Video-Based Identification of Sequential Body Impact Location Using Pose Estimation.

Ethan Henley, Brianna Reilly, Katherine Mayes, Damon Kuehl, Nicole E.-P. Stark, Steve Rowson

I. INTRODUCTION

Falls are a leading cause of injury in older adults, resulting in millions of injuries each year, ranging from wrist and hip fractures to traumatic brain injury [1]. The location of impact during a fall is a key determinant of injury type and severity, influencing biomechanical loading pathways and clinical outcomes [2]. Direct visual capture of falls provides a unique opportunity to better understand these mechanisms and to inform fall-prevention strategies and injury thresholds. Current approaches to analysing fall videos often rely on manual annotation or simplified assumptions, which limits scalability and introduces subjectivity in identifying impact locations [3]. As a result, there is a critical need for automated and objective methods to determine body impact locations from real-world fall footage. Video-based pose estimation offers a scalable alternative, by extracting body kinematics from video data. While prior work has leveraged pose estimation for fall detection and activity recognition [7], foundational fall-biomechanics studies have used pose estimates to characterise body configuration and to validate visually observed first contacts [3,5]. These studies relied on manual coding and lacked the automation and predictive power of machine learning. To our knowledge, no prior study has used machine learning to predict which body region first contacts the ground from pose-derived kinematics. This study advances from manual coding to automated, ML-driven prediction, supporting reconstruction of injury mechanisms and inputs for injury biomechanics models.

II. METHODS

We analysed 100 fall videos from the Le2i dataset [6] (320×240 pixels, 24 fps), with 455 impact events manually annotated across six body regions (head, shoulder, elbow, hand, hip, knee) and the first ground contact identified per region. Pose estimation used YOLO11-Large [9], extracting COCO 17-keypoint coordinates and confidence scores per frame; individuals were tracked using Hungarian matching on confidence-weighted centroids, and a reference body height (90^{th} percentile shoulder-to-ankle distance) was computed per video for normalisation.

To identify impact locations, we formulated the task as a multi-class classification problem, predicting the body region contacting the ground at a fixed time horizon (~208 ms) prior to impact. An initial image-space approach performed poorly (26% accuracy), highlighting two challenges: (1) sequential impacts place multiple regions in contact at later stages of a fall, and (2) camera perspective distorts vertical positioning, limiting pixel-based height features. We therefore developed a unified single classifier for all impact events: first contacts rely on pose-derived features, while subsequent contacts add prior-contact indicators of which regions have already struck the ground. This increases training data (421 events versus ~90 for first contacts alone) and enables knowledge transfer: the model learns body-part-specific pose signatures from abundant subsequent contacts and applies them to first contacts. Three feature groups were extracted: camera-invariant kinematics (joint angular velocities, segment speeds, body configuration), keypoint confidence dynamics capturing occlusion as body parts approach the ground, and temporal contact context. A gradient boosting classifier (HistGradientBoosting) was trained and evaluated using stratified five-fold cross-validation (CV). Performance was summarised using precision, recall, and the F1-score: for a given body region, recall is the proportion of actual impacts there that the model correctly identified, precision is the proportion of impacts it assigned to that region that were correct, and the F1-score combines the two into a single balanced average; the macro F1-score averages F1 equally across the six regions. Differences in kinematic features between fall types are reported as Cohen's d, where 0.2, 0.5, and 0.8 denote small, moderate, and large effects.

E. Henley is a Ph.D. student in Biomedical Engineering at Virginia Tech (ethan02@vt.edu), B. Reilly is a BS student in Neuroscience at Virginia Tech, K. Mayes and D. Kuehl are emergency physicians at Carilion Clinic, S. Rowson is a Professor in Biomedical Engineering, and N. E.-P. Stark is a Research Assistant Professor in Biomedical Engineering at Virginia Tech.

