Keystroke-Aligned Body Motion Patterns for Short-Burst Continuous Smartphone Authentication: A Proof-of-Concept Study Using Motion Capture

Nicholas Cariello, Lam Nguyen, Rosemary Gallagher, Isaac Kurtzer, Kiran S. Balagani, and Paolo Gasti

***Abstract*—Continuous smartphone user authentication systems aim to verify a user's identity by monitoring behavioral patterns during typical device usage. Current approaches based on keystroke dynamics have demonstrated robust performance over extended authentication windows ranging from 20 to 120 seconds, with recent advances reducing these requirements to 5–10 seconds. However, many common smartphone interactions, such as typing a URL in a mobile browser, entering search queries on the home screen, or composing brief text message responses, occur in very short bursts lasting only 1–2 seconds. Consequently, the current state of the art in smartphone keystroke dynamics remains unsuitable for authenticating users during these prevalent short-burst textual interactions.**

Building upon our previous research that demonstrated the effectiveness of integrating body motion features with swipe-based behavioral biometrics, this proof of concept study explores whether similar performance enhancements can be achieved for keystroke dynamics within short authentication windows. We investigate a multimodal approach that combines traditional keystroke dynamics with laboratory-grade 3D motion capture body movement data and smartphone motion features, utilizing keystroke events as temporal anchors for feature extraction.

We evaluate our methodology on our publicly available dataset comprising of 42 users and demonstrate that the integration of phone motion and body motion features with keystroke dynamics achieves an equal error rate (EER) of 1.5% over 1-second authentication windows when utilizing all available features. This represents a substantial improvement over keystroke-only approaches, which typically achieve EERs of 9.5% to 11.9% under similar conditions. Additionally, we identify the specific body regions and feature types most valuable for short-window authentication, providing a roadmap for future implementations using smartphone sensors and wearable devices.

***Index Terms*—Behavioral biometrics, continuous authentication, keystroke dynamics, motion capture, multimodal authentication, smartphone security**

I. Introduction

CONTINUOUS authentication systems represent a paradigm shift from traditional point-of-entry authentication methods, aiming to provide ongoing verification of user identity through the seamless monitoring of behavioral patterns during natural smartphone interactions. This approach addresses fundamental limitations of conventional authentication mechanisms, which create security vulnerabilities during the extended periods between initial login and subsequent usage sessions. Behavioral biometrics, particularly keystroke dynamics, have been promising as a foundation for continuous authentication due to their ability to capture user-specific interaction patterns without requiring explicit authentication actions from the user [1], [2].

Despite their potential, keystroke dynamics approaches face inherent limitations [3], [4] that significantly impact their practical deployment in smartphone environments. To achieve acceptably low authentication error rates, these systems typically require substantial collections of user interaction samples [5], [6], necessitating extended observation periods that can span 20 to 120 seconds of continuous typing activity [5], [6]. This requirement creates a fundamental tension between security and usability: longer observation periods provide more robust authentication but allow potential adversaries to remain undetected through larger numbers of interactions, while simultaneously creating user experience friction through delayed authentication decisions.

Recent research efforts have focused on addressing these limitations by developing techniques that can reduce the required authentication windows to more practical durations of 5 to 10 seconds [7]–[9]. While these advances represent meaningful progress toward practical deployment, they still fail to address a critical gap in smartphone authentication scenarios. Smartphone users frequently engage in very brief typing interactions that complete within 1 to 2 seconds, including activities such as entering URLs, typing search queries, or composing brief text messages [10].

Recent research efforts have focused on addressing these limitations by developing techniques that can reduce the required authentication windows to more practical durations of 5 to 10 seconds [7]–[9]. While these advances represent meaningful progress toward practical deployment, they still fail to address a critical gap in smartphone authentication scenarios. Smartphone users frequently engage in very brief typing interactions that complete within 1 to 2 seconds, including activities such as entering URLs, typing search queries, or composing brief text messages [10]. This prevalence of short-burst interactions motivates our investigation into whether body motion features, temporally aligned with keystroke events, can enable reliable authentication within these very short windows.

A. Research Objectives

This proof-of-concept study evaluates whether body motion features can enhance keystroke-based authentication within

N. Cariello, L. Nguyen, R. Gallagher, I. Kurtzer, K. S. Balagani, and P. Gasti are with the New York Institute of Technology, Old Westbury, NY 11568 USA (e-mail: ncariell@nyit.edu; lnguye25@nyit.edu; rgalla01@nyit.edu; ikurtzer@nyit.edu; kbalagan@nyit.edu; pgasti@nyit.edu).

very short (1–2 second) windows. Using high-precision motion capture equipment, we aim to identify which body motion features and regions provide discriminative value, establishing a foundation before pursuing implementation with commodity sensors.

Our investigation focuses on extracting motion signals anchored to typing events through what we term *keystroke alignment*: we use each keystroke event as a temporal anchor around which we define a *keystroke event perimeter*. We then extract features within this perimeter from synchronized phone and body motion sensor data. This event-driven approach, illustrated in Figure 1, focuses computational resources on moments when authentication-relevant motion patterns are most pronounced.

We explore various approaches to temporal alignment between keystroke events and feature computation, investigating different window sizes and alignment strategies to optimize feature extraction. We analyze the discriminative value of different body regions, examining whether the body motion features that proved effective for swipe authentication maintain their utility for typing interactions.

As a proof-of-concept study, our body movement features are extracted from 3D motion capture data to establish a precise reference measurement regarding feature discriminability. We do not expect that, in a real-world setting, 3D motion capture equipment will be used to collect body posture/movement features for the purpose of continuous authentication on smartphones. Rather, we expect that the most promising features for keystroke events, identified in this work, will then be collected using smartphone sensors and commercially-available wearable devices (see Section IV), ultimately leading to comparable low authentication latency and low error rates.

B. Summary of Contributions

Our previous research demonstrated that the integration of body motion features with behavioral biometrics successfully addresses the challenge of short authentication windows, achieving reliable user authentication within 1–2 second timeframes for swipe-based interactions [11]. However, prior research focused exclusively on swiping, and is therefore not applicable to typing interactions.

