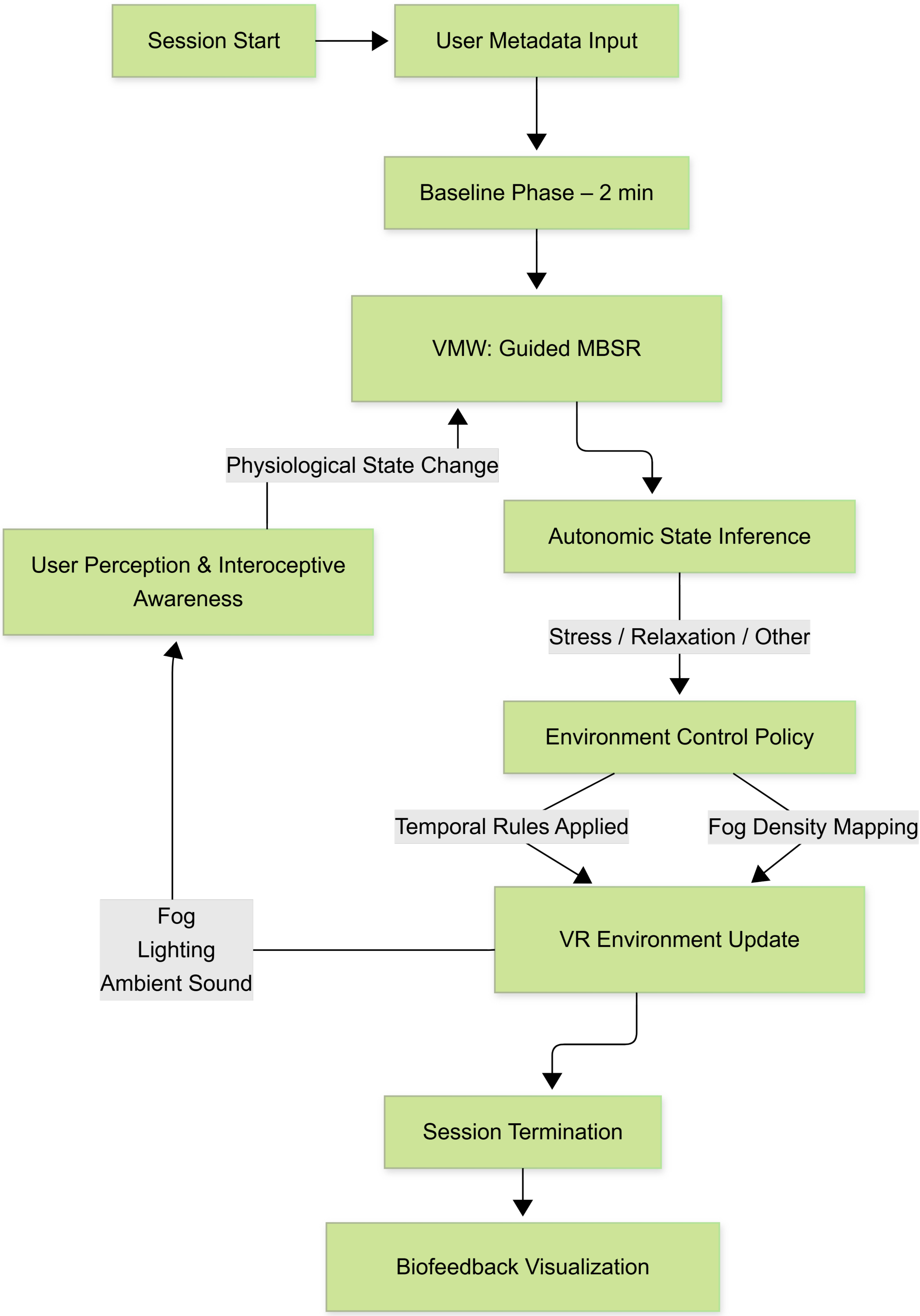


Figure A.1. Closed-loop workflow of the *Virtual Meditative Walk* (VMW) system

Appendix B. Physiological Feature Set

Table B.1. Physiological features used as model inputs. Metadata columns, dataset identifiers, subject identifiers, task names, window timestamps, original labels, and target labels were excluded from model training