III. INITIAL FINDINGS

Each fall comprised one to six sequential impacts (median inter-impact gap two frames, ~83 ms). After excluding 12 videos with unreliable pose estimation, the final dataset comprised 421 impact events from 88 videos. Across all impact events, the proposed model achieved an accuracy of 58.7% ± 1.5% and a macro F1-score of 0.57 ± 0.01, outperforming chance (17%) and the majority-class baseline (26%). Performance was stable across 20 random CV splits (range: 55.1%–61.3%). Within the unified model, first-contact events were classified at 57.9% ± 3.8% accuracy, while subsequent contacts achieved 58.9% ± 1.4% (TABLE I). For subsequent contacts, prior-contact features were highly influential, with sequence timing and contact history comprising eight of the top ten predictive features.

TABLE I: PER-CLASS PERFORMANCE (CV FOLD)

Region	n	Recall	F1
Head	44	55%	0.52
Shoulder	50	68%	0.64
Elbow	56	34%	0.34
Hand	110	75%	0.72
Hip	89	63%	0.60
Knee	72	49%	0.47

Analysis of impact sequences revealed that hand (48%) and knee (43%) impacts accounted for 91% of first contact events, consistent with protective reflexes during forward falls [2]. In contrast, head, shoulder, and elbow impacts occurred almost exclusively as subsequent contacts.

The most discriminative features were trunk angular velocity (0.117), centre-of-mass velocity (0.099), and knee speed (0.096). Fall types showed distinct patterns: hand-first falls by joint extension (knee +0.30, hip +0.13 rad/s) and knee-first falls by flexion (knee −0.44, hip −0.42 rad/s), reflecting protective arm bracing versus lower-limb collapse.

IV. DISCUSSION

Video-based pose estimation can identify body impact locations during falls, enabling scalable extraction of biomechanical data from real-world footage. Because impact location governs load transmission and injury patterns [2–3], automated identification is an important step toward quantifying exposure conditions, and generalises to domains such as sports and transportation, where impact location informs injury risk modelling and reconstruction [11–12].

This study makes four key contributions. First, falls exhibit predictable sequential contact patterns, with hand and knee typically contacting first, followed by hip, then shoulder and head, supporting the need to model falls as multi-impact sequences. Leveraging all events (n = 421) rather than isolating first contacts improved first-contact accuracy from 45% to 58%, demonstrating knowledge transfer from later impacts. Feature analysis confirmed the model relies on kinematics absent contact history and switches to sequence-based features once contacts are known. Second, camera perspective limits traditional 2D image features, whereas camera-invariant kinematics improve performance. Third, joint angular velocities encode biomechanical signatures, distinguishing protective bracing (extension) from collapse (flexion) [4]. Fourth, keypoint confidence dynamics provide a novel signal reflecting occlusion patterns.

Clinically, the observed distal-to-proximal sequence (hand/knee → hip → shoulder/head) provides 80–170 ms of lead time before head impact, which may enable active protection strategies [8]. These findings support future integration with biomechanical models to estimate impact forces and injury risk from predicted contact sequences.

Limitations include video resolution and dataset size. Low resolution reduces keypoint accuracy (head confidence ~ 0.65 at impact), and 14% of videos were unusable. Signal-to-noise constraints also limited feature separability (trunk angular velocity Cohen's d = 0.36, a small-to-moderate effect). Larger datasets, higher-resolution video, and 3D pose estimation [10] are likely to improve performance.

V. REFERENCES

[1] World Health Organisation. Falls. Fact Sheet, 2021. [2] Nevitt, M.C. et *al.,* J Am Geriatr Soc 1993. [3] Yang, Y. *et al.,* J Bone Miner Res. 2020. [4] Hsiao, E.T. *et al.,* J. Biomech. 1998. [5] Robinovitch, S.N. *et al.,* Lancet 2013. [6] Charfi, I. *et al*., J. Electron. Imaging 2013. [7] Ramirez, H. *et al.,* IEEE Access 2021. [8] Tamura, T. *et al.,* IEEE Trans. Inf. Tech. Biomed., 2009. [9] Jocher, G. *et al.,* Ultralytics YOLO11 2024. [10] Goel, S. *et al*., ICCV 2023. [11] Cortes, N. *et al.,* Am. J. Sports Med. 2017. [12] Han, Y. *et al*., Int. J. Crashworthiness 2019.