Swipe and keystroke events have inherently different temporal structures. Swipe gestures are *continuous, sequential* motions with defined start and end points, where features can be extracted over the gesture's natural duration. In contrast, keystroke events are *discrete, bursty* occurrences that create a "ripple effect" in motion patterns, where each tap propagates motion as the user continues to type.

This burstiness introduces unique challenges not present in swipe authentication: (1) there are no natural boundaries defining when the authentication-relevant motion begins and ends; (2) consecutive keystrokes are often so close that the ripple effect from the first keystroke overlaps with the ripple effect from the next keystroke; and (3) the temporal characteristics of these ripples vary based on typing speed and style. To capture this complexity, we introduce the concept of *keystroke event perimeters* as temporal windows centered around individual keystroke events. This is a completely different feature extraction paradigm than our previous swipe-based work, where feature windows were defined by gesture boundaries rather than configurable perimeters.

As such, this paper includes the following key contributions:

- **Proof-of-concept demonstration**: We establish, through controlled motion capture experiments, that body motion features hold substantial discriminative value for keystroke authentication in very short (1–2 second) authentication windows, providing a foundation for future real-world implementations.
- **Keystroke event perimeters**: We introduce the concept of keystroke event perimeters as configurable temporal windows that capture the ripple effect of motion around discrete keystroke events. This differs from swipe-based approaches where gesture boundaries define feature windows, and addresses the unique challenge of extracting features from short-burst interactions.
- **Short-window performance**: With EERs as low as 1.5%, our technique is the first to achieve reliable authentication during common short-burst textual interactions (URL entry, search queries, brief messages) on smartphones. This represents a 77% improvement over our previous swipe-based results and substantially outperforms all existing keystroke and smartphone IMU approaches at comparable windows.
- **Body region analysis**: We identify the most impactful body motion features and regions for keystroke authentication, ultimately demonstrating that center-body features (arms, shoulders, elbows) provide the strongest discriminative signals.
- **Public dataset**: We republished our refined and standardized dataset containing smartphone IMU data, 3D body motion capture data, and smartphone events from 82 users, totaling 61,047 keystrokes. The dataset is publicly available at https://www.nyit-lamp.com/dataset/dataset4/.

C. Organization

The remainder of this paper is organized as follows. We review related work in Section II. In Section III, we describe our methodology, including the dataset used, the feature extraction process, and the classifiers employed. We then address the feasibility of leveraging body motion features in Section IV. Section V presents our experimental results, including performance across different postures, authentication windows, and n-graph configurations. Finally, we discuss varied related topics in Section VI, and present our conclusions and outline future work in Section VII.

II. Related Work

Over the past decade, continuous authentication has evolved from early keystroke-only schemes to sophisticated multimodal systems leveraging time-series deep learning. Table I provides a quantitative comparison of key approaches, revealing the progression from long-window methods to recent short-window attempts. We structure our review around three research streams: keystroke dynamics for mobile authentication, time-series behavioral biometrics with multimodal fusion, and human activity recognition methods that inform our feature extraction approach.

Keystroke Dynamics for Mobile Authentication

The adaptation of keystroke dynamics from desktop to mobile environments has progressed through distinct phases. For instance, Shen et al. [8] made a notable advancement this transition with an active authentication framework achieving EERs between 1.7% and 9% using 11-second windows, establishing that smartphone touch and keystroke behaviors contain sufficient discriminative information but require long observation periods. Li et al. [7] addressed data scarcity through augmentation techniques, reducing windows to 5 seconds at 4.7% EER.

Recognition that most mobile typing occurs in very short bursts prompted re-evaluation of these approaches. Senarath et al. [13] revealed that authentication performance degrades sharply below 5-second windows, even with anomaly detectors tuned for small samples. Dutta et al. [2] demonstrated through a 160-participant evaluation that classical Scaled Manhattan Distance remained competitive below 5 seconds (2.48–3.60% EER), suggesting that sophisticated architectures may not outperform well-tuned traditional methods when data is limited. This was a contributing factor in our classifier selection, where we aimed to balance performance with computation and energy overhead. Roy et al. [12] showed that combining low-sampling-rate keystroke features with soft biometrics could achieve <3% EER at 2.5-second windows. However, when authentication windows drop below 2.5 seconds, these approaches degrade in performance. This leaves common smartphone interactions lasting only 1–2 seconds—such as URL entry, search queries, and brief messages—without authentication coverage comparable to what existing methods achieve at longer windows.

Time-Series Modeling and Multimodal Fusion

As limitations became apparent, research pursued advanced time-series modeling and sensor fusion strategies. Nguyen et al. [17] demonstrated that spatio-temporal dual-attention transformers could extract discriminative representations from keystroke and IMU sequences by learning which sensor channels and time steps contribute most to authentication. Senarath et al.'s [16] BehaveFormer extended this through stacked attention modules and cross-modal fusion, exploiting correlations between typing rhythm and device micro-movements. These transformer-baed architectures achieved promising results at 3-second windows, though performance at 1–2 second windows were only slightly better than keystroke-only or multimodal baseline approaches (Table I).

Sensor fusion has consistently outperformed unimodal approaches. Kumar et al. [6] pioneered fusion of keystroke rhythms with device motion, achieving 11.45% EER at 20–40 second windows. Shen et al. [9] advanced multimodal fusion to 1.8% EER with 5–7 second windows. Sitová et al.'s [14] HMOG framework achieved 1.5% EER by integrating touch with inertial sensors, though requiring 25-second windows. Ray-Dowling et al. [3] confirmed through systematic benchmarking that multimodal fusion universally outperforms single-modality systems, matching HMOG's 1.5% EER at 25 seconds. Ray et al. [15] demonstrated that even in low-motion contexts, IMU sensors capture identity-discriminative micro-movements, achieving 2.4% EER during seated form-filling. The consistent finding is that fusion improves accuracy, but phone-sensor-only approaches plateau at 5–25 second windows for high accuracy.

Human Activity Recognition and Body Motion Features

The HAR community has developed methods for modeling IMU signals. Pathirage et al.'s [20] TEZARNet projects IMU sequences into learned latent spaces, demonstrating effective handling of signal variability, which is a challenge shared by behavioral biometrics when users exhibit day-to-day variation. Chandirakumar et al. [21] showed that attention mechanisms can effectively fuse heterogeneous sensor streams, directly applicable to our fusion of keystroke events with smartphone and body motion data. De Silva et al. [22] demonstrated that attention maps can identify which temporal sub-sequences contribute most to recognition, paralleling our interest in understanding which moments around keystroke events carry the most discriminative body motion information.

