Real-time emotion recognition in virtual reality from behavioral motion[*]

Nonna Kulishova [*,†] and Volodymyr Sylantiev[†]

Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

Abstract

This work presents a real-time, motion-only pipeline that recognizes affect in VR using consumer hardware (Meta/Oculus Quest 2). The approach relies solely on behavioral kinematics – head and hand 6DoF streams – captured at 90 FPS, avoiding additional physiological sensors. Raw pose sequences are segmented into 2 seconds sliding windows with overlap, preprocessed via smoothing, outlier handling, and gesture debouncing, then transformed into a compact feature set (kinematics, posture, gesture dynamics, spectral descriptors, and summary statistics). The feature space is standardized and reduced with PCA that preserve 95% variance, supporting both unsupervised structure discovery (Mini-Batch K-Means, DBSCAN) and supervised mapping (classification) to five emotion categories (joy, focus, boredom, anxiety, stress) using Random Forests, SVM, and a lightweight 1D-CNN for mapping evaluation.

Keywords

Virtual Reality; Affective Computing; Behavioral Motion; Emotion Recognition; Clustering; Classification.

1. Introduction

The rapid advancement of virtual reality (VR) technologies has fundamentally transformed how people interact with digital environments, delivering compelling immersive experiences that blur the boundary between the real and the virtual world. VR has emerged as a powerful platform – ranging from video games and training simulators to telemedicine and educational applications – where user engagement and satisfaction are key determinants of success.

A major obstacle to improving such experiences is understanding the emotional and psychological state of users while they interact with VR content. In VR where sensory load is carefully controlled the user's emotional state can directly affect task performance, sense of presence, and cognitive workload. Accordingly, integrating affective computing – the capacity of systems to recognize, interpret, and respond to human emotions – is necessary for building emotionally adaptive virtual environments [1][2][3][4].

Although traditional emotion-recognition systems largely rely on physiological measures (e.g., electroencephalography, heart-rate variability, electrodermal activity) or facial-expression analysis, these methods often require additional equipment, involve invasive sensors, or exhibit limited generality across users and contexts. In contrast, behavioral movement analysis offers a non-invasive, scalable alternative. How a user moves – head tilts, hand gestures, body posture, and movement patterns – can contain rich information about emotional and cognitive state.

The aim of this work is to develop a real-time system for recognizing a user's emotional state in virtual reality that relies exclusively on natural behavioral motion captured with an Oculus Quest 2 headset or compatible VR devices. To achieve this aim, we implement a staged process that includes: high-rate motion data acquisition; construction of informative features from these data; clustering of the discovered movement patterns; and classification of emotional states – joy, stress, boredom, focus, and anxiety.

[*] *CIAW-2025: Computational Intelligence Application Workshop, September 26-27, 2025, Lviv, Ukraine*

[1*] Corresponding author.

[†] These authors contributed equally.

nonna.kulishova@nure.ua (N. Kulishova); volodymyr.sylantiev1@nure.ua (V. Sylantiev)

0000-0003-1142-4100 (N. Kulishova); 0009-0009-5616-9431 (V. Sylantiev)

System design accounts for strict real-time constraints (target rendering at 90 FPS and end-to-end latency below 30 ms), ease of integration into existing VR applications, and ergonomics and scalability considerations.

We expect that completing these stages will align motion-based interaction with automatic emotion recognition in VR, thereby contributing to affective computing and human-centered immersive systems.

2. Literature Review

2.1. Affective computing in VR

Research on recognizing emotional states in immersive environments combines elements of several disciplines: affective computing, human–computer interaction (HCI), motion analysis, and machine learning. Affective computing – an interdisciplinary field initiated by Rosalind Picard in 1997 [5] – seeks to equip machines with the ability to perceive, interpret, and respond to human emotions. In user-centered systems, an emotionally aware response is a key component of the overall experience [5].

Within VR, traditional affective computing methods often rely on physiological signals to recognize emotions. The most common data sources are: electrodermal activity (EDA) or galvanic skin response (GSR), which measures sweat gland activity as an indicator of arousal [6][7]; heart rate (HR) and heart rate variability (HRV), which reflect sympathetic nervous system activity [8][9]; electroencephalography (EEG), which records electrical brain activity associated with various cognitive and emotional states [10][11].

These signals are informative and have demonstrated effectiveness in both clinical and entertainment VR applications [12][13]. However, they typically require wearable sensors, careful calibration, and noise control, which complicates their practical use in active VR scenarios. As an alternative and promising direction in affective computing, behavioral signals – particularly motion data – have shown potential for inferring emotional state without additional invasive equipment [14]. Head pose dynamics, hand gestures, and overall movement patterns correlate with engagement, anxiety, and relaxation [15].

Contemporary VR studies consider behavioral analytics for detecting stress based on indicators such as abrupt head movements, increased gesture frequency, or constrained posture. Behavioral data have several advantages: they can be collected passively without extra hardware; they are robust to momentary occlusions; and they scale across users and contexts.

Meta (Oculus) Quest 2 is selected as the main device for data acquisition because it combines a performant platform with inside-out head and hand tracking, making it possible to capture high-frequency motion data – position and orientation of the head and controllers, as well as derived kinematic features – for multidimensional analysis of user behavior [16][17][18]. We propose to develop this approach by formalizing a system that uses behavioral signals – specifically head and hand motion – to infer emotions in real time. The system is designed to function online, which is critical in dynamic VR environments.