Signal	Feature group	Features
EDA	Statistical features	EDA_mean, EDA_geometric_mean, EDA_harmonic_mean, EDA_standard_deviation, EDA_mean_absolute_deviation, EDA_variance, EDA_median, EDA_maximum, EDA_minimum, EDA_25th_percentile, EDA_75th_percentile, EDA_range, EDA_interquartile_range, EDA_variation_coefficient, EDA_central_moment, EDA_skewness, EDA_kurtosis
EDA	Nonlinear features	EDA_approx_entropy, EDA_sample_entropy, EDA_SD1, EDA_SD2
EDA	Domain-specific features	SCL, SCR Amplitude, SCR Duration, SCR Frequency
HRV	Statistical features	HRV_mean, HRV_geometric_mean, HRV_harmonic_mean, HRV_standard_deviation, HRV_mean_absolute_deviation, HRV_variance, HRV_median, HRV_maximum, HRV_minimum, HRV_25th_percentile, HRV_75th_percentile, HRV_range, HRV_interquartile_range, HRV_variation_coefficient, HRV_central_moment, HRV_skewness, HRV_kurtosis
HRV	Nonlinear features	HRV_approx_entropy, HRV_sample_entropy, HRV_SD1, HRV_SD2
HRV	Time-domain and respiratory-related features	SDNN, SDANN, AVNN, RMSSD, PNN50, RSA Index, Deep Breathing Difference (DBD), Exhalation/Inspiration Ratio (EI)
HRV	Frequency-domain features	Total Power, VLF Power, LF Power, HF Power, LF/HF Ratio

Appendix C. Dataset Composition and Label Distributions

The following figures provide additional detail about the composition of the combined WESAD and StressID dataset, the target label distributions, and subject-level stress prevalence.

