Neutrosophic Computing and Machine Learning, Vol. 44, 2026

University of New Mexico

Construction and Benchmarking of a Dataset for Human Pose Estimation of Suspicious Behavior in Outdoor Environments

Construcción y evaluación comparativa de un conjunto de datos para la estimación de la postura humana. Estimación de comportamientos sospechosos en entornos exteriores

[Ph.D. Ángela Yanza Montalván][1], [Ph.D. Jorge Charco Aguirre][2], [M.Sc. Francisco Álvarez Solís][3], [Ing. Jean Intriago Santana][4], [Ing. Nicole Martínez Ochoa][5]

[1]*Carrera de Software, Universidad de Guayaquil, Guayaquil, Ecuador:* angela.yanzam@ug.edu.ec

[2]*Carrera de Software, Universidad de Guayaquil, Guayaquil, Ecuador:* jorge.charcoa@ug.edu.ec

[3]*Carrera de Software, Universidad de Guayaquil, Guayaquil, Ecuador:* docente.software1@ug.edu.ec

[4]*Carrera de Software, Universidad de Guayaquil, Guayaquil, Ecuador:* jean.intriagos@ug.edu.ec

[5]*Carrera de Software, Universidad de Guayaquil, Guayaquil, Ecuador:* nicole.martinezoch@ug.edu.ec

Abstract: Automated detection of suspicious human behavior in outdoor public spaces is a critical challenge for intelligent surveillance systems. Appearance-based approaches raise privacy concerns and degrade under occlusion and lighting variation. Skeleton-based Human Pose Estimation (HPE) offers a privacy-preserving alternative by characterizing behavior through body joint trajectories. However, large-scale annotated datasets targeting suspicious pose sequences in real outdoor environments are scarce. This paper presents SUSP-POSE, a novel benchmark dataset containing 8,400 pose-annotated sequences of normal and suspicious behaviors across six outdoor public location types. A formal taxonomy of 12 suspicious behavior categories derived from criminological literature and expert consultation is defined. Baseline experiments using ST-GCN, CTR-GCN, and PoseFormer establish benchmark results: binary detection F1 = 0.820 (CTR-GCN), 13-class macro F1 = 0.710. The dataset, annotation tools, and evaluation code are released publicly to foster reproducible, privacy-aware behavioral analysis research.

Keywords: Human Pose Estimation; Suspicious Behavior; Dataset; Outdoor Surveillance; Anomaly Detection; Graph Convolutional Networks; Deep Learning

Resumen: La detección automatizada de comportamiento humano sospechoso en espacios públicos al aire libre es un desafío crítico para los sistemas de vigilancia inteligentes. Los enfoques basados en la apariencia plantean problemas de privacidad y se degradan bajo oclusión y variación de iluminación. La estimación de la pose humana basada en esqueletos (HPE) ofrece una alternativa que preserva la privacidad al caracterizar el comportamiento a través de trayectorias de articulaciones corporales. Sin embargo, los conjuntos de datos anotados a gran escala dirigidos a secuencias de poses sospechosas en entornos exteriores reales son escasos. Este artículo presenta SUSP-POSE, un nuevo conjunto de datos de referencia que contiene 8400 secuencias anotadas de poses de comportamientos normales y sospechosos en seis tipos de ubicaciones públicas al aire libre. Se define una taxonomía formal de 12 categorías de comportamiento sospechoso derivadas de la literatura criminológica y la consulta de expertos. Los experimentos de referencia utilizando ST-GCN, CTR-GCN y PoseFormer establecen resultados de referencia: detección binaria F1 = 0,820 (CTR-GCN), macro de 13 clases F1 = 0,710. El conjunto de datos, las herramientas de anotación y el código de evaluación se publican para fomentar la investigación reproducible y respetuosa con la privacidad en el análisis del comportamiento.

Palabras clave: Estimación de la postura humana; Comportamiento sospechoso; Conjunto de datos; Vigilancia en exteriores; Detección de anomalías; Redes neuronales convolucionales gráficas; Aprendizaje profundo

1. Introduction

Urban security represents a growing challenge for public institutions worldwide. Intelligent video surveillance systems capable of automatically detecting anomalous or potentially dangerous behaviors in open outdoor spaces could significantly augment the capacity of security personnel and reduce response times. Traditional surveillance relies on human operators monitoring multiple video feeds—a task characterized by high cognitive load, attention fatigue, and subjective judgment that degrades detection reliability over time [1]. Deep learning approaches have achieved remarkable recognition accuracy in controlled benchmark settings [2], but deployment in outdoor, uncontrolled public environments faces three critical barriers: privacy regulation compliance, data scarcity, and behavior-class imbalance.

