Automated Pose Classification for Behavioral Analysis in Therapeutic Settings

Liubov Revutska[1,2], Bei-Xuan Lin[1,2], Emily B. Wake[1,2], Nastia Junod[3], Jennifer Glaus[3], Olga Sidiropoulou[3], Kerstin J. Plessen[3], Micah M. Murray[1,2*], Naomi K. Middelmann[1,2,3*], Matthew J. Vowels[1,2*]

[1]The Radiology Department, Lausanne University Hospital Center and University of Lausanne (CHUV-UNIL), Lausanne, Switzerland
[2]The Sense Innovation and Research Center, Lausanne and Sion, Switzerland
[3]Division of Child and Adolescent Psychiatry, Department of Psychiatry, CHUV-UNIL, Lausanne, Switzerland
*These authors share senior authorship and contributed equally to this work

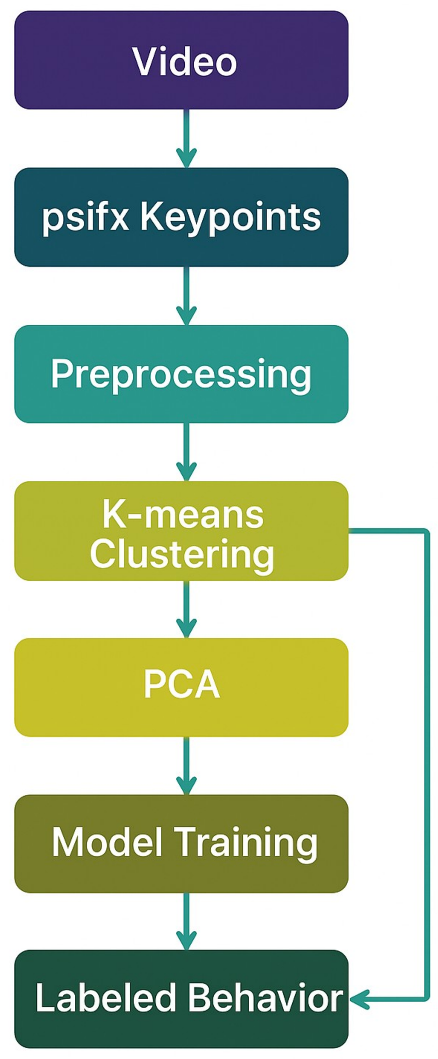
the sense
innovation and research center

Background

Quantitative pose analysis enables objective, reproducible assessment of nonverbal behaviors in real-world therapeutic settings. Scalable analysis of these embodied interactions can support data-driven understanding of therapy dynamics and functional efficacy for individual patients.
We leveraged the *psifx*[1] framework to extract structured pose and segmentation features from video recordings, forming the foundation for automated behavioral classification.

Aims

- **Develop** a modular, end-to-end pipeline for automated classification of therapist and patient poses in therapy sessions.
- **Validate** the model's performance across diverse interaction scenarios.
- **Establish** a scalable framework for large-scale, data-driven analysis of embodied therapeutic interactions.