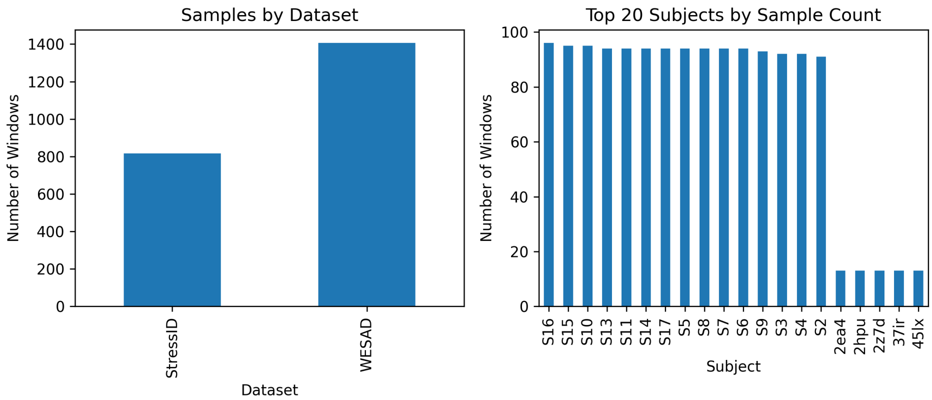


Figure C.2. Composition of the combined WESAD and StressID dataset

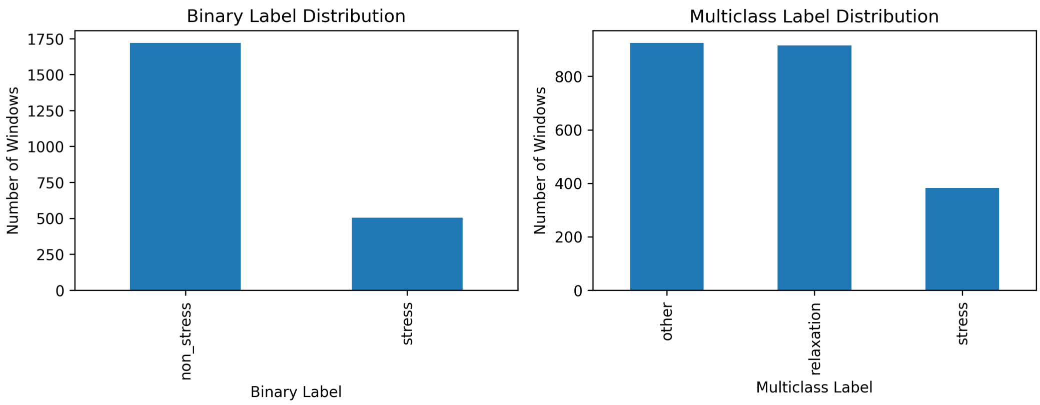
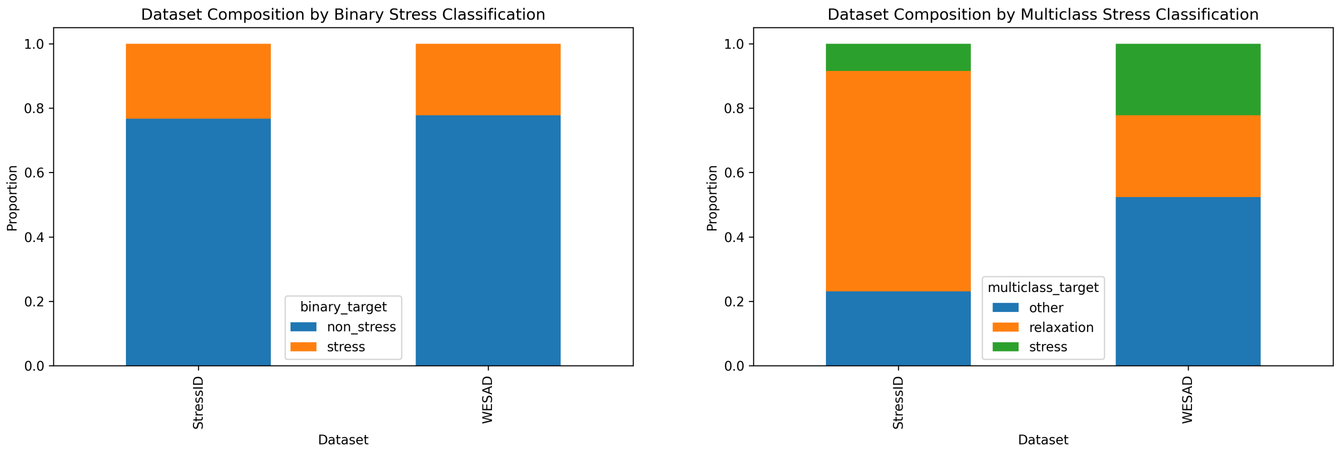


Figure C.3. Binary and multiclass label distributions in the combined dataset



(a) Binary stress classification target

(b) Multiclass physiological state classification target

Figure C.4. Dataset composition by target structure. The left panel shows the binary stress classification target, and the right panel shows the multiclass physiological state classification target

Appendix D. Exploratory Feature Analysis

The following figures provide additional feature-level analysis, including feature-label correlations, correlation heatmaps, and feature distributions across binary and multiclass targets.