Gait-based authentication established that body motion contains biometric value. Derawi et al. [18] achieved 20.1% EER using smartphone accelerometers during walking, though gait approaches require continuous locomotion. Our previous work [11] demonstrated that full-body motion features dramatically improve swipe-based authentication, achieving 6.4% EER in 1–2 second windows compared to 15–20% for swipe-only approaches. Our key innovation was using device interactions as temporal anchors for synchronized body motion extraction. However, swipes and keystrokes differ fundamentally: swipes are continuous gestures while keystrokes are discrete events. Whether body motion integration translates to keystroke authentication, and whether the precise temporal anchors keystrokes provide yield even better performance, motivates this work.

III. Methodology

To comprehensively evaluate the effectiveness of body movement features in conjunction with keystroke dynamics for short-window smartphone authentication, we utilized the publicly available dataset from our previous research [11]. This dataset provides a unique opportunity to investigate how body motion features that demonstrated effectiveness for swipe-based authentication perform when adapted to keystroke dynamics[1].

Figure 1 illustrates our end-to-end methodology, from data collection through authentication decisions. The process begins with synchronized capture of keystroke events, smartphone IMU data, and 3D motion capture data. Keystroke

[1] The study was approved by the NYIT's Institutional Review Board under protocol BHT-1290, and all subjects provided informed consent. [19]

TABLE I
QUANTITATIVE COMPARISON OF KEYSTROKE AND MULTIMODAL AUTHENTICATION APPROACHES

Study	Year	Window	EER (%)	Modalities	Classifier	N	Dataset	Key Innovation
Keystroke-Only Approaches								
Shen et al. [8]	2016	11s	1.7–9.0	Keystroke + Touch	Logistic Regr.	71	Custom	Active authentication
Li et al. [7]	2019	5s	4.7	Keystroke	Data augment.	100	HMOG	Addressed data scarcity
Dutta et al. [2]	2025	<5s	2.48–3.60	Keystroke (2 Hz)	SM	160	Custom	Multi-input types
Roy et al. [12]	2025	2.5s	<3.0	KS + Soft Bio.	Multiple	—	—	Low sampling rate
Senarath et al. [13]	2023	1–3s	6.08	Keystroke	RF	31,400	AaltoDB	Short window analysis
Multimodal Fusion Approaches								
Kumar et al. [6]	2016	20–40s	11.45	Keystroke + IMU	Fusion	28	Custom	Combined typing + motion
Shen et al. [9]	2018	5–7s	1.8	Multi-sensor	SVM	102	Custom	Multiple sensors
Sitová et al. [14]	2016	25s	1.5	Touch + IMU	Multiple	100	HMOG	HMOG framework
Ray-Dowling et al. [3]	2022	25s	1.5 / 0.2	Motion + Touch	Fusion	100, 115	HMOG, BB-MAS	Systematic benchmark
Ray et al. [15]	2021	—	2.4 / 6.9	Accel. + Gyro.	Score fusion	—	—	Seated form-filling
Deep Learning Approaches								
Senarath (BehaveFormer) [16]	2023	Not specified	1.80–12.04	Keystroke	STDAT	60,527	AaltoDB, HMOG, HuMIdb	Dual attention transformer
Senarath (BehaveFormer) [16]	2023	Not specified	2.95–3.62	KS + IMU	STDAT	527	HMOG, HuMIdb	KS + IMU fusion
Nguyen et al. [17]	2024	Not specified	2.95	KS + IMU	STDAT	428	HuMIdb	Joint temp.-spatial attn.
Body Motion Integration								
Derawi et al. [18]	2010	∼26s	20.1	Accel. (gait)	—	31	Custom	Gait recognition
Cariello et al. [11]	2025	1–2s	6.4	Swipe + Body + IMU	Multiple	39	NYIT-SAD [19]	Swipe-based full-body
This Work	**2025**	**1–2s**	**1.5–1.9**	**KS + Body + IMU**	**RF, SM**	**42**	**NYIT-SAD**	**KS-based, ultra short bursts**

events serve as temporal anchors that define extraction windows (keystroke event perimeters) for computing statistical features from both phone and body motion signals. These multimodal features are then used to train user-specific authentication models, which generate accept/reject decisions based on incoming test samples.

A. Dataset Characteristics and Participant Demographics

The dataset employed in this study consists of multimodal biometric data collected from 82 participants using carefully controlled experimental protocols designed to capture naturalistic smartphone usage patterns across different postural conditions. Data collection utilized an iPhone XR smartphone equipped with a custom typing application designed to precisely record keystroke events and synchronized movement data from built-in accelerometer and gyroscope sensors.

Concurrent with smartphone data collection, full body motion data was captured using a Vicon Nexus System motion capture setup, which represents precise three-dimensional movement tracking. Participants wore 39 reflective markers positioned on their bodies based on the established Vicon Nexus Plug-In-Gait model [23], which provides comprehensive coverage of major body segments and joints.

Of the 82 participants, data from 41 (walking) and 42 (sitting) users were usable; the remainder contained only swiping data. For each posture, users with two sessions (29 walking, 30 sitting) were used for training and testing, while single-session users (12 in each condition) provided additional impostor data, with Session 1 serving as training and Session 2 as testing, this ensures no data leakage between training and testing sets. Each session contained approximately 12 typing trials of fixed 60-second duration. The trials were evenly distributed across two postural conditions: six trials collected while the participant maintained a seated position, and six trials collected while the participant walked at a comfortable pace in a square pattern on the floor.

On average, each user generated 111 keystrokes per trial when walking, and 126 when sitting. Additionally, on average, each user generated 1,458 keystrokes while sitting, and 1,253 keystrokes while walking across their available sessions. The dataset included adults of various ages, varying between 18 and 74 years old. Participants aged 18–32 years contributed 386 trials totaling 61,047 keystrokes (34,448 sitting / 26,599 walking), averaging 156 keystrokes per trial (164 sitting, 144 walking). Older participants aged 48–74 years produced 455 trials with 34,377 keystrokes total (18,674 sitting / 15,703 walking), averaging 75 keystrokes per trial (80 sitting, 71 walking). No participants were between 33 and 47 years old.