2.2. Motion analysis in human–computer interaction (HCI)

Prior work shows that kinematic parameters – speed, acceleration, curvature of trajectories – can reflect cognitive activity and the user's emotional state [19][20][21]. In traditional HCI settings (e.g., desktop), motion analysis has been applied to mouse dynamics, touch gestures, and gaze for authentication, workload estimation, and adaptive UI design [22][23]. In immersive media, the expansion of three-dimensional interaction opens new avenues for interpreting user state through movement.

Analyzing HCI studies focused on embodied interaction, researchers highlight the relationship between movement patterns and user performance in VR tasks [24]. Data-driven models capture characteristic patterns, such as increased head rotation variance during stress, or smoother hand trajectories during focused attention [25][26]. Motion complexity has also been linked to mental load: higher task difficulty can be reflected in increased jerk, trajectory irregularity, and reduced smoothness, enabling the construction of workload estimators based on these features.

In collaborative or social VR scenarios, indicators such as gaze direction, interpersonal distance, and turn-taking gestures help reveal social and affective aspects of interaction in HCI. Systems can predict intent and anticipate user actions based on motion patterns – for example, when a user attempts to grasp an object or interact with a virtual menu [27].

2.3. Machine learning for emotion recognition

ML methods are typically divided into supervised and unsupervised approaches. Supervised learning trains a model on labeled data to map input features to target emotional categories. Unsupervised learning, by contrast, seeks to discover latent structure – clusters or manifolds – without explicit labels, helping to identify underlying behavioral patterns that may correspond to affective states [28].

Common supervised algorithms for classification include:

- Support Vector Machines (SVM), effective for high-dimensional spaces, especially in small-sample regimes with robust margin-based generalization; SVMs have shown reliability in classifying emotions from facial expressions and motion patterns [29];
- Random Forest (RF), a tree-based ensemble method resistant to overfitting and noise; RF effectively predicts arousal levels and distinguishes behavioral markers such as body-movement frequency or gesture rate [30];
- Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN), particularly effective for spatiotemporal data; CNNs capture local patterns and temporal context in time series, including head and hand motion velocities in VR [31];
- Recurrent Neural Networks (RNN) [32] and LSTM networks [33], designed for sequential data where temporal dependencies are critical; they can track the dynamics of emotional state over time.

Selecting an algorithm for online multiclass affective classification requires balancing accuracy, latency, interpretability, and robustness. The decision depends on dataset specifics and latency constraints [6][7].

Emotional patterns can change substantially in dynamic environments – due to individual differences and context shifts – leading to "concept drift." Models trained offline may degrade over time when user behavior changes, reducing performance on previously learned emotions [34][35]. We therefore propose incremental online learning as a practical compromise: it supports adaptation to new data distributions while maintaining a stable representation of previously learned classes. Additionally, unsupervised clustering is recommended as a preliminary stage to identify behavioral movement segments; this facilitates subsequent supervised learning and provides a deeper understanding of affective states.

In this work, the focus is on the kinematic behavioral data in VR which are well suited to online acquisition and low-latency inference. Combining established ML models with behavioral motion data extends the boundaries of affective computing in VR. This integration enables emotionally adaptive systems that can respond to affective cues with low latency and high accuracy.

3. System Architecture

The system for recognizing emotional states in VR is organized as a low-latency, modular pipeline that prioritizes minimal algorithmic complexity, efficient resource usage, and unobtrusiveness for the user. System high level architecture is presented on figure 1

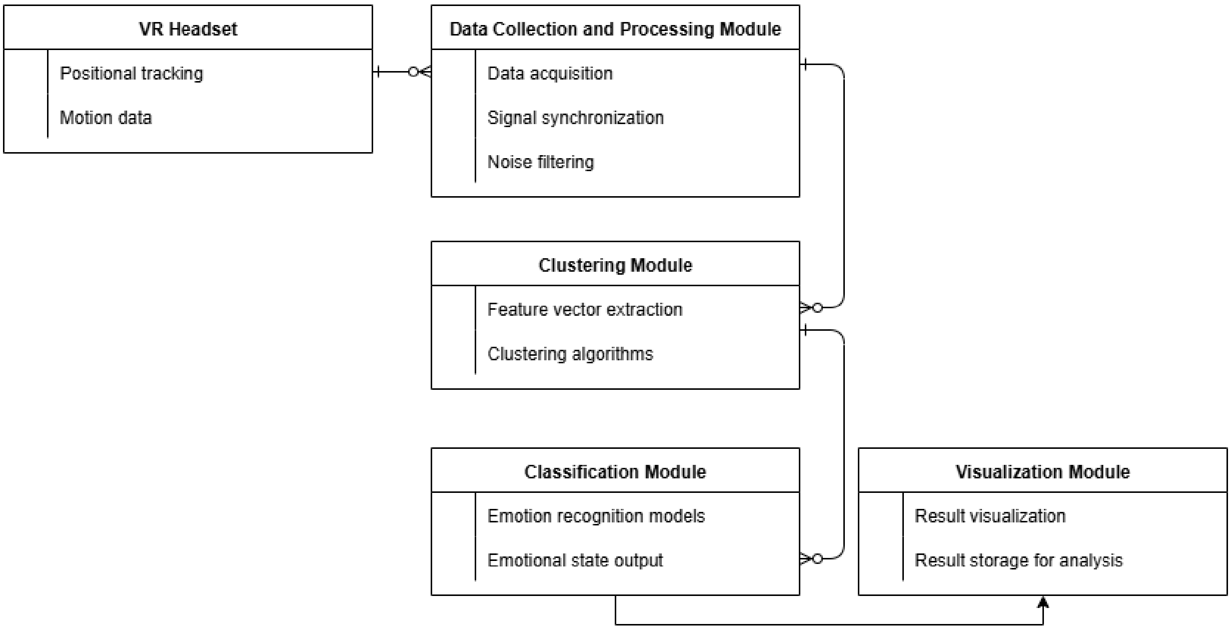

Figure 1: System high level architecture.