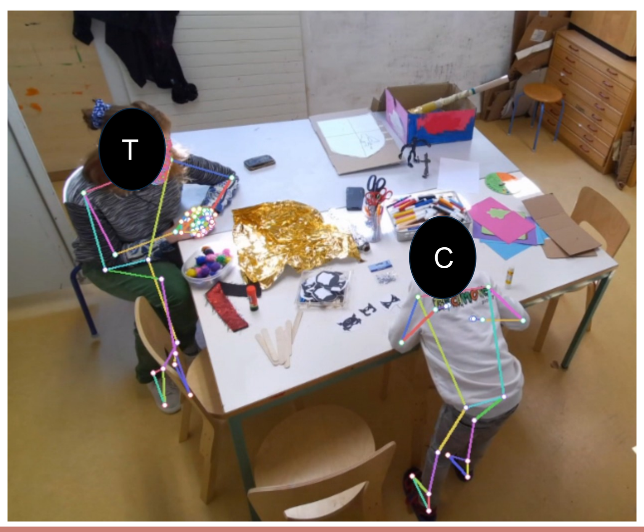
Methods

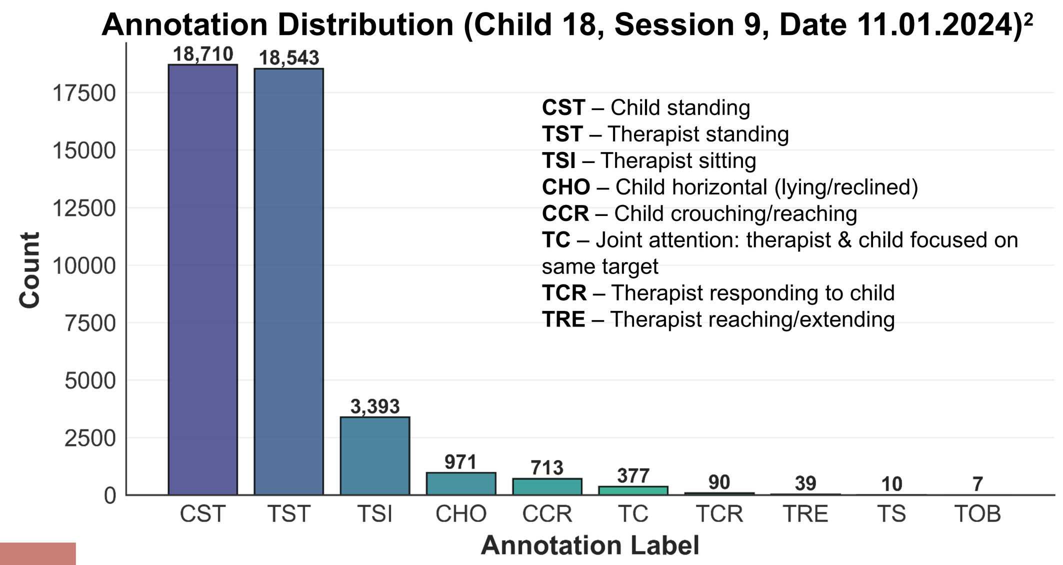
- **Feature Extraction:** Pose skeletons and segmentation masks were obtained using the psifx framework and consolidated into standardized JSON files.
- **Preprocessing:** Temporal alignment, interpolation of missing keypoints, and multi-person consistency checks ensured robust input data.
- **Feature Engineering:** Derived motion and posture descriptors were computed, followed by PCA for dimensionality reduction (50 components, 95% variance retained).
- **Model Training:** Random Forest and XGBoost classifiers were optimized and evaluated on annotated video segments.

Results

Video Processing & Annotations

- Raw pose features were extracted using *psifx*.
- Robust preprocessing was applied, including temporal alignment and handling of missing keypoints and multi-person consistency.