B. Keystroke-Aligned Feature Extraction Methodology

The adaptation of our previously developed feature extraction methodology from swipe-based to keystroke-based authentication required fundamental modifications to address the different temporal characteristics of these interaction modalities. We developed the concept of keystroke event perimeters, which represent temporal windows of fixed duration centered around individual keystroke events.

To determine optimal perimeter durations, we systematically evaluated perimeter lengths of 100 milliseconds, 400 milliseconds, and 800 milliseconds, investigating how perimeter duration affects feature quality and authentication performance. The selection of appropriate perimeter durations involves balancing several competing considerations: shorter perimeters reduce the likelihood of overlap between consecutive keystroke events but may not capture sufficient motion data, while longer perimeters provide more comprehensive motion data but increase the probability of temporal overlap.

Phone Motion Features. For each keystroke event perimeter, we extract comprehensive statistical representations of both smartphone motion sensor data and motion capture body posture data. From smartphone accelerometer and gyroscope

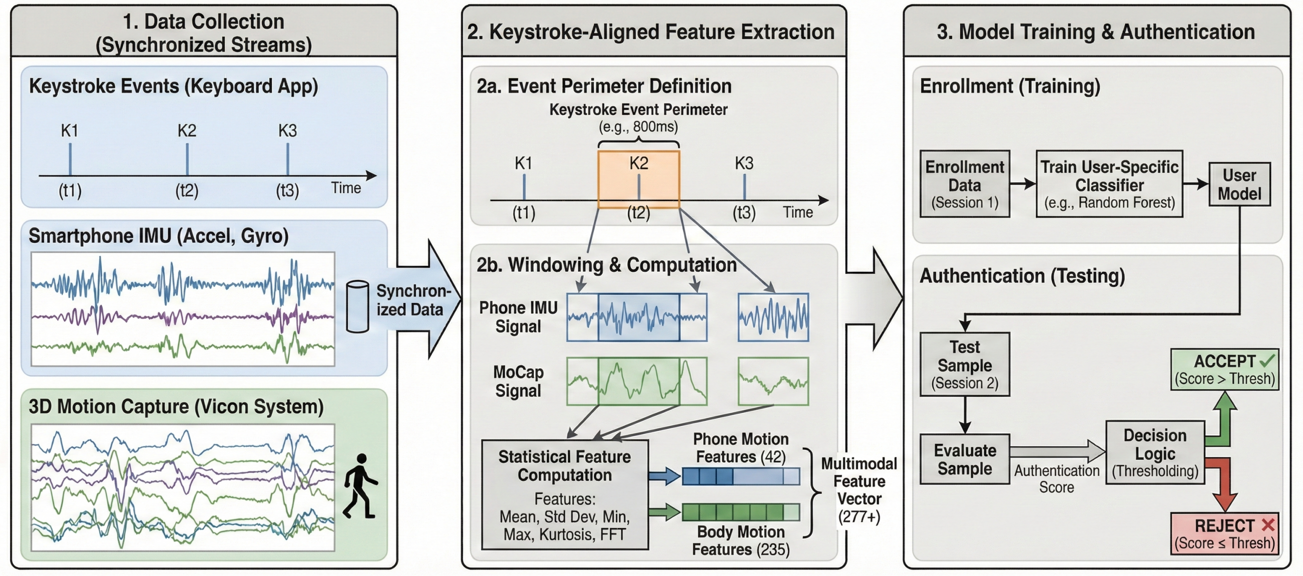

1. Data Collection (Synchronized Streams)
Keystroke Events (Keyboard App)
K1
K2
K3
(t1)
(t2)
(t3)
Time
Smartphone IMU (Accel, Gyro)
Synchron-ized Data
3D Motion Capture (Vicon System)
2. Keystroke-Aligned Feature Extraction
2a. Event Perimeter Definition
Keystroke Event Perimeter (e.g., 800ms)
K1
K2
K3
(t1)
(t2)
(t3)
Time
2b. Windowing & Computation
Phone IMU Signal
MoCap Signal
Statistical Feature Computation
Features: Mean, Std Dev, Min, Max, Kurtosis, FFT
Phone Motion Features (42)
Body Motion Features (235)
Multimodal Feature Vector (277+)
3. Model Training & Authentication
Enrollment (Training)
Enrollment Data (Session 1)
Train User-Specific Classifier (e.g., Random Forest)
User Model
Authentication (Testing)
Test Sample (Session 2)
Evaluate Sample
Authentication Score
Decision Logic (Thresholding)
ACCEPT ✓ (Score > Thresh)
REJECT ✗ (Score ≤ Thresh)

Fig. 1. End-to-end process flow of our keystroke-aligned motion authentication system. Keystroke events serve as temporal anchors that define extraction windows for synchronized phone IMU and motion capture data. Statistical features computed over these windows are used to train user-specific classifiers for authentication. K1, K2, and K3 represent individual keystroke events, and t1, t2, and t3 represent the points in time that the events occur.

signals, we compute five statistical measures for each sensor axis: mean values, standard deviations, kurtosis values, minimum values, and maximum values. Additionally, we compute frequency domain features using Fast Fourier Transform (FFT) analysis.

Motion Capture Features. Motion capture features undergo similar statistical analysis, with the same five statistical measures computed for each of the 47 body posture features during each keystroke event perimeter. Following the analytical framework established in our previous research, we organize motion capture features into three anatomically meaningful body regions defined in Table II. We discuss the biomechanical rationale for selecting these body motion features in Section III-C.

Combined Feature Set. Our feature extraction method resulted in 42 phone movement features (6 raw sensor features $\times$ 7 statistical representation features), and 235 motion capture features (47 body posture features $\times$ 5 statistical representation features) associated with each keystroke event. The total number of features used in our experiments corresponds to the number of selected keystrokes multiplied by the number of features available per keystroke event. When more than one keystroke is selected, we include the flight time between consecutive keystrokes as an additional distinct feature. An individual keystroke event feature vector is represented in Figure 2 For a single keystroke we have 277 features. For n-graphs, when $n > 2$, the size of the feature vector is therefore $n \times (42 + 235) + 1$. As a result, for a digraph, which has two keystrokes, we have 556 features. The additional feature is the flight time between the two keystrokes.