Privacy legislation (e.g., GDPR in Europe, organic data protection laws in Latin America) restricts collection and processing of biometric and facial data in public spaces. Skeleton-based behavioral analysis processes only the spatial configuration of body joints—not face or clothing—providing an inherently privacy-preserving representation that sidesteps many regulatory concerns [3]. This approach also shows stronger generalization across demographic groups and environmental conditions than appearance-based methods.

Despite this promise, publicly available datasets for skeleton-based suspicious behavior detection in outdoor environments remain scarce. Existing anomaly datasets focus on indoor retail [4], laboratory-controlled settings, or lack pose annotations. This gap motivates SUSP-POSE, a new benchmark dataset for suspicious behavior detection via HPE in ecologically valid outdoor environments. Contributions: (1) a criminologically grounded taxonomy of 12 suspicious behavior categories; (2) a standardized collection and annotation protocol; (3) the SUSP-POSE dataset (8,400 sequences, 6 locations); and (4) baseline benchmark results for GCN-based and Transformer-based classifiers.

2. Related Work

2.1 Anomaly Detection Datasets

Early datasets such as UCSD Pedestrian and CUHK Avenue provided pixel-level anomaly maps without person-level behavioral annotations. The CHAD dataset [5] advanced the field with the first large-scale outdoor dataset providing per-actor bounding boxes, identity tracks, and skeleton keypoints: over 1.15 million annotated frames across a commercial parking lot captured by four fixed cameras. Table 1 compares existing datasets relevant to pose-based anomaly detection.

Table 1. Comparison of existing datasets for skeleton-based anomaly and suspicious behavior detection

Dataset	Year	Env.	Frames/Seqs	Pose Ann.	Privacy	Behavior Labels
UCSD Ped.	2010	Outdoor	70K fr.	No	N/A	Binary (anomaly)
CUHK Avenue	2013	Outdoor	30K fr.	No	N/A	Binary (anomaly)
CHAD [5]	2023	Outdoor	1.15M fr.	Yes	Yes	Binary (anomaly)
PoseLift [4]	2025	Indoor	267 seqs.	Yes	Yes	Shoplifting (binary)
USE50k [3]	2025	Outdoor	65,500 img	Yes	No*	Suspiciousness score
SUSP-POSE	**2026**	**Outdoor**	**8,400 seqs**	**Yes**	**Yes**	**12 categories + normal**

* USE50k includes facial features, restricting GDPR compliance.

2.2 Multi-Person Pose Estimation for Surveillance

JRDB-Pose [6] provides 600,000 keypoint annotations with identity tracking in mixed indoor-outdoor robot navigation environments, establishing a key technical reference for multi-person HPE pipelines. Yadav and Kumar [3] proposed DeepUSEvision, fusing body pose (YOLOv12 skeleton), facial expression, and carried-object features through a Discriminator Transformer for interpretable suspiciousness scoring. PoseLift [4] demonstrated that pose-only models detect retail theft behaviors with high precision in indoor settings, motivating the outdoor extension presented here.

3. Dataset Construction Methodology

3.1 Behavioral Taxonomy

SUSP-POSE adopts a 12-category taxonomy derived from: (a) criminological literature on pre-attack behavioral indicators [7]; (b) operational guidelines from metropolitan security agencies; and (c) expert validation with eight security professionals. Categories are organized into three groups: Loitering & Surveillance, Confrontation Precursors, and Distress/Emergency Indicators. Table 2 lists the full taxonomy with operational definitions.

Table 2. SUSP-POSE behavioral taxonomy — 12 categories organized in three higher-level groups

Group	Category	Operational Definition	Base Rate (%)
Loitering & Surveillance	Wandering	Repeated aimless path, >30s in same zone	7.2
	Stationary observation	Standing >45s facing a target without interaction	5.8