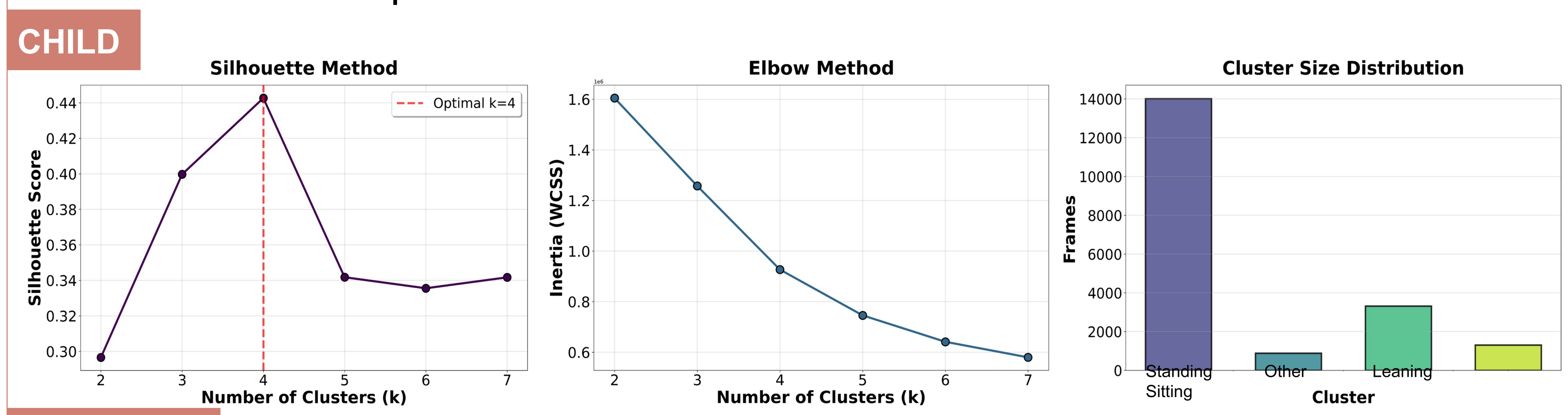




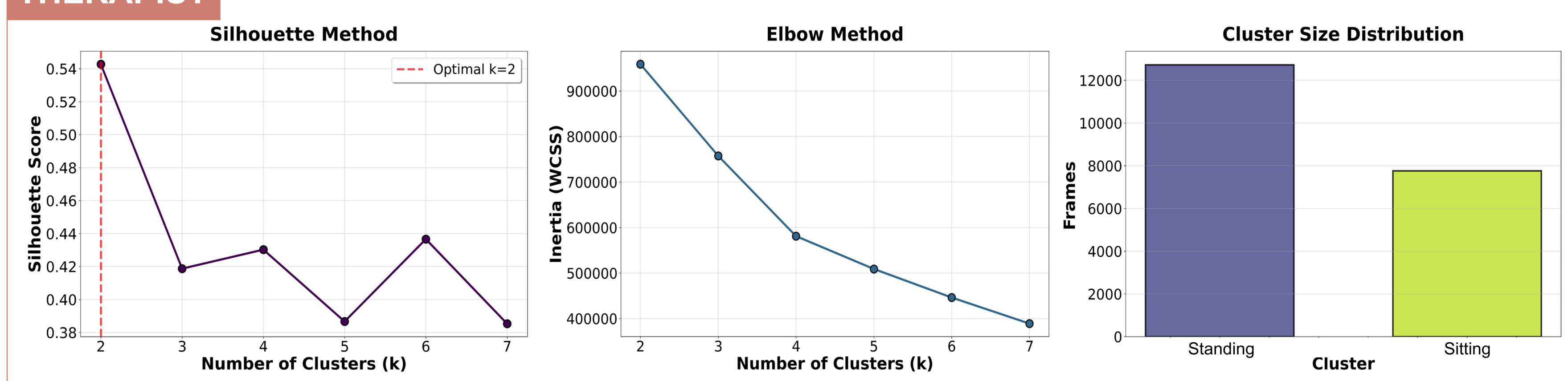
Posture Clusterization

- Normalized pose vectors were clustered using *K-means* to identify recurring body configurations (sitting, standing, leaning, and transitional postures).
- Cluster validity was assessed with the *silhouette method* (which measures how well-defined each cluster is) and the *elbow method* (which helps determine the optimal number of distinct clusters) to ensure meaningful posture groupings reflective of therapeutic behavior.