We further divide our features into the classes outlined in Table III, which represent the varying combinations of phone motion data with the entirety of the motion capture feature set, as well as the varying body regions of the motion capture feature set.

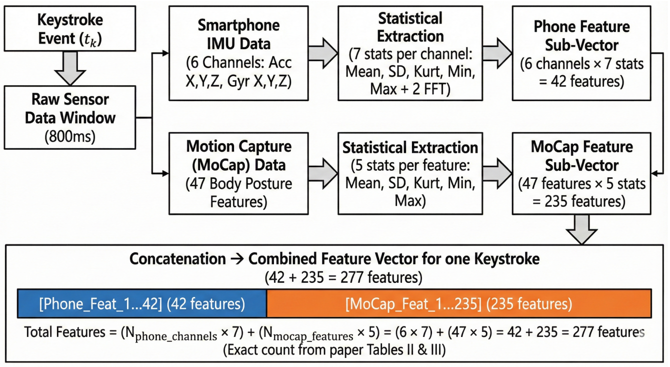

Keystroke Event (t_k)
Raw Sensor Data Window (800ms)
Smartphone IMU Data (6 Channels: Acc X,Y,Z, Gyr X,Y,Z)
Statistical Extraction (7 stats per channel: Mean, SD, Kurt, Min, Max + 2 FFT)
Phone Feature Sub-Vector (6 channels × 7 stats = 42 features)
Motion Capture (MoCap) Data (47 Body Posture Features)
Statistical Extraction (5 stats per feature: Mean, SD, Kurt, Min, Max)
MoCap Feature Sub-Vector (47 features × 5 stats = 235 features)
Concatenation → Combined Feature Vector for one Keystroke (42 + 235 = 277 features)
[Phone_Feat_1...42] (42 features)
[MoCap_Feat_1...235] (235 features)
Total Features = (N_phone_channels × 7) + (N_mocap_features × 5) = (6 × 7) + (47 × 5) = 42 + 235 = 277 features
(Exact count from paper Tables II & III)

Fig. 2. Example keystroke event feature vector composition showing phone motion features (left) and motion capture body posture features (right) extracted within a keystroke event perimeter. Note that this example illustrates a single keystroke event; for n-graphs, features from multiple consecutive keystroke events are concatenated, along with inter-keystroke timing features.

TABLE II
BODY-REGION MOTION CAPTURE FEATURE GROUPS USED IN REGION-LEVEL FUSION EXPERIMENTS.

Region (Count)	**Description**
Upper Body (12)	Head-to-phone and clavicle-to-phone distance, head-to-phone angles, neck-to-forehead angles, and shoulder-to-elbow joint angles.
Center Body (17)	Upper- and lower-spine angles, sternum-to-phone and clavicle-to-elbow distances, elbow-to-wrist, and forearm-to-wrist angles.
Lower Body (18)	Hip-to-phone and hip-to-elbow distances, hip angles, knee flexion angles, and ankle joint angles.

C. Biomechanical Rationale of Body Motion Feature Selection

Our feature selection from the motion capture system is

TABLE III
FEATURE CLASS CONFIGURATIONS USED IN EXPERIMENTS.

Feature Class	Description
Phone Only	Phone accelerometer, gyroscope, and keystroke dynamics features.
MoCap Only	Motion capture features across all body regions.
Phone + MoCap	Combined phone and all motion capture features.
Phone + Region	Combined phone features with MoCap features from a single body region (see Table II).

grounded in two principles from motor neuroscience and biomechanics. First, the nervous system utilizes multiple *reference frames* for motor planning: eye-centered, shoulder-centered, hand-centered, among others [24]–[26]. Measuring body-segment relationships from multiple anatomical landmarks (head, clavicle, sternum, hips) therefore captures complementary aspects of motor organization that no single reference point could provide. Second, each keystroke creates a force that propagates bidirectionally through the user's kinetic chain; how individuals absorb and stabilize these perturbations reflects habituated motor patterns that are difficult to consciously modify.

These principles explain the discriminative value of our features. *Center body features* (clavicle-to-elbow distances, forearm-wrist angles) capture the primary kinematic linkages in typing and showed consistently high mutual information (MI $>$ 2.0) across both postures. *Upper body features* (head-phone distance, shoulder angles) encode device-viewing and stabilization strategies. *Lower body features* serve posture-dependent roles: during sitting, knee and ankle angles establish stable baselines (MI $>$ 2.8); during walking, they encode gait characteristics. Statistical descriptors (mean, min, max, standard deviation, kurtosis) capture both static posture and dynamic movement variability. A comprehensive detailed biomechanical justification along with detailed transformation from raw 3D marker positions to body posture features (e.g., how "Left Knee Spherical Elevation" is computed from marker coordinates) in the supplemental material.

D. Keystroke Matching and N-Graph Configuration Analysis

The discrete nature of keystroke events enables several different approaches to temporal matching and sequence analysis. We systematically evaluated four distinct keystroke event matching strategies:

Character agnostic matching (Any Keystroke): Treats all keystroke events generically regardless of the specific key pressed, maximizing available samples.

Character specific matching (Keystroke): Maintains correspondence between keystroke events and single specific keys pressed.

Digraph matching: Considers sequences of two keystroke events, including inter-keystroke timing intervals.

Trigraph matching: Considers sequences of three keystroke events with timing relationships.