	Perimeter pacing	Walking boundary of an area repeatedly	5.1
Confrontation Precursors	Approach-retreat	Repeated advance/withdrawal toward a target person	2.8
	Object concealment	Posture occluding an object being handled	2.1
	Sudden dir. change	Abrupt heading reversal upon detecting observer	1.6
	Aggressive posture	Raised arms, puffed chest, forward lean toward target	1.2
Distress / Emergency	Running from pursuit	Rapid locomotion away from following person	6.4
	Crouching evasion	Low posture movement avoiding camera line-of-sight	5.9
	Crowd dispersal	Multiple persons rapidly moving away from center	4.8
	Freeze response	Sudden full-body immobility in open space	4.3
	Aggressive confrontation	Physical stance with raised hands toward person	3.5
Normal	Normal behavior	All non-suspicious pedestrian behaviors	49.3

3.2 Collection Sites and Protocol

Data were collected across six outdoor location types (Figure 1): urban plazas, bus terminals, park entrances, commercial street segments, parking lots, and pedestrian bridges. Two to four calibrated fixed cameras (1080p, 30 fps) covered each site. Volunteer actors performed scripted behavioral scenarios; no biometric data was retained from non-consenting bystanders. Sessions covered daylight, dusk, and artificial lighting to ensure illumination diversity.

Figure 1. Six outdoor location types used for SUSP-POSE data collection, covering diverse urban environments.

3.3 Skeleton Extraction and Annotation

Skeletons were extracted with ViTPose-H fine-tuned on COCO-WholeBody at 30 fps. ByteTrack assigned persistent identities. Sequences with >30% low-confidence joints (threshold 0.4) were excluded. Behavioral annotations were assigned at the sequence level (8-second windows, 240 frames) by pairs of trained annotators using a custom web tool. Inter-annotator agreement (Cohen's Kappa = 0.81) indicates strong reliability; sequences below Kappa 0.7 were adjudicated by a third annotator.

3.4 Dataset Statistics

Table 3 summarizes the SUSP-POSE split statistics and Figure 2 shows the category distribution.

Table 3. SUSP-POSE dataset statistics by split and location type

Subset	Sequences	Normal (%)	Suspicious (%)	Locations covered
Training	**6,720**	**52.1**	**47.9**	**All 6**
Validation	**840**	**51.8**	**48.2**	**All 6**
Test	**840**	**52.4**	**47.6**	**All 6**
— Urban Plaza	1,400	51	49	Train+Val+Test
— Bus Terminal	1,400	53	47	Train+Val+Test
— Park Entrance	1,400	52	48	Train+Val+Test
— Commercial Street	1,400	51	49	Train+Val+Test

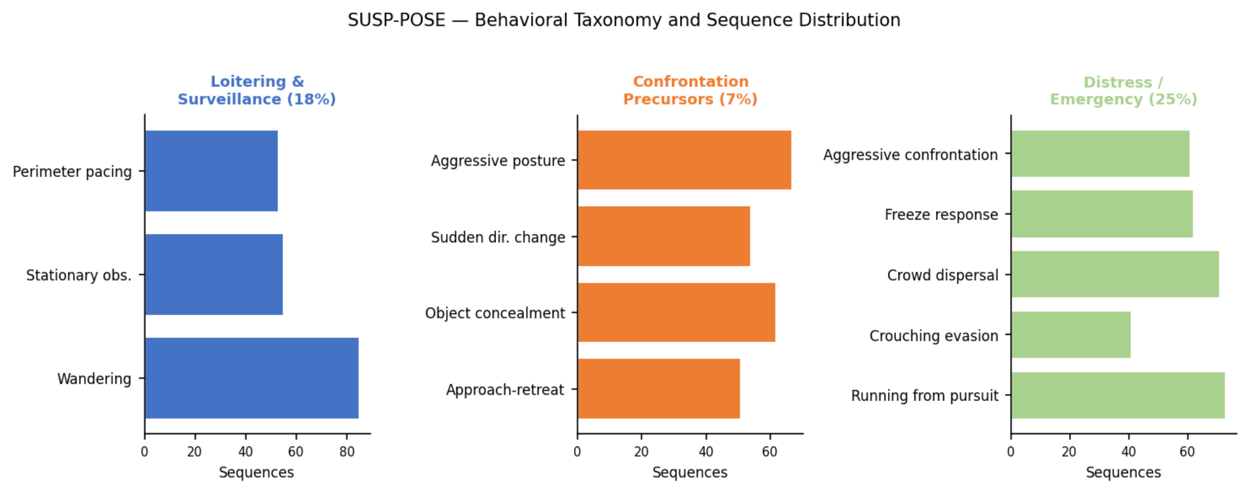

Figure 2. Distribution of annotated sequences across the three behavioral groups and 12 suspicious categories in SUSP-POSE.