The system consists of five core components:

- Motion data acquisition module;
- Preprocessing and feature-extraction module;
- Clustering mechanism for discovering behavioral patterns;
- Affective classification mechanism;
- Real-time feedback interface.

For ease of integration and simplicity of use, the system employs the Oculus (Meta) Quest 2 platform, which provides an optimal balance among performance, cost, and sensor accuracy. Data streams collected in real time via the Oculus SDK.

Each motion frame is accompanied by a precise timestamp, enabling exact alignment of data across sensors for synchronization.

To support both streaming and batch processing, the system uses a sliding-window approach with overlap. This preserves the temporal context while enabling real-time processing.

Each window segment contains an n x m matrix (n = window length in frames, m = number of variables) with base signals (node positions and orientations; head and hands kinematics) and derived features (velocity, acceleration, gesture-change indicators).

To improve accuracy and reduce noise, the module uses:

- Exponential moving average (EMA), applied to all numeric columns within the window;
- Outlier detection to exclude spikes caused by tracking losses or brief occlusions, with imputation by the median of the corresponding variable within the same window;
- Gesture stabilization that ignores short, non-salient signals and reduces isolated state flips without blurring the boundaries of meaningful events.
- These measures increase the reliability of behavioral-pattern detection, especially in dynamic and potentially noisy VR environments.

For training, validation, or debugging, the module supports session logging in CSV/Parquet formats. Each session contains:

- Metadata: pseudonymized user ID, task scenario, VR-application context;
- Start and end timestamps;
- Annotated markers (optional): emotion labels, events.

The data-collection module is optimized for low-latency operation (stable high-frequency capture and efficient buffering), forming a reliable foundation for all real-time affective inference tasks.

4. Feature Vector Construction

Constructing the feature vector directly influences both unsupervised cluster discovery and supervised classification for emotional-state estimation.

The process analyzes time-segmented motion windows and transforms raw 6-DoF signals into a compact feature vector that describes the dynamic properties of head and hand movements.

The features are grouped by functional categories listed in Table 1.

A comprehensive processing of the initial feature set was carried out, including smoothing, normalization, windowing, and extraction of both time-domain indices (moments, rates, ranges) and spectral characteristics (dominant frequencies, spectral energy) [36].

Table 1
Features functional categories

Category		Characteristics	Notation
A	Kinematic features (physical movement of the user)	Mean speed of the head and of each hand; Standard deviation of acceleration; Maximum angular velocity (head rotation); Range of motion (Euclidean distance traveled within the window); Jerk (change in acceleration reflecting movement abruptness).	$\bar{v}$ σa ωmax $R = max(x) - min(x)$ $j = da/dt$
B	Spatial orientation and posture	Histograms of pitch, yaw, roll (distribution of head orientation); Spatial relations between hands and headset (e.g., mean distance between a hand and the head); Height asymmetry of the hands (may indicate tension or dominance).	$h(\theta)$ $dH{-}H$ $\Delta y = yL - yR$
C	Gesture dynamics	Number of gesture-state changes within a window; Distribution of gesture durations (dwell time in states such as grip, pointing, rest, etc.); Frequency of specific gestures (e.g., waving, pointing, clenched fist).	Ntrans tgesture ρgesture

D	Frequency-domain features	Fast Fourier Transform (FFT) coefficients;
		Dominant frequency components (characterize movement periodicity);
		Spectral energy density (distribution of movement energy over frequencies);
			fdom
			Ef
			ρgesture

		Rhythmic descriptors.	
E	Frequency-domain features	Mean, median, variance;	μ, $\bar{x}$, $\sigma2$
		Skewness;	$\gamma1$
		kurtosis (“peakedness” of the distribution).	$\gamma2$

To stabilize model training, z-score standardization is applied; to reduce redundancy, Principal Component Analysis (PCA) is used. The initial data analysis revealed vector fields within the range $[-5.9, +2.8]$, reflecting the normalized nature of the data.

The rationale for the emotional relevance of the features is based on interdisciplinary findings from psychology and nonverbal communication showing that gestures and posture are key channels of emotional transmission [37]. High acceleration and frequent changes in body direction are reliable markers of stress and anxiety; conversely, a limited range of motion and infrequent, muted gestures correlate with boredom and disengagement. Open-hand gestures and a forward-leaning posture are characteristic of engagement and joy, ensuring effective emotional and communicative expression [37].

5. Clustering of Behavioral Movements

A key element of the clustering module is the discovery and formation of behavioral movement patterns. This is necessary for two main reasons: first, it enables detecting unlabeled recurring movement patterns that may indicate emotional state; second, it simplifies input-data analysis by structuring similar behavioral sequences before subsequent classification. The unsupervised stage is intended to uncover latent emotional dynamics without immediate labeling or predefined emotional templates.

The clustering module processes the feature vectors corresponding to 2-second motion segments and groups them into meaningful behavioral clusters using scalable, low-latency algorithms.