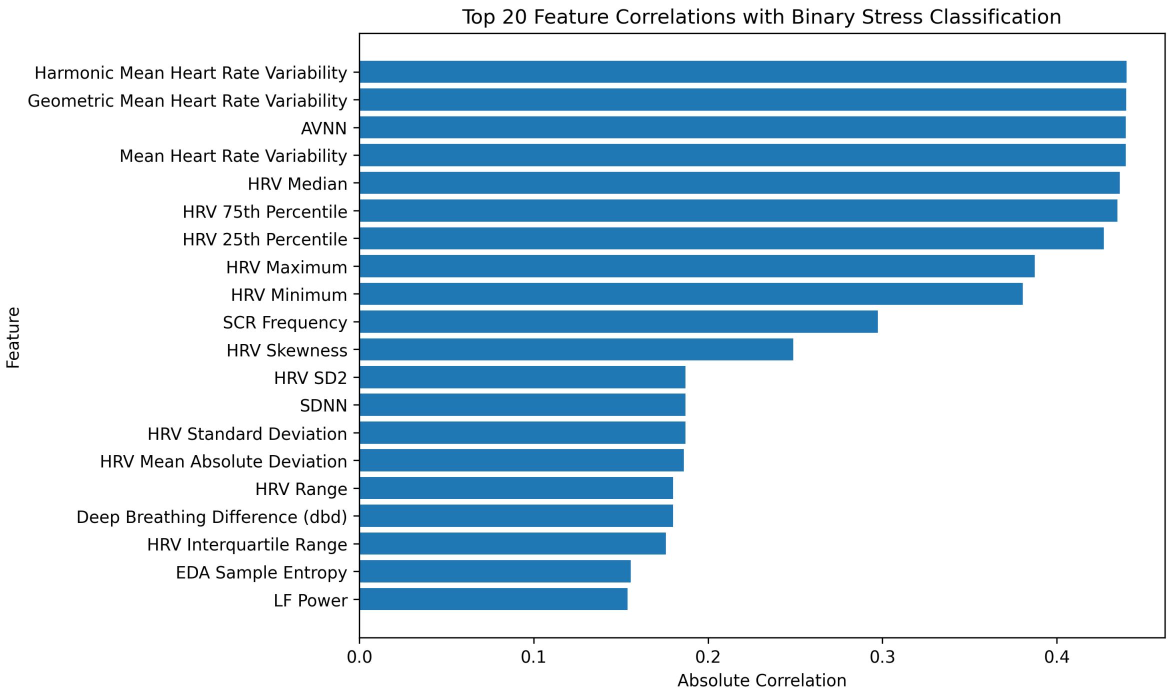


Figure D.5. Top correlated features for binary stress classification

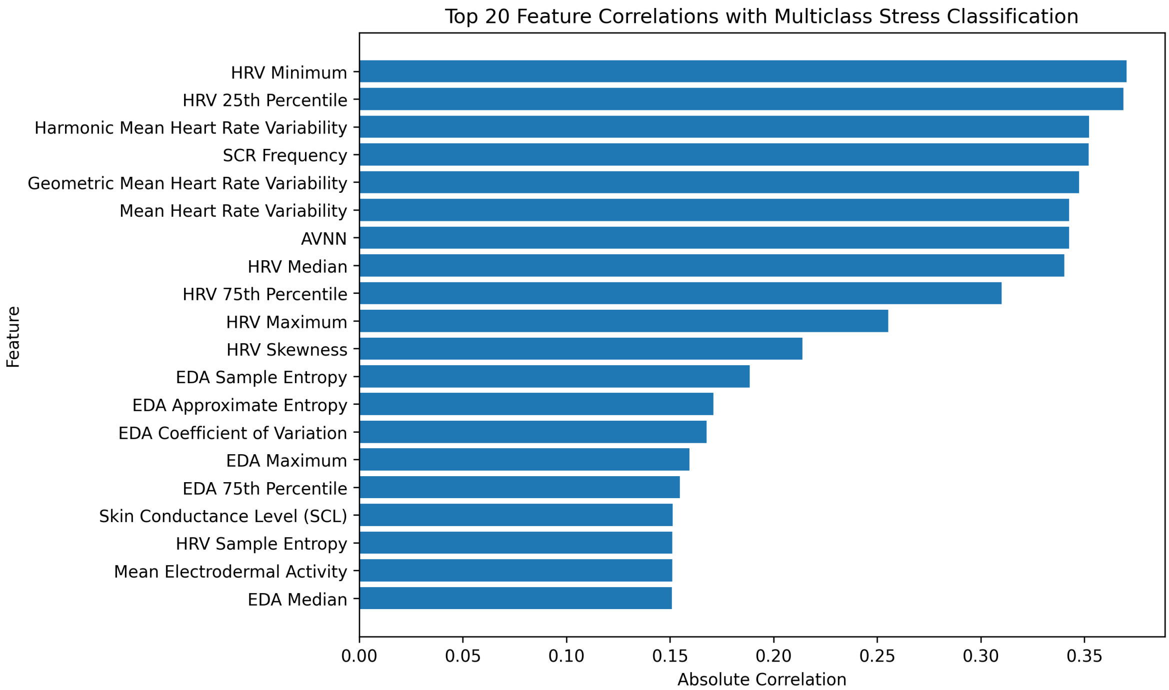
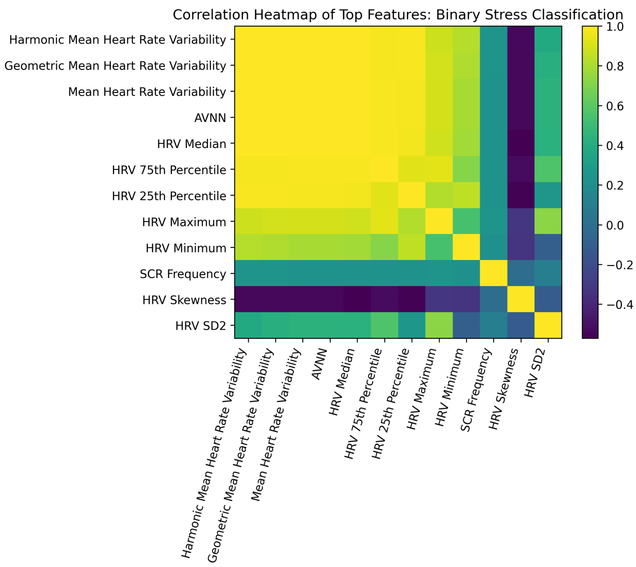
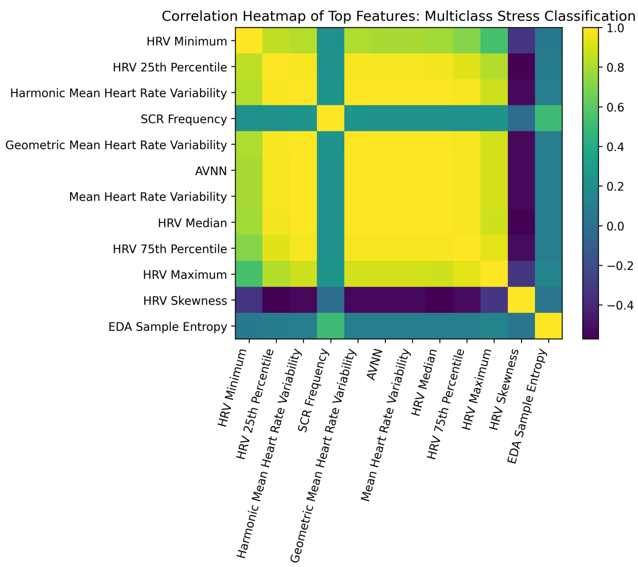