E. Classifier Implementation and Evaluation Methodology

We performed authentication experiments using four classifiers: Scaled Manhattan Distance (SM), Random Forest (RF), One-Class Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). Our classifier selection was guided by two considerations discussed in Section II: (1) demonstrated effectiveness with behavioral signals, particularly in short-window scenarios where Dutta et al. [2] showed that traditional methods remain competitive with more complex architectures when data is limited; and (2) practical deployment constraints, as these classifiers incur substantially lower memory, computation, and energy overhead compared to deep learning approaches [27], [28]. This is particularly valuable on mobile devices, where such resources are at a premium. We follow a standard verifier-based approach for each classifier, where every user had a model trained specifically for them [29]. SM is a one-class, distance-based classifier that computes the distance between a "template" feature vector (derived from enrollment data) and a "test" feature vector, with each feature dimension scaled by the standard deviation of the training data. Specifically, for n-dimensional template and test feature vectors $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, with per-feature standard deviations $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ computed from training data, the scaled Manhattan distance is $d = \sum_{i=1}^{n} |x_i - y_i|/\sigma_i$. This scaling ensures that features with higher variability do not dominate the distance calculation [30]. Random Forest is a two-class classifier that is an ensemble machine learning algorithm using a collection of decision trees to make predictions based on random feature subsets, and uses the outcomes of the trees to provide a balanced prediction [31]. One-Class Support Vector Machine is a one-class classifier that identifies the smallest hyperplane encompassing the training data, effectively distinguishing between normal and anomalous data points [32]. K-Nearest Neighbors is a two-class classifier that uses the distance between the test and training data to determine the closest neighbors, and then uses the majority class of those neighbors to classify the test data [33].

RF, SVM, and KNN were implemented using the Scikit-learn library [34]. For the two-class classifiers, we trained each user's model on their genuine data from Session 1 and data from five randomly selected impostors. These five impostors were chosen to represent a small subset of the user group that allows the model to avoid overfitting and were excluded from testing to prevent leakage. We then tested the model using the genuine user's data from Session 2, along with data from all remaining users (excluding the five randomly selected for training) collected from their Session 2 and the users with only one session.

We tuned the models using a standard grid search algorithm and applied specific preprocessing steps tailored to each method, and selected the best hyperparameter combination of the tuned model for each user based on the lowest Half Total Error Rate (HTER). HTER is defined as HTER $=$ (FAR $+$ FRR)$/2$, where FAR (False Acceptance Rate) is the proportion of impostor attempts incorrectly accepted, and FRR (False Rejection Rate) is the proportion of genuine attempts incorrectly rejected. HTER provides a single metric that balances both error types and represents general authentication performance independent of threshold selection. For Random Forest, we explored different numbers of trees (ranging from 100 to 1000 in steps of 100) and varied the `max_features` parameter between "*sqrt*", and 2 through 4. For KNN, we searched over the number of neighbors (2 through 9). For SVM, we implemented outlier filtering with Isolation Forests [35] at 10% contamination and reduced the dimensionality of the feature set with Principal Component Analysis (PCA) [36]. We then conducted the grid-search over two key SVM hyperparameters: ν and γ. The ν parameter (nu) provides an upper bound on the fraction of training errors and a lower bound on the fraction of support vectors, effectively controlling the trade-off between model complexity and training error, where lower values produce tighter decision boundaries. The γ parameter defines the influence radius of each training sample when using the RBF kernel, where smaller values create smoother, more generalized decision boundaries while larger values allow the model to capture finer local structure. We searched with ν set to 0.05, 0.1, 0.2, and 0.3, γ set to 0.001, 0.01, 0.1, and "*scale*", and the number of PCA components set to 2, 3, and 4.[2]

To generate scores, we used `predict_proba` for RF and KNN, and `decision_function` for SVM. For Scaled Manhattan Distance, we calculated each score s_i from the corresponding distance as $s_i = 1 - d_i/M$, where d_i is the distance between two vectors, and M is the maximum distance between all vectors. This provides a normalized set of scores to be used when calculating performance.

We deliberately selected lightweight classifiers to minimize computational overhead. To ensure secure and private processing, either locally on the user's device avoiding transmitting the data or via privacy-preserving protocols [14], [37], [38] low-resource models are essential for reducing power consumption and latency, particularly under frequent authentication intervals (every 1–2 seconds).

IV. Practical Body Motion Feature Collection

While our proof of concept evaluation utilized laboratory-grade motion capture equipment to establish the theoretical potential of body motion integration, practical deployment requires consideration of how the most discriminative features can be captured using commodity hardware available in modern smartphones and wearable devices. As in our previous research, we directly addressed the challenges of the practicality of these features by presenting a comprehensive analysis of how contemporary smartphone and wearable device capabilities can be leveraged to capture the body motion features identified as most valuable for keystroke authentication enhancement. The previous analysis is summarized here along with additional findings.

Recent advances in smartphone-based pose estimation and body tracking have demonstrated remarkable progress toward capturing the types of body motion features that our research identifies as most valuable for authentication enhancement. Ahuja et al. [39] developed "Pose-on-the-Go", which represents a significant advancement in smartphone-only full-body pose estimation using exclusively built-in sensors including cameras, inertial measurement units, and touchscreen interactions combined with inverse kinematics algorithms. Liang et al. [40] demonstrated an alternative approach through their development of smartphone-mounted fisheye camera systems for tracking both near-field hand gestures and full-body movements. Their system successfully extracts features including head orientation, head-to-camera distance measurements, and body-hand-phone connectivity relationships with accuracy levels that approach those achieved by our motion capture analysis. Kim et al. [41] explored a hybrid approach that combines wide-angle RGB cameras with narrow-field depth sensors and accelerometer data, all integrated within standard smartphone form factors. Their "OddEyeCa" system achieved body landmark tracking with average errors of 4.3 centimeters, which approaches the precision requirements for capturing the discriminative features identified in our research.

More recently, Xu et al. [42] propose *MobilePoser*, a real-time system for full-body pose estimation and 3D global translation that operates exclusively on IMUs embedded in commodity mobile devices—namely smartphones, smartwatches, and wireless earbuds. Unlike conventional motion capture systems that require intrusive, multi-sensor setups (e.g., Xsens with 17 dedicated IMUs), MobilePoser is designed to function with a dynamic and minimal configuration of just one to three ubiquitous devices. The system supports various real-world usage patterns, including a smartphone in the user's pocket, a smartwatch on either wrist, and a unified IMU stream derived from a pair of earbuds. These devices, despite their lower-fidelity sensors and heterogeneous sampling rates (e.g., 60 Hz for the watch, 25 Hz for the AirPods), are fused through a multi-stage neural network architecture that first infers joint positions and orientations, and then estimates global translation via a hybrid of foot-contact heuristics and direct velocity regression. To enhance realism and reduce artifacts like jitter and foot sliding, the system incorporates a physics-based optimizer. MobilePoser achieves mean per-joint vertex errors as low as 10.6 cm and operates at 60 fps on an iPhone 15 Pro, demonstrating both spatial accuracy and computational feasibility on edge devices. Notably, the system generalizes across sparse and variable sensor placements, achieving improved accuracy over prior work (e.g., IMUPoser) even when limited to a single device. This reinforces that full-body kinematic tracking with translation is practically achievable in-the-wild using only the mobile devices users already carry, without the need for external cameras or specialized wearables.