5.1. Goal of clustering and approach

The primary goal of the clustering module is to create behavioral clusters, i.e., groups of movement sequences with similar dynamical and spatial characteristics. These clusters can later be labeled (offline) with emotion tags based on observer ratings or user self-reports, thereby producing training data for classification.

Clustering must meet several critical requirements: support incremental learning for real-time adaptation; impose low computational load so that a high frame rate is maintained in the VR environment (at least 90 FPS); and detect complex, non-linear, and asymmetric structures characteristic of natural human motor behavior, which rarely conforms to ideal geometry.

5.2. Comparison of clustering algorithms

To meet the module’s objectives, four principal clustering methods were considered and implemented for motion segmentation: k-means; density-based DBSCAN; hierarchical agglomerative clustering (HAC); and online vector quantization (OVQ).

K-means [38] is a baseline clustering algorithm that partitions data into k groups by distance to cluster centers (means). It iteratively updates centroids to minimize the sum of squared distances between points and their assigned centers. Effective when clusters are approximately spherical with similar variance, k-means is simple and fast for clearly separated, symmetric clusters, but degrades on complex or noisy data. For real-time use, the optimized Mini-Batch k-means variant is often applied to reduce computational load.

Gaussian Mixture Models (GMM) is a probabilistic approach that represents data as a mixture of multivariate normal components with different parameters (means, covariance matrices, and weights). Expectation–Maximization (EM) iteratively estimates point-to-cluster probabilities (E-step) and updates component parameters (M-step) until convergence. GMM performs "soft" clustering (probabilistic membership), can model clusters of arbitrary elliptical shape and orientation, and works well on complex data. Limitations include the need to predefine the number of clusters, sensitivity to initialization, and reduced robustness in the presence of outliers or heavy noise [39].

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [40] groups points that are sufficiently dense in space. It does not require a pre-set number of clusters and can automatically identify both dense regions (clusters) and sparse regions (outliers). DBSCAN is effective for behavioral patterns with uneven density or pauses in motion, though it is sensitive to hyperparameter choice, which strongly affects quality.

Hierarchical Agglomerative Clustering (HAC) [41] merges the closest pairs of objects or clusters based on pairwise distances until all objects are combined into one cluster. Results can be displayed as a dendrogram that visualizes hierarchical structure and supports selecting the desired level of detail. HAC is informative but computationally expensive, and is typically used for offline analysis or small datasets.

Online Vector Quantization (OVQ) [42] incrementally updates cluster centers in response to streaming inputs. Each incoming vector is approximated by the nearest centroid, which is then adjusted by the new sample. Owing to its simplicity and efficiency, OVQ suits resource-constrained, real-time environments and supports adaptation to concept drift. Its lightweight nature makes it a good fit for VR platforms with changing behavioral patterns.

5.3. Analysis of clustering evaluation methods

To evaluate clustering effectiveness, the following internal metrics were used.

5.3.1. Silhouette Coefficient

Silhouette Coefficient (measures cluster separability) [43]:

$$S(i)=\frac{b(i)-a(i)}{max\{a(i),b(i)\}} \tag{1}$$

where *a(i)* is the average distance from point *i* to all other points in its own cluster; *b(i)* is the minimal average distance from *i* to all points in the nearest neighboring cluster.

5.3.2. Calinski–Harabasz Index

Calinski–Harabasz Index [44] (ratio of between- to within-cluster dispersion):

$$CH=\frac{BGSS(k-1)}{WGSS(n-k)} \tag{2}$$

with BGSS

$$BGSS=\sum_{k=1}^{k} n_k \| c_k - c \|^2 \tag{3}$$

where *BGSS* – between-cluster sum of squares, with n_k is the number of observations in cluster *k*, c_k is the centroid of cluster *k*, *c* is the dataset centroid;

$$WGSS_k=\sum_{i=1}^{n_k} \| x_{ik} - c_k \|^2 \tag{4}$$

where $WGSS_k$ – within-cluster sum of squares for cluster *k*,

$$WGSS=\sum_{k=1}^{K_{\square}} W\,G\,S\,S_k \tag{5}$$

where *k* is the number of clusters.

5.3.3. Davies–Bouldin Index

Davies–Bouldin Index [45] (lower is better):

$$DB=\frac{1}{k}\sum_{i=1}^{k} m\,a\,x\left(\frac{\Delta(x_i)+\Delta(x_j)}{\delta(x_i,x_j)}\right) \tag{6}$$

where $\Delta(x_k)$ is the inner-cluster distance within cluster x_k; $\delta(x_i,\ x_j)$ is the inter-cluster distance between the centroids of clusters x_i and x_j.

5.4. Identification of emotional patterns and integration the classification module

To improve emotion identification, we adopted a movement-pattern-oriented approach and obtained a balanced dataset with diverse emotional states exhibiting varied behavioral manifestations. A controlled VR session was recorded in which the participant performed movements typical for each state (e.g., active dance-like movements for joy; calm and smooth for focus; soft/slow for boredom; nervous and rapid for anxiety; aggressive for stress), and the behavior was captured for subsequent annotation. Using these data, we analyzed and classified movement patterns (by amplitude, speed, and spectral characteristics) for each emotion and formulated a mapping rule based on their statistical profiles, providing a more objective and representative attachment of clusters to emotional classes for subsequent supervised learning and automatic recognition.