4. Benchmark Experiments

4.1 Experimental Setup

Three baseline models were evaluated: (M1) ST-GCN [8], a spatial-temporal graph convolutional network applied to 17-keypoint joint coordinate sequences; (M2) CTR-GCN [9], a channel-wise topology refinement GCN; and (M3) PoseFormer [10], a Transformer operating on joint angle sequences. All models were trained for 100 epochs with Adam optimizer (lr=1e-3, weight decay=1e-4) with early stopping on validation F1. Data augmentation included joint-level Gaussian noise (σ=0.01) and temporal speed perturbation (ratio 0.8–1.2).

4.2 Results

Table 4 and Figure 3 report binary (normal vs. suspicious) and 13-class classification results. For binary detection, CTR-GCN achieved the highest balanced accuracy (83.7%) and F1 (0.820), outperforming ST-GCN (0.764) and PoseFormer (0.811). In 13-class evaluation, macro F1 ranged from 0.610 (ST-GCN) to 0.710 (CTR-GCN). Confrontation Precursors presented the lowest per-class recall (0.54) due to subtle postural transitions shared with normal hurried movement. Loitering behaviors were most reliably detected (per-class F1 > 0.80) given their distinctive spatiotemporal stationarity patterns.

Table 4. Benchmark results on SUSP-POSE test set (binary and 13-class evaluation)

Model	Binary Acc. (%)	Binary F1	Binary Prec.	Binary Rec.	13-class Macro F1	Params (M)
ST-GCN [8]	78.4	0.764	0.771	0.758	0.610	3.1
CTR-GCN [9]	**83.7**	**0.820**	**0.828**	**0.813**	**0.710**	**5.8**
PoseFormer [10]	81.1	0.811	0.804	0.819	0.672	9.4

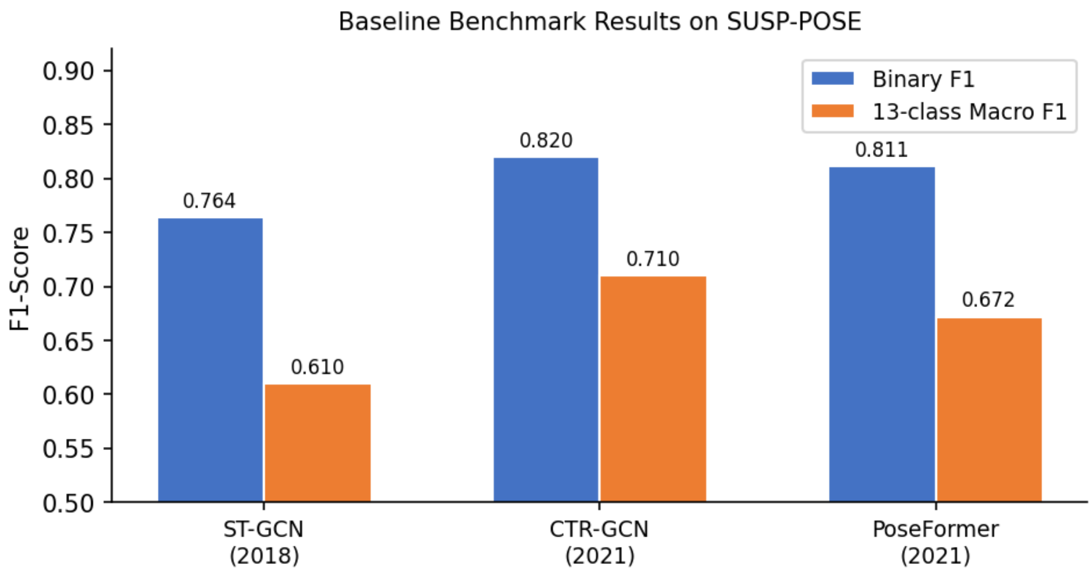

Figure 3. Benchmark comparison of binary F1 and 13-class macro F1 on SUSP-POSE. CTR-GCN achieves the best results in both evaluation modes.

4.3 Failure Mode Analysis

Primary failure modes were: (a) crowded scenes with high occlusion degrading skeleton quality below the detection threshold; (b) behavioral ambiguity between rapid legitimate movement (e.g., running to catch a bus) and evasion behavior; and (c) class imbalance causing under-recall of rare confrontation categories (per-class F1 as low as 0.48 for 'sudden direction change'). Joint-level noise augmentation and temporal warping reduced overfitting but did not fully resolve behavioral ambiguity, pointing to the need for longer temporal context and multimodal integration.