V. Results

In this section, we present the performance of motion capture features combined with phone motion features when aligned with keystroke events. We compare different feature-tuning methods and classifiers, and evaluate the most impactful features on authentication performance and latency. We explore unlabeled keystroke events, which we define as a generalized keystroke not associated with any characters,

[2] Before setting the limits for the grid search, we experimented with high values of γ, which did not perform well.

and match solely based on the keystroke event itself. We evaluate the performance of unlabeled keystroke events, single keystroke matches, digraph matching, and trigraph matching.

We evaluated the performance of each feature class (Table III) with respect to authentication latency, defined as the time required to collect sufficient input data for an authentication decision. For each feature class, we present results across three latency intervals: very short (1 and 2 seconds), short (3 and 5 seconds), and long (10 and 15 seconds); we divide our results by posture (sitting vs. walking). We use EER when comparing performance across different feature classes and authentication windows. EER is a standard biometric performance metric that indicates a system's error rate at the point where the FAR and the FRR are equal. As a result, EER provides a meaningful trade-off between FRR and FAR. We evaluated the impact of authentication window length on EER, particularly in relation to user behavior during typing tasks. The EERs presented in this paper are average EERs across all users included in each experiment.

A. Aspects of Keystroke Event Feature Tuning

To determine the best configuration of keystroke events to use for authentication, we analyzed the performance of our system across different keystroke feature combinations, matching methods, and keystroke event perimeter size and report results by modality (walking and sitting) and by classifier. We explore the use of keystroke event matching, single keystroke character matching, digraph matching, and trigraph matching. Keystroke event matching is where single keystroke events are decoupled from their corresponding character, and become unlabeled, allowing them to be matched generally, regardless of the specific key pressed. Additionally, we compared the performance of varying keystroke event perimeter sizes to determine the strongest configuration for all users in the shortest authentication window possible.

Keystroke Event Feature Extraction. To determine whether extracting features during keystroke events results in better performance than extracting the same features independently from keystrokes, we implemented two procedures: (1) we selected “events” uniformly at random throughout each session, so that they would not correlate with typing events, and extracted features; and (2) we selected “events” at fixed intervals in each trial. With walking, randomized events and fixed intervals performed worse than keystroke-matched events at 1-second authentication windows across all classifiers; with 2.42% and 2.82% EER respectively, compared to 1.47% EER with keystroke-matched events using the Random Forest classifier. A similar trend was observed with the Scaled Manhattan Distance classifier (3.39% and 3.92% vs. 2.61%), KNN (10.34% and 9.60% vs. 9.84%), and SVM (18.00% and 17.73% vs. 14.98%) which are all represented in Figure 3. These results are representative of our findings across all modalities and feature combinations, and demonstrate that matching movement signals with keystroke events directly improves authentication performance. This aligns with findings from the HAR literature, where De Silva et al. [22] showed that certain temporal sub-sequences contribute disproportionately to recognition; in our context, the moments surrounding keystroke events appear to carry the most discriminative motion information.

Keystroke Matching Configurations. We analyzed the performance of our system across different n-graph configurations, where we matched on any keystroke event, single characters, digraphs, and trigraphs. At a one-second window, the results reveal substantial variation across classifiers, though single-keystroke strategies, with matching on any keystroke event consistently outperforming all other matching strategies. In sitting conditions, Scaled Manhattan Distance achieves 6.1% EER with keystroke events and 12.9% with single-character matches, while digraphs and trigraphs perform far worse at 35.6% and 43.2%. Random Forest delivers the best overall performance, with error rates between 1.54% and 2.9% across all n-graph types. KNN is notably weaker, with 10.4% for single-character and roughly 19% for digraphs, trigraphs and any keystroke event. Meanwhile SVM yields uniformly higher error, ranging from 23.2% to 25.7%. Walking trials follow a similar pattern: SM again performs strongly with 2.6% for keystroke events and 8.9% for single-character, compared to 35.0% and 37.6% for digraphs and trigraphs. Random Forest remains robust, between 1.5% and 3.3% across strategies. KNN performs moderately, from 9.8% on keystroke events to 14% on digraphs and trigraphs, while SVM sits in the mid-range with error between 15.0% and 18.0%. Collectively, these findings demonstrate that higher-order n-graphs systematically degrade performance, particularly under SM and KNN, whereas Random Forest and SM with single-keystroke features provide the most reliable outcomes (see Figure 4).

Additionally, we reviewed the number of samples available when using keystroke event matching in comparison to digraph matching, and observed a 49.6% decrease in the number of samples when walking, and 41.5% decrease in the number of samples while sitting. The number of samples decreased even further when using trigraphs, with a decrease of 90.7% when walking and 88.7% when sitting. The sample loss between keystroke event matching, and keystroke character matching is only 12.5% when walking and 10.4% when sitting. This demonstrates that generic keystroke event matching provides a notable increase in the number of samples available for authentication, while maintaining a low EER.