Based on the results, we built a complete process combining Gaussian Mixture Models (GMM) clustering with a domain-specific empirical emotion-labeling scheme. Instead of generic algorithms that often introduce noise or converge to uniform outcomes, GMM was used to identify three key clusters in a 10-minute VR dataset:

- Cluster 0 (334 samples): concentrated interaction with low variability and a stable movement pattern.
- Cluster 1 (90 samples): calm/static standing with minimal intensity.
- Cluster 2 (139 samples): vigorous, active movements with high intensity and variability.

For each cluster, detailed statistics were computed, including mean amplitude, speed, dispersion, pattern-stability index, and the dynamism of changes between neighboring frames. A domain-specific mapping was designed: Cluster 0 ⟶ “focus/engagement” (confidence 0.92), Cluster 1 ⟶ “boredom” (confidence 0.88), Cluster 2 ⟶ “joy/activity” (confidence 0.90). Each assignment was justified by real user observations and then calibrated by cluster size and pattern stability. A temporal check confirmed the logical sequence of emotional transitions, e.g., “focus ⟶ focus ⟶ boredom ⟶ joy ⟶ joy.”

The final dataset was augmented with emotion label, emotion name, and mapping confidence. The structure (563 windows ⟶ 337 cleaned samples) includes 24 PCA components and the cluster identifier, yielding a ready-to-train multi-class set with three balanced emotion classes. This clustering-and-labeling scheme establishes a solid basis for subsequent supervised classifiers (Random Forest, SVM, 1D-CNN) and for real-time emotion recognition in VR systems.

6. Classification Pipeline

After grouping behavioral patterns and determining emotional templates, the critical stage is emotion classification, which ensures online inference, stability under noise, and low latency suitable for VR applications.

6.1. Goal of classification and approach

The main task of classification is to implement a mapping function that transforms a high-dimensional feature vector x ∈ $\mathbb{R}^n$ into a categorical label from the finite set $\{e_1, \ldots, e_k\}$, where e_i is one of the predefined affective classes [46].

In our case, the model recognizes five emotional states:

- Joy / happiness
- Focus / engagement
- Boredom
- Anxiety
- Stress

Here x_i is the resulting feature vector representing a 24-dimensional space of principal components combined with the original normalized features and corresponding weights; y_i is the associated emotion label (a scalar).

To solve the classification problem, three models were employed due to their balance of accuracy, interpretability, and suitability for real-time use:

- Random Forest (RF)
- Support Vector Machine (SVM)
- One-dimensional Convolutional Neural Network (1D-CNN)

Random Forest is an ensemble method based on a set of decision trees trained on bootstrap samples with feature subsampling; the final prediction is obtained by majority voting. RF is robust to noise, handles heterogeneous features well, and is often used as a baseline in tasks related to behavioral analysis.

The Support Vector Machine (SVM) is a kernel model effective for high-dimensional spaces and relatively small datasets; by maximizing the margin between classes, it provides stable generalization. However, inference speed may decrease on very large or streaming datasets.

The one-dimensional Convolutional Neural Network (1D-CNN) is a deep model that captures local temporal patterns in motion time series, providing high accuracy – especially with GPU acceleration – while maintaining low latency in windowed inference.

The cluster-based labeling obtained in the unsupervised stage provides training data for the supervised classifiers (RF, SVM, 1D-CNN). This mapping enables objective alignment of movement patterns with affective classes and supports validation of the models' generalization and practical applicability.

6.2. Evaluation metrics for multi-class classification

Evaluation metrics for multi-class classification use the standard definitions:

6.2.1. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

where *TP* – is the number of true positives, *FN* – the number of false negatives [47].

6.2.2. Precision

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

where *TP* – is the number of true positives, *FP* – the number of false positives [48].

6.2.3. Recall

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

where *TP* – is the number of true positives, *FN* – the number of false negatives [48].

6.2.4. F1 Score

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

6.2.5. Latency

$$L = \frac{1}{N} \sum_{i=1}^{N} t_{pred,i} \tag{11}$$

where t_i is the processing time of the *i*-th sample, and *N* is the total number of samples [49].

6.2.6. Robustness

$$R = 1 - \frac{|M_{original} - M_{noisy}}{M_{original}} \tag{12}$$

where $M_{original}$ is the model metric on clean data and M_{noisy} the metric on data with injected noise [50].

To ensure compatibility with real-time VR environments, the final (1D-CNN) model was implemented with optimized inference so that the end-to-end per-window processing time remains below the target threshold (less than 30 ms, enabling updates above 30 Hz).

7. Evaluation

To assess the effectiveness, reliability, and suitability of the proposed system, an experimental study was conducted. The analysis covered three main aspects:

- Classification accuracy – how precisely the system recognizes emotional states
- Computational efficiency – whether the system meets VR response-time and frame-rate requirements
- Robustness and generalizability – how well the system adapts to different users and scenarios

Equipment: Oculus Quest 2 (90 FPS), connected via Oculus Link to a PC with an NVIDIA RTX 4060 GPU.

Emotion labeling: self-assessment using the SAM (Self-Assessment Manikin) scale.

Classification accuracy: the parameters of the models (RF, SVM, 1D-CNN) were evaluated using stratified 5-fold cross-validation, details presented in table 2.