Figure D.6. Top correlated features for multiclass physiological state classification



(a) Binary stress classification

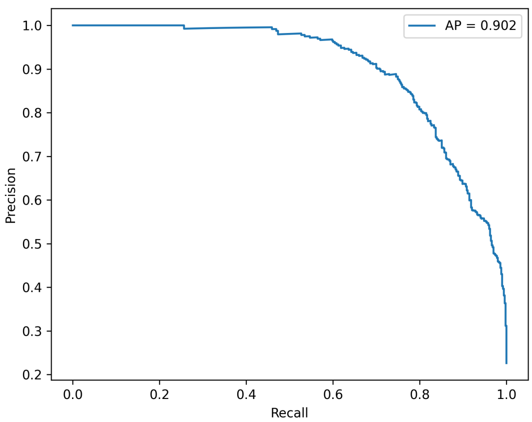


(b) Multiclass physiological state classification

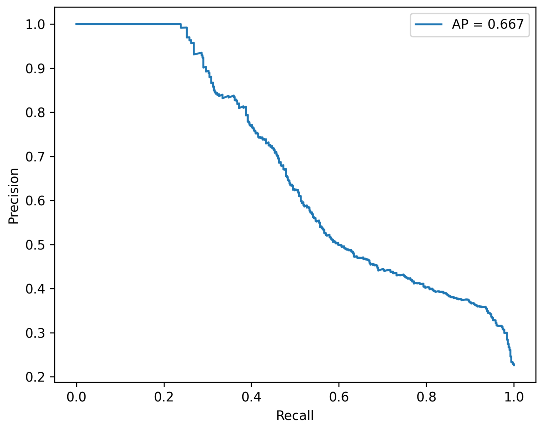
Figure D.7. Correlation heatmaps of selected features for binary stress classification and multiclass physiological state classification

Appendix E. Precision-Recall Analysis

The following figures provide additional precision-recall diagnostics for the binary and multiclass classification tasks. These plots are included because the stress class was smaller than the non-stress class, making precision-recall behavior useful for interpreting model performance beyond accuracy and ROC-AUC.