CHILD

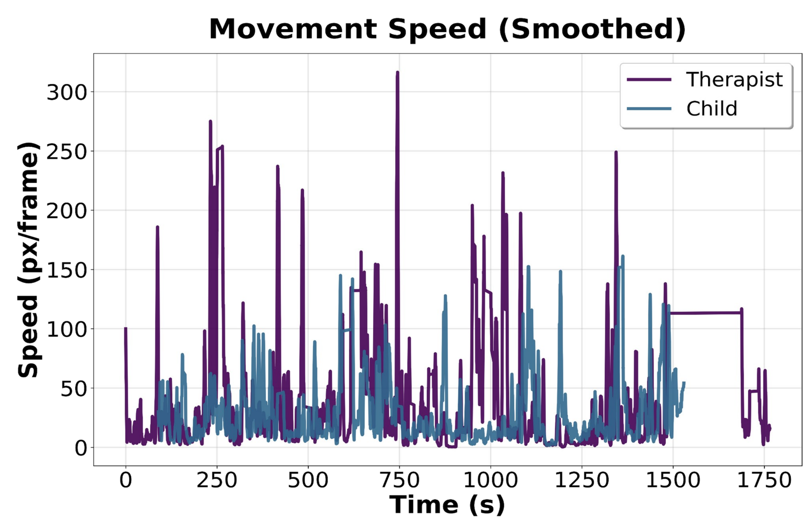


THERAPIST

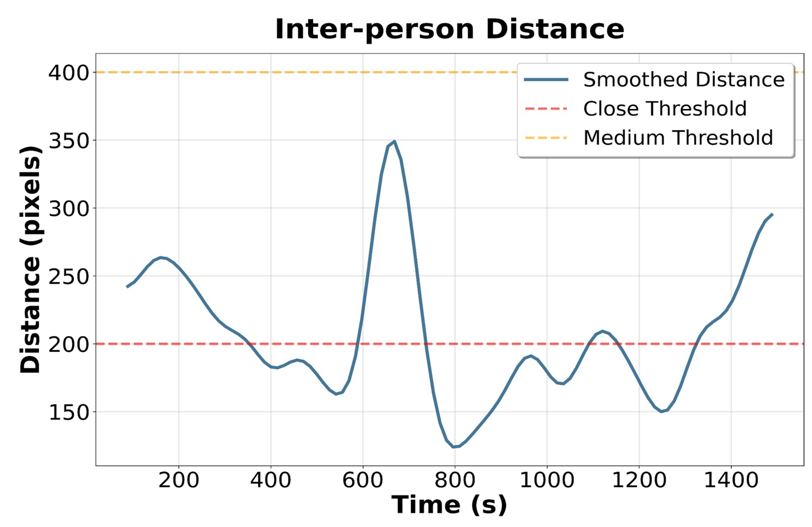


Optimal k was selected by maximizing the Silhouette Score, which provides a quantitative measure of cluster quality (cohesion and separation), overriding the visual Elbow point when scores conflict.

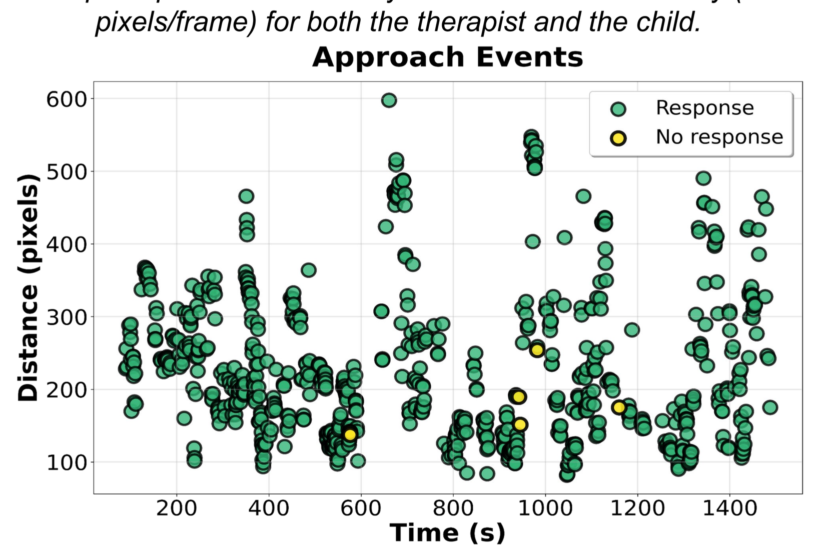
Feature Selection and Model Performance



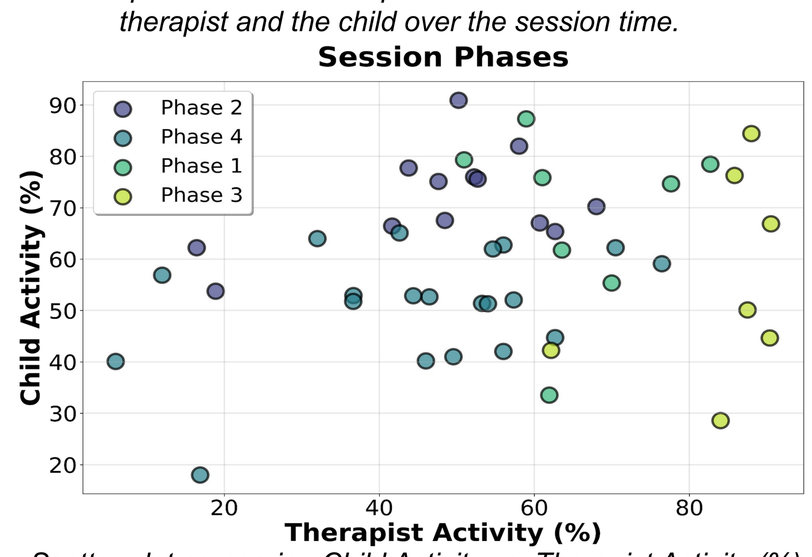
Temporal plot of the frame-by-frame movement velocity (in pixels/frame) for both the therapist and the child.