Table 2
Classification accuracy

Model	Accuracy	Precision	Recall	F1 Score	Latency (ms)
RF	0.976	0.969	0.968	0.968	0.001-0.005
SVM	1.000	1.000	1.000	1.000	0.001-0.005
1D-CNN	0.976	0.968	0.968	0.968	0.001-0.01

The results confirm that emotional information is distributed across multiple movement modalities, with gesture dynamics being particularly informative. The system achieves high emotion-recognition accuracy using only movement data, meets real-time requirements (<30 ms) and demonstrated stability and efficiency in a practical VR environment.

8. Comparative Analysis with Baseline Methods

To position the proposed emotion-recognition system within the broader field, we compare it with conventional baseline approaches that use physiological signals – such as galvanic skin response (GSR), heart rate and heart rate variability (HR/HRV), and electroencephalography (EEG) – to determine a person's emotional state. Despite their high accuracy, these methods have several limitations: they require additional hardware and careful calibration; they reduce comfort and mobility due to cabled or wearable sensors; they increase setup time and cost; and they complicate interactive real-time feedback in active VR scenarios. Table 3 shows comparison of a user emotion-recognition model based on biosignals (baseline) versus a movement-based model (proposed).

Table 3
Comparison of a biosignals model versus a movement-based model

Criterion	Movement-based model (proposed)	Biosignal-based model (baseline)
Invasiveness	Low – no extra sensors required	High – electrodes/sensors needed
Dependence on equipment	Built-in Meta Quest 2 tracking	Specialized sensors (ECG, GSR, EEG)
Setup time	< 2 minutes	10–15 minutes (including calibration)
Real-time operation	Native (compatible with 90 FPS)	Depends on signal-processing pipeline
Comfort / mobility	Full mobility, wireless	Limited by cables and wearables
Scalability	High (consumer hardware)	Moderate (rising cost and complexity)
Generalizability	Robust across users/scenarios	Often requires per-user tuning
Maintenance	None	Sensor cleaning, charging required
Privacy	Motion only, no biometrics	Medical data; requires protection (GDPR, IRB)

The movement-based model built on Meta Quest 2's integrated tracking provides a fast, low-friction solution that preserves mobility and minimizes privacy risk by using only motion data. In contrast, biosignal approaches typically deliver higher diagnostic power in controlled or clinical settings but at the cost of practicality in consumer-scale, real-time VR use. For clarity, we also provide a scenario-to-approach mapping. It should be noted that movement- and biosignal-based models are not mutually exclusive. Hybrid systems, which combine physiological measurements with compact sets of motion features, often achieve higher accuracy than any single-modality approach. Future work includes integrating physiological indicators in an optional hybrid configuration while retaining the proposed movement-only pathway for scenarios demanding the best balance of accuracy, convenience, and scalability.

9. Limitations and Future Work

Despite the promising results and practical suitability of the proposed motion-based VR emotion-recognition system, several constraints remain that delimit the scope of the present conclusions and open prospects for further research and refinement.

9.1. Limitations

9.1.1. Subjectivity in emotion labeling

The emotion labels used in training and evaluation – obtained from self-reports and/or observer annotations – are susceptible to inter-rater variability and contextual bias. Reducing labeling subjectivity and ensuring consistency across sessions and users remains a difficult task in affective recognition. Also, broader validation across demographic and psychological spectra is still required. The robustness of the model for different user groups and contexts needs confirmation on a wider sample.

9.1.2. Limited set of affective categories

At present, the system recognizes only five states. Extending the taxonomy (including dimensional models) would better reflect the complexity of emotional experience. Also, the system relies on clearly expressed, classifiable behavioral patterns. Minimal-movement interactions and "quiet" states may provide limited signal for inference. Motion alone may be an insufficient source of emotional information, especially for such "quiet" conditions. To improve adaptation to new users and changing environments, we will explore semi-supervised schemes and continuous updates that enable emotion recognition with minimal user involvement.

10. Conclusion

The study demonstrates that user emotions in VR can be recognized accurately using motion data alone, without auxiliary physiological sensors. The full pipeline operates within strict real-time constraints (under 30 ms per window) and integrates smoothly with interactive VR applications, enabling timely affect-aware responses during immersion. The approach segments short-horizon windows of 6DoF head/hand motion, applies robust preprocessing, constructs a principled descriptor set, compresses it via PCA, reveals latent behavioral structure with clustering, and maps segments to five affective categories with lightweight classifiers. DBSCAN outperforms MBKM under internal validity metrics, consistent with non-spherical, noisy manifolds of human movement. Across varied tasks and participants, the models generalize to new users and scenarios while maintaining stable behavior and computational efficiency. These properties make the approach practical for deployment in real VR environments and establish a reliable basis for building emotionally adaptive experiences. Future work should extend to semi/self-supervised objectives for label efficiency, personalization with safe online updates, domain adaptation across devices/contexts without sacrificing the core simplicity of motion-only sensing.

Declaration on Generative AI

The authors did not use any generative AI tools.

References

[1] X. Zhang, Q. Yan, S. Zhou, L. Ma, S. Wang. Analysis of Unsatisfying User Experiences and Unmet Psychological Needs for Virtual Reality Exergames Using Deep Learning Approach. Information, 2021, pp. 480-486. doi:10.3390/info12110486.

[2] R. G. Africa, M. Rosario Gonzalez-Rodriguez, M. Carmen Diaz-Fernandez. Salient features and emotions elicited from a virtual reality experience: the immersive Van Gogh exhibition. Quality & Quantity (2023) doi:10.1007/s11135-023-01752-2.