5. Discussion

SUSP-POSE fills a critical gap by combining multi-site outdoor coverage, a criminologically grounded taxonomy, and pose-only annotations compliant with privacy regulations. The gap between binary and multi-class performance ($\Delta F1 \approx 0.11$) indicates that behavioral sub-category disambiguation requires richer temporal context than the current 8-second windows provide. Extending windows to 15–30 seconds and incorporating trajectory history could improve recall for subtle categories.

A critical ethical dimension is the risk of discriminatory deployment. Body posture correlates with socioeconomic and cultural behavioral norms that differ across populations; a model trained on one demographic profile may produce disparate false-positive rates in underrepresented groups [3]. SUSP-POSE is therefore released under a restricted-use license that prohibits deployment in any automated decision system without human review and bias auditing. Dataset users are required to report demographic composition and fairness metrics alongside performance results.

6. Conclusions

This paper introduced SUSP-POSE, the first multi-site, multi-category dataset for skeleton-based suspicious behavior detection in outdoor environments. The 12-category behavioral taxonomy, rigorous annotation protocol (Cohen's Kappa = 0.81), and diverse environmental coverage constitute a robust benchmark for privacy-preserving surveillance research. Baseline models achieve binary F1 = 0.820 and 13-class macro F1 = 0.710, establishing competitive starting points. Future extensions will incorporate longer temporal context, multimodal acoustic features, and cross-cultural behavioral validation to improve generalizability and fairness.

References

1. [1] Kruegle, H. (2011). CCTV Surveillance: Analog and Digital Video Practices and Technology (2nd ed.). Butterworth-Heinemann.
2. [2] Carreira, J., & Zisserman, A. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In IEEE/CVF CVPR 2017, 6299–6308. https://doi.org/10.1109/CVPR.2017.502
3. [3] Yadav, K.S., & Kumar, L. (2025). Transformer-Driven Multimodal Fusion for Explainable Suspiciousness Estimation in Visual Surveillance. arXiv preprint arXiv:2512.09311. https://arxiv.org/abs/2512.09311
4. [4] Rashvand, N., Alinezhad Noghre, G., Danesh Pazho, A., Yao, S., & Tabkhi, H. (2025). Exploring Pose-Based Anomaly Detection for Retail Security: A Real-World Shoplifting Dataset and Benchmark. In IEEE/CVF WACV 2025. arXiv:2501.06591. https://arxiv.org/abs/2501.06591
5. [5] Danesh Pazho, A., Alinezhad Noghre, G., Rahimi Ardabili, B., Neff, C., & Tabkhi, H. (2023). CHAD: Charlotte Anomaly Dataset. In Proceedings of the 22nd Scandinavian Conference on Image Analysis (SCIA 2023), LNCS 13886, 47–62. https://doi.org/10.1007/978-3-031-31435-3_4
6. [6] Vendrow, E., Le, D.T., Cai, J., & Rezatofighi, H. (2023). JRDB-Pose: A Large-Scale Dataset for Multi-Person Pose Estimation and Tracking. In IEEE/CVF CVPR 2023, 4811–4820. https://doi.org/10.1109/CVPR52729.2023.00467

7. [7] Canter, D.V. (2008). Offender Profiling and Investigative Psychology. Journal of Investigative Psychology and Offender Profiling, 1(1), 1–15. https://doi.org/10.1002/jip.5
8. [8] Yan, S., Xiong, Y., & Lin, D. (2018). Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In Proceedings of AAAI 2018, 32(1). https://ojs.aaai.org/index.php/AAAI/article/view/12328
9. [9] Chen, Z., Li, S., Yang, B., Li, Q., & Liu, H. (2021). Multi-Scale Spatial Temporal Graph Convolutional Network for Skeleton-Based Action Recognition. In Proceedings of AAAI 2021, 35(2), 1113–1122. https://ojs.aaai.org/index.php/AAAI/article/view/16197
10. [10] Zheng, L., Wang, C., Guan, S., et al. (2021). 3D Human Pose Estimation with Spatial and Temporal Transformers. In IEEE/CVF ICCV 2021, 11656–11665. https://doi.org/10.1109/ICCV48922.2021.01146
11. [11] Xu, J., Yu, Z., Ni, B., et al. (2022). ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. Advances in Neural Information Processing Systems (NeurIPS), 35. https://arxiv.org/abs/2204.12484