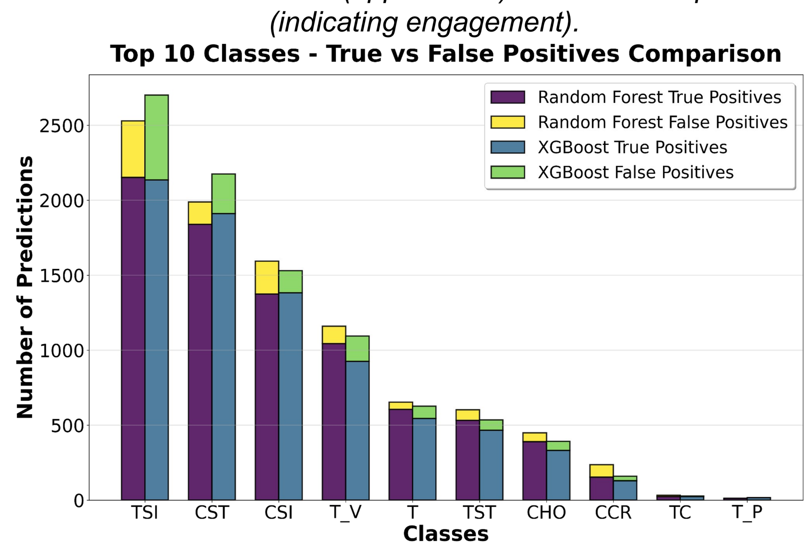
Graph of the smoothed pixel distance between the therapist and the child over the session time.



Scatter plot of approach distance vs. time, showing if nonverbal invitations to interact (approaches) received a response (indicating engagement).



Scatter plot comparing Child Activity vs. Therapist Activity (%) across 30-second time bins: Phase 1 - Mutual Engagement, Phase 2 - Quiet Observation, Phase 3 - Therapist Leading, Phase 4 - Child Exploring.



Metric	Random Forest	XGBoost
F1 Score	0.876	0.845
Precision	0.879	0.850
Recall	0.877	0.849

- The high F1 score (for Random Forest model 0.876) is crucial because the therapist and child have fundamentally different roles and thus distinct behavioral and postural profiles in a session.
- The ability to accurately distinguish their postures (the Therapist Sitting (TSI) from the Child Standing (CST)) confirms the system is performing the foundational task necessary for all subsequent analysis.
- This accurate classification is the first step toward correlating specific role-based nonverbal behaviors with therapeutic outcomes.

Discussion

- **Our findings** demonstrate that automated pose-based classification can provide a scalable, objective foundation for analyzing embodied therapeutic interactions. High model performance validates the effectiveness of our current pipeline for static pose recognition.
- **Ongoing work** focuses on improving temporal tracking and integrating supervised sequence models for behavior dynamics.
- **Future extensions** include multimodal integration (e.g., gaze, speech, facial expression) to enable richer behavioral mapping and deeper insights into therapist–child interaction patterns.[3]

References

1. Rochette, G., Rochat, M. & Vowels, M. (2024). psifx — Psychological and Social Interactions Feature Extraction Package. arXiv. https://doi.org/10.48550/arXiv.2407.10266
2. Lin, B.-X. et al. Video-Based Quantification of Child–Therapist Interaction Across Art Therapy Sessions. Champéry Retreat (2025).
3. Middelmann, N. K. et al. (2025) The ADVANCE Toolkit: Automated Descriptive Video Annotation in Naturalistic Child Environments. Behavior Research Methods, in press; OSF Preprint (2024). https://doi.org/10.31219/osf.io/y73mb