[3] G. Lampropoulos, E. Keramopoulos, K. Diamantaras, G. Evangelidis. Augmented Reality and Virtual Reality in Education: Public Perspectives, Sentiments, Attitudes, and Discourses. Educ. Sci., 2022, pp. 780-798. doi:10.3390/educsci12110798.

[4] B. G. P. Linares-Vargas, S. E. Cieza-Mostacero. Interactive virtual reality environments and emotions: a systematic review. Virtual Reality (2025). doi:10.1007/s10055-024-01049-1.

[5] R. Picard. Affective Computing. Cambridge, MA, MIT Press, 1997.

[6] Y. Cai, X. Li, J. Li. Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review. Sensors (2023), 2455. doi:10.3390/s23052455.

[7] A. Dzedzickis, A. Kaklauskas, V. Bucinskas. Human Emotion Recognition: Review of Sensors and Methods. Sensors (2020) 592. doi:10.3390/s20030592.

[8] N. Hinricher, S. Konig, C. Schroer and C. Backhaus. Effects of virtual reality and test environment on user experience, usability, and mental workload in the evaluation of a blood pressure monitor. Front. Virtual Real. (2023) 190. doi:10.3389/frvir.2023.1151190.

[9] S. Grossberg. Recurrent Neural Networks. Scholarpedia (2013) 188. doi:10.4249/scholarpedia.1888.

[10] E. Gkintoni, A. Aroutzidis, H. Antonopoulou, C. Halkiopoulos, From Neural Networks to Emotional Networks: A Systematic Review of EEG-Based Emotion Recognition in Cognitive Neuroscience and Real-World Applications. Brain Sci. (2025) 220. doi: 10.3390/brainsci15030220.

[11] H. A. Hamzah, K. K. Abdalla. EEG-Based Emotion Recognition Datasets for Virtual Environments (2024). doi:10.1155/2024/6091523.

[12] J. Marin-Morales, J. L. Higuera-Trujillo, A. Greco, J. Guixeres, C. Llinares, E.P. Scilingo, M. Alcaniz and G.Valenza. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. Scientific Reports (2018) 13657. doi:10.1038/s41598-018-32063-4.

[13] M. Li, J. Pan, Y. Li, Y. Gao, H. Qin and Y. Shen. Multimodal physiological analysis of impact of emotion on cognitive control in VR, IEEE Transactions on Visualization and Computer Graphics, 2024, pp. 2044-2054. doi:10.1109/TVCG.2024.3372101.

[14] J.-P. Tauscher, F. W. Schottky, S. Grogorick, P. M. Bittner, M. Mustafa and M. Magnor. Immersive EEG: Evaluating electroencephalography in virtual reality, in: Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces, 2019, pp. 91–199. doi:10.1109/VR.2019.8797858.

[15] D. Andreoletti, M. Paoliello, L. Luceri, T. Leidi, A. Peternier, S. Giordano. A framework for emotion-driven product design through virtual reality. Inf Technol Manag, 2022, pp. 42–61. doi:10.1007/978-3-030-98997-2_3.

[16] M. Slater, C. Cabriera, G. Senel, D. Banakou, A. Beacco, R. Oliva, J. Gallego. The sentiment of a virtual rock concert. Virtual Reality, 2023, pp. 651–675. doi:10.1007/s10055-022-00685-9.

[17] D. Abdlkarim, M. Di Luca, P. Aves, M. Maaroufi, S.-H. Yeo, R.C. Miall, P. Holland, J.M. Galea. A methodological framework to assess the accuracy of virtual reality hand-tracking systems: A case study with the Meta Quest 2. Behaviour Research Methods, 2023, pp. 1052-1063. doi:10.3758/s13428-022-02051-8.

[18] A. Carnevale, I. Mannocchi, M. S. Hadj Sassi, M. Carli, G. De Luca, U.G. Longo, V. Denaro, E. Schena. Virtual Reality for Shoulder Rehabilitation: Accuracy Evaluation of Oculus Quest 2. Sensors, 2022, 5511. doi:10.3390/s22155511.

[19] J. W. Rankin, W. M. Richter and R. R. Neptune. The influence of obesity on walking and cycling biomechanics and muscle activation patterns. Clinical Biomechanics, 2014, pp. 1021-1027. doi:10.1007/s00221-012-3357-4.

[20] E. E. Caron, L. R. Marusich, J. Z. Bakdash, R. J. Ballotti, A. M. Tague, J. S. A. Carriere, D. Smilek, D. Harter, S. Lu and M.G. Reynolds. The Influence of Posture on Attention. Experimental Psychology, 2023, pp. 295-307. doi:10.1027/1618-3169/a000567.

[21] E. Wiese, A. Wykowska, J. Zwickel and H. J. Muller. I See What You Mean: How Attentional Selection Is Shaped by Ascribing Intentions to Others. PLoS ONE, 2012. doi:10.1371/journal.pone.0045391.

[22] K. Tzafilkou and N. Protogeros. Mouse behavioral patterns and keystroke dynamics in End-User Development: What can they tell us about users' behavioral attributes? Computers in Human Behavior, 2021, pp. 34-47. doi:10.1016/j.chb.2018.02.012.

[23] D. Tian, S. Zhang, S. Chen, Y. Zhang, K. Peng, H. Zhang and D. Wang. Tracking dynamic flow: Decoding flow fluctuations through performance in a fine motor control task. IEEE Transactions on Affective Computing, 2023. doi:10.48550/arXiv.2310.12035.