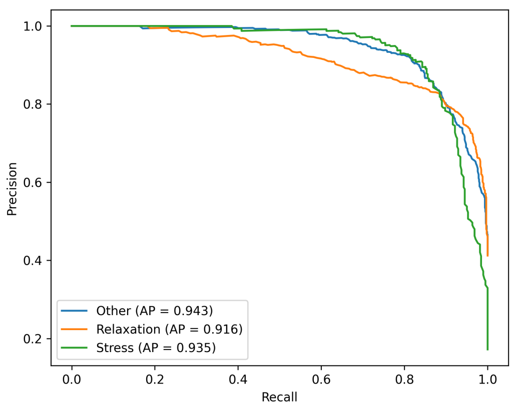


(a) Cross-sample precision-recall curve

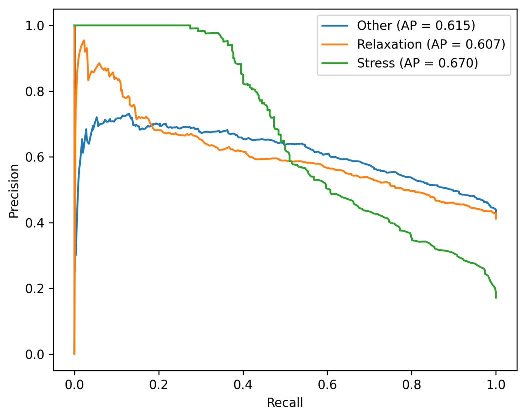


(b) LOSO precision-recall curve

Figure E.8. Precision-recall curves for binary stress classification



(a) Cross-sample precision-recall curves

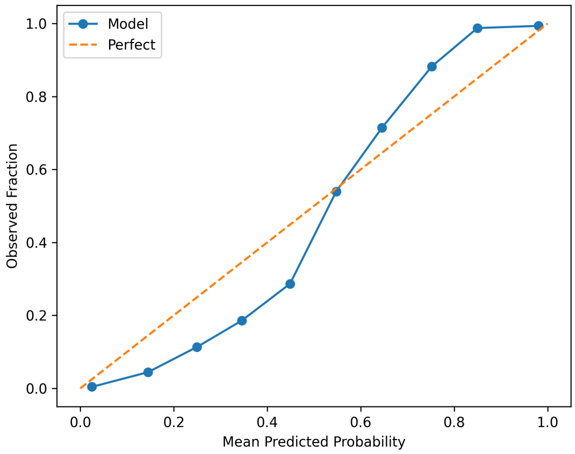


(b) LOSO precision-recall curves

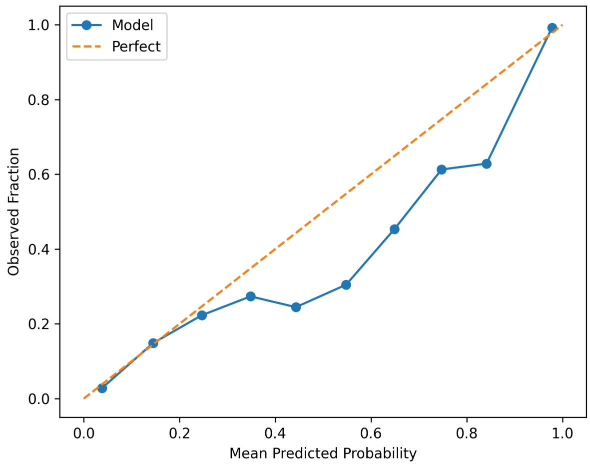
Figure E.9. Precision-recall curves for multiclass physiological state classification

Appendix F. Probability Calibration Analysis

The following figures show calibration behavior for the final Extra Trees models. These plots are included as additional diagnostics for interpreting whether predicted probabilities align with observed class frequencies.