[24] H. Herrebroden, A. R. Jensenius, T. Espeseth, L. Bishop and J. K. Juoskoski. Cognitive load causes kinematic changes in both elite and non-elite rowers. Human Movement Science, 2023, pp. 103-113. doi:10.1016/j.humov.2023.103113.

[25] B. S. Hasler, G. Hirschberger, T. Shani-Sherman and D. A. Friedman, D. A. Virtual Peacemakers: Mimicry Increases Empathy in Simulated Contact with Virtual Outgroup Members. Cyberpsychology, Behavior, and Social Networking, 2024, pp. 766-771. doi:10.1089/cyber.2014.0213.

[26] T. Pejsa, M. Gleicher and B. Mutlu. Who, Me? How Virtual Agents Can Shape Conversational Footing in Virtual Reality. Intelligent Virtual Agents, 2017, pp. 334-344. doi: 10.1007/978-3-319-67401-8_45.

[27] N. M. Gamage, D. Ishtaweera, M. Weigel and A. Withana. So Predictable! Continuous 3D Hand Trajectory Prediction in Virtual Reality, in: Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology, 2021, pp. 522-533. doi:10.1145/3472749.3474753.

[28] M.S. Akhtar, D. Ghosal, A. Ekbal, P. Bhattacharyya and S. Kurohashi. All-in-One: Emotion, Sentiment and Intensity Prediction Using a Multi-Task Ensemble Framework. IEEE Transactions on Affective Computing, 2022, pp. 285-297. doi:10.1109/TAFFC.2019.2926724.

[29] A. Mariette, K. Rahul. Support Vector Machines for Classification. Efficient Learning Machines, 2015, pp. 39–66. doi:10.1007/978-1-4302-5990-9_3.

[30] L. Breiman. Classification and Regression Trees. New York: Routledge. 2017. doi:10.1201/9781315139470.

[31] Y. LeCun, Y. Bengio, G. Hinton. Deep learning. Nature, 2015, pp. 436–444. doi:10.1038/nature14539. ISSN 1476-4687.

[32] S. Grossberg. Recurrent Neural Networks. Scholarpedia. 2013. doi:10.4249/scholarpedia.1888.

[33] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, J. Schmidhuber. A Novel Connectionist System for Improved Unconstrained Handwriting Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, pp. 855–868. doi:10.1109/tpami.2008.137.

[34] J. Gonzalez, L. Prevost. Personalizing emotion recognition using incremental random forests. EUSIPCO, 2021. doi:10.23919/EUSIPCO54536.2021.9616296.

[35] C. Du, K. Fu, J. Peng, S. Zhao, X. Chen, H He. EmoGrowth: Incremental Multi-label Emotion Decoding with Augmented Emotional Relation Graph. ICML, 2025. URL: https://openreview.net/forum?id=b2fhCbhe62.

[36] Y. Bodyanskiy and N. Kulishova. Multidimensional Neuro-fuzzy System and Fuzzy Coding for a Constant Length Facial Landmark Set Formation. 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2021, pp. 166-171. doi:10.1109/IDAACS53288.2021.9660907.

[37] V. Vinayagamoorthy, A. Steed and M. Slater. Building Expression into Virtual Characters. Eurographics Conference State of the Art Reports, 2006, pp. 21-61. doi:10.1109/TVCG.2005.79.

[38] J. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations. Fifth Berkeley Sympo-sium on Mathematical Statistics and Probability (1967) 281–297.

[39] R. Zhao, Y. Li and Y. Sun. Statistical convergence of the EM algorithm on Gaussian mixture models. Electronic Journal of Statistics, 2020, pp. 632-660. doi:10.1214/19-EJS1660.

[40] M. Ester, H.-P. Kriegel, J. Sander, X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. Second International Conference on Knowledge Discovery and Data Mining, 1996, pp. 226–231. doi:10.5555/3001460.3001507.

[41] Jr. Ward, H. Joe. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association, 1963, pp. 236–244. doi:10.2307/2282967.

[42] R. M. Gray. Vector Quantization. IEEE ASSP Magazine. 1984, pp. 4–29. doi:10.1109/massp.1984.1162229.

[43] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 1987, pp. 53-65. doi:10.1016/0377-0427(87)90125-7.

[44] T. Calinski and J. Harabasz. A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods, 1974, pp. 1-27. doi:10.1080/03610927408827101.

[45] D. L. Davies and D. W. Bouldin. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1979, pp. 224-227. doi:10.1109/TPAMI.1979.4766909.

[46] Y.V. Bodyanskiy, N. Y. Kulishova and V. P. Tkachenko. Feature vector generation for the facial expression recognition using neo-fuzzy system. Radio Electronics, Computer Science, 2018. doi:10.15588/1607-3274-2018-3-10.

[47] N. Yager and T. Dunstone. The biometric menagerie. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, pp. 220–230. doi:10.1109/TPAMI.2008.291.

[48] L. Itti, C. Koch and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, pp. 1254–1259. doi:10.1109/34.730558.

[49] K. Fawagreh and M. M. Gaber. Resource-efficient fast prediction in healthcare data analytics: A pruned Random Forest regression approach. Computing, 2020, pp. 1187–1198. doi:10.1007/s00607-019-00785-6.

[50] Y. Huang, I. King, T.Y. Liu and M. Steen. WWW '20: Proceedings of The Web Conference 2020. The Web Conference, 2020.doi:10.1145/3366423.