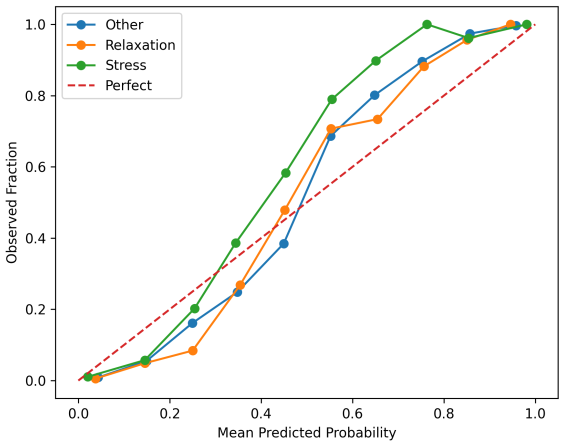


(a) Cross-sample calibration plot

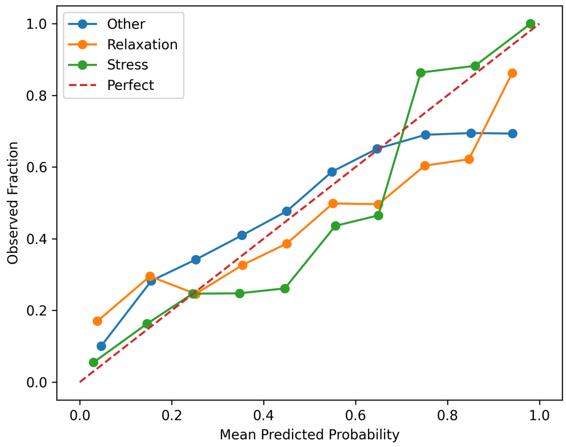


(b) LOSO calibration plot

Figure F.10. Calibration plots for binary stress classification



(a) Cross-sample calibration plot

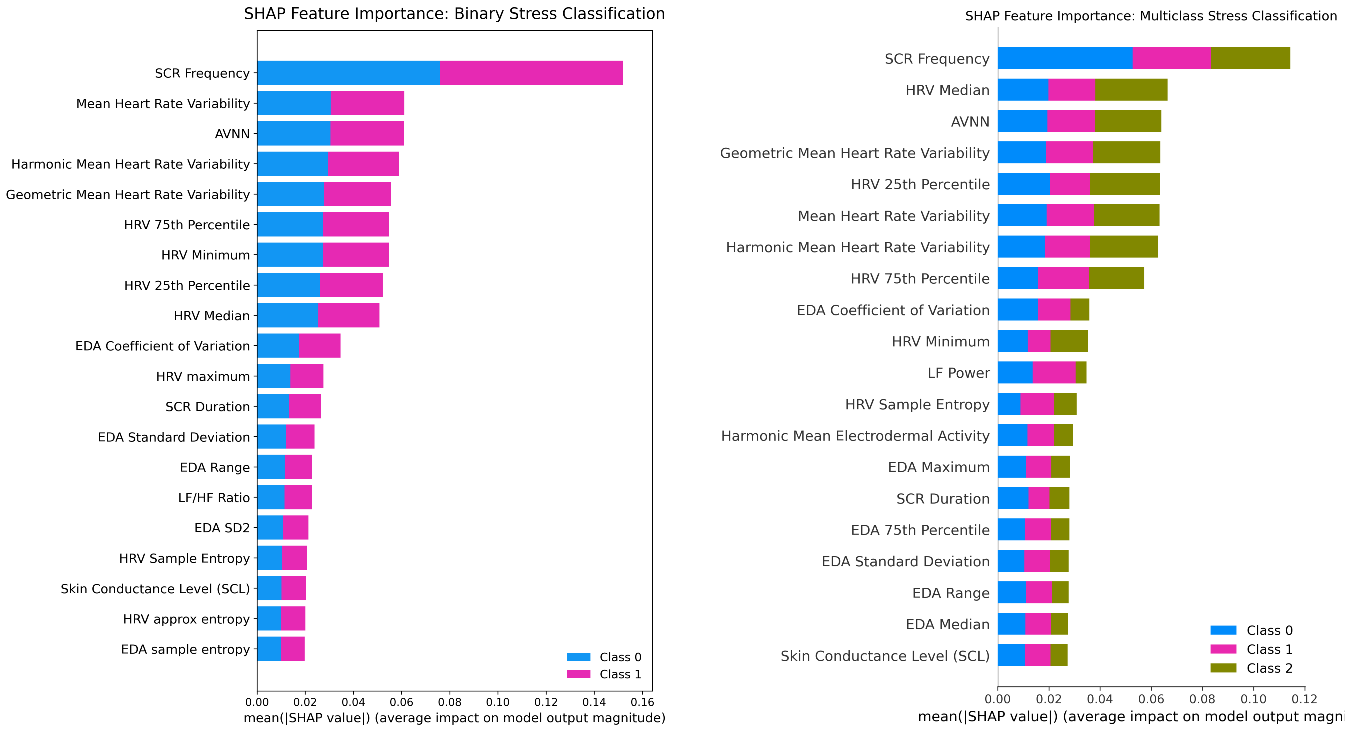


(b) LOSO calibration plot

Figure F.11. Calibration plots for multiclass physiological state classification

Appendix G. SHAP-Based Model Interpretation

The following figures provide additional model interpretation analyses for the final Extra Trees models.

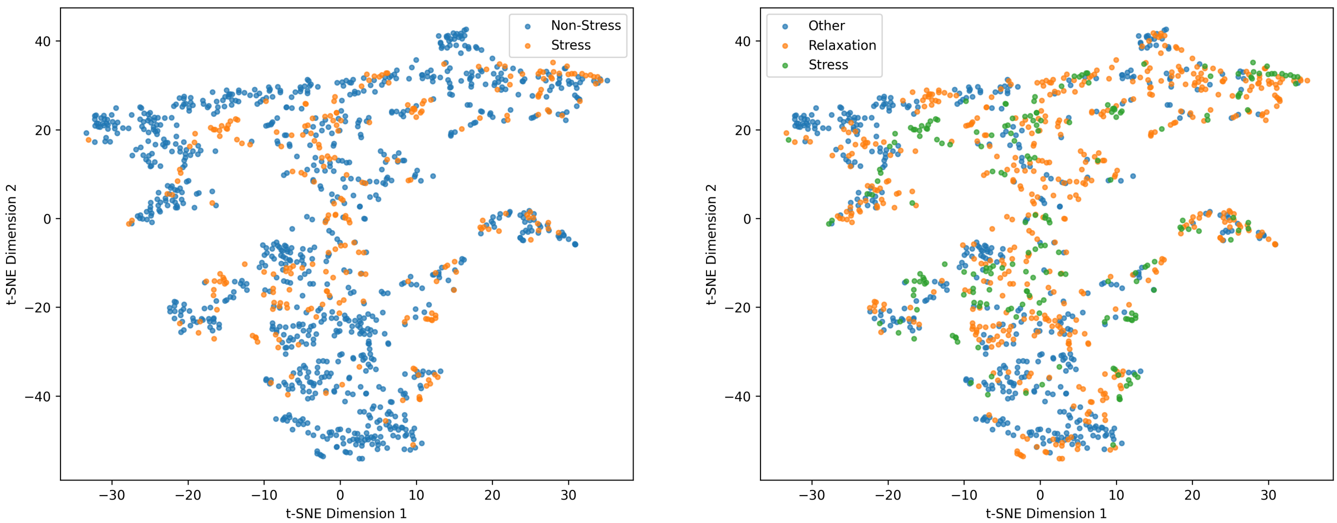


(a) Binary stress classification (b) Multiclass physiological state classification

Figure G.12. SHAP feature importance for the final Extra Trees models. The left panel shows the binary stress classification model, and the right panel shows the multiclass physiological state classification model

Appendix H. Feature-Space Visualization

The following t-SNE plots provide an additional visualization of the feature-space structure for the binary and multiclass classification t asks. T hese p lots a re i ncluded a s e xploratory d iagnostics a nd a re n ot u sed a s direct evidence of model performance.



(a) Binary stress classification (b) Multiclass physiological state classification

Figure H.13. t-SNE embeddings of the physiological feature space for binary and multiclass classification tasks