Keystroke Event Perimeter Length. To assess the impact of keystroke event perimeter length on performance, we used perimeter lengths of 100 ms, 400 ms, and 800 ms. We then calculated an average EER across all users for respective authentication windows of 1, 2, 3, 5, 10, and 15 seconds. We observed that the performance of varying keystroke event perimeters clustered tightly across the different modalities, with 800 ms representing the middle ground of these clusters across all classifiers we tested. We present these results in Figure 5. Performance between event perimeter sizes is notable when comparing phone motion features alone. However, once we integrated motion capture features, performance improve-

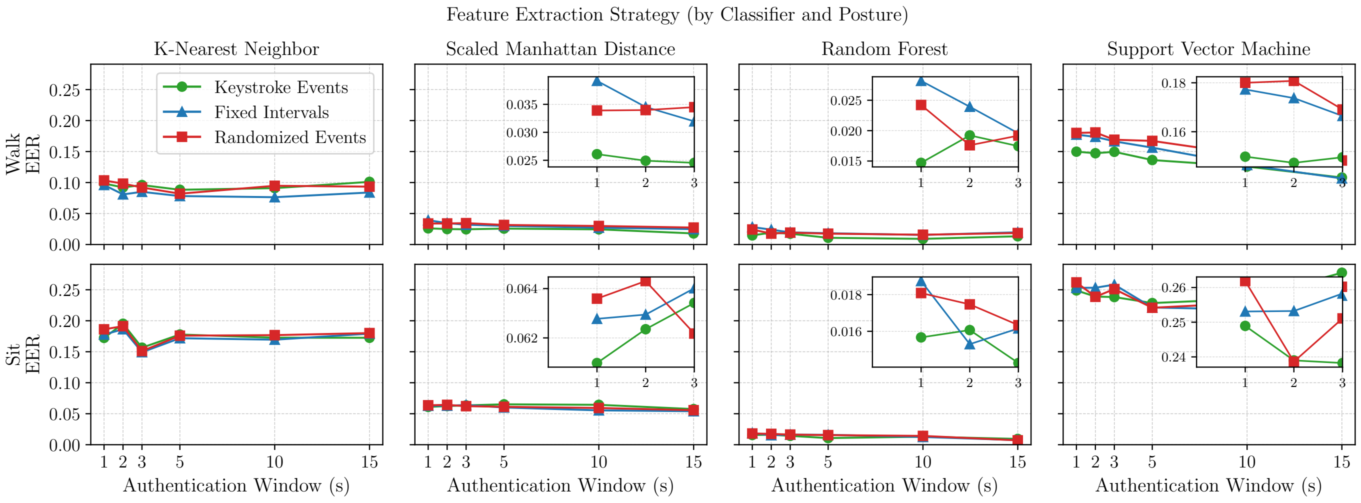


Fig. 3. EERs across increasing authentication windows of varied feature extraction indicators. Using an 800 ms event perimeter with smartphone and motion capture features for the walking and sitting modality using KNN, SM, RF, and SVM Classifiers.

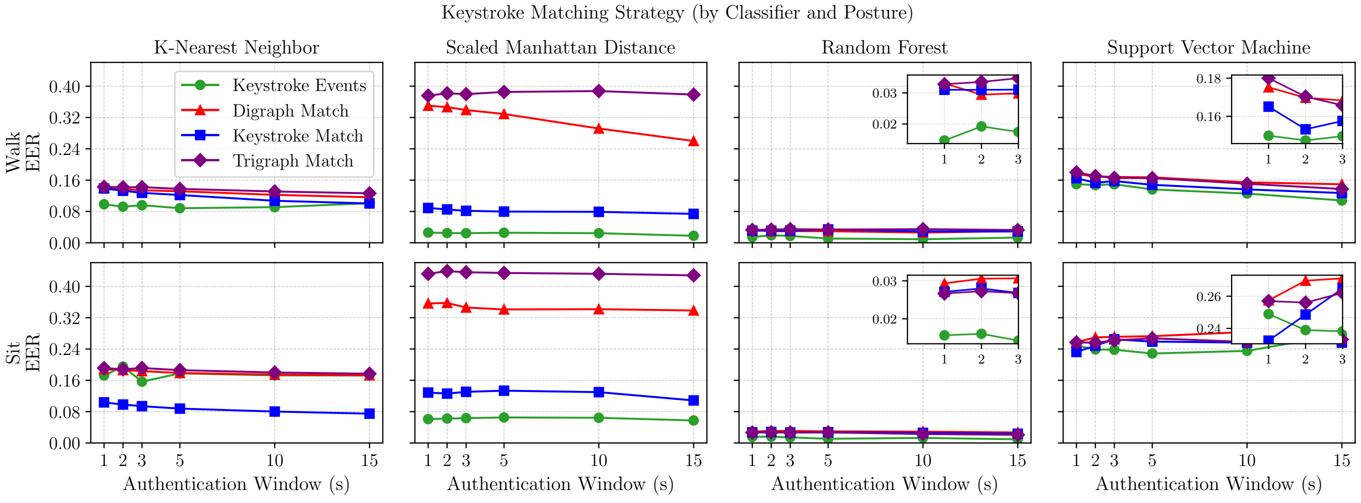


Fig. 4. EERs across increasing authentication windows of varying methods of keystroke matching for the walking and sitting modalities with a 800 ms keystroke event perimeter, using using KNN, SM, RF, and SVM Classifiers.

ment across perimeters became minimal, and as such evaluate our experiments with an 800 ms keystroke event perimeter.

B. Fusion of Smartphone and Motion Capture Features

We analyzed the performance of our system across different feature configurations and authentication window lengths. We found that integrating motion capture features consistently improves performance across all authentication window lengths. When sitting, the addition of motion capture features from all body regions resulted in an EER of 1.57% at 1-second, compared to 9.56% for phone motion features alone with the Random Forest classifier. A similar improvement was observed for walking, where EER was reduced to 1.47% with motion capture features compared to 11.97% with phone features alone. This trend was consistent across classifiers: with the Scaled Manhattan Distance classifier, sitting performance improved from 18.37% to 6.24% and walking from 17.40% to 3.31%; with KNN, sitting improved from 23.63% to 17.87% and walking from 22.66% to 9.93%. SVM classifiers struggled to perform well overall, though even there motion capture features reduced error rates in walking scenarios (23.19% to 16.90%). These results show that the inclusion of motion capture features provides a pronounced reduction in EER across classifiers and authentication window lengths, demonstrating that even at very short authentication windows in both sitting and walking scenarios, motion capture features significantly contribute to authentication accuracy.

We also explored the contribution of the body-region-specific motion capture features to authentication accuracy. To this end, we divided features into upper-, center- and lower-body regions (see Section III). We report these results across authentication window lengths in Table IV for walking and sitting of our best performing classifier across experiments, Random Forest.

This article has been accepted for publication in IEEE Transactions on Biometrics, Behavior, and Identity Science. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBIOM.2026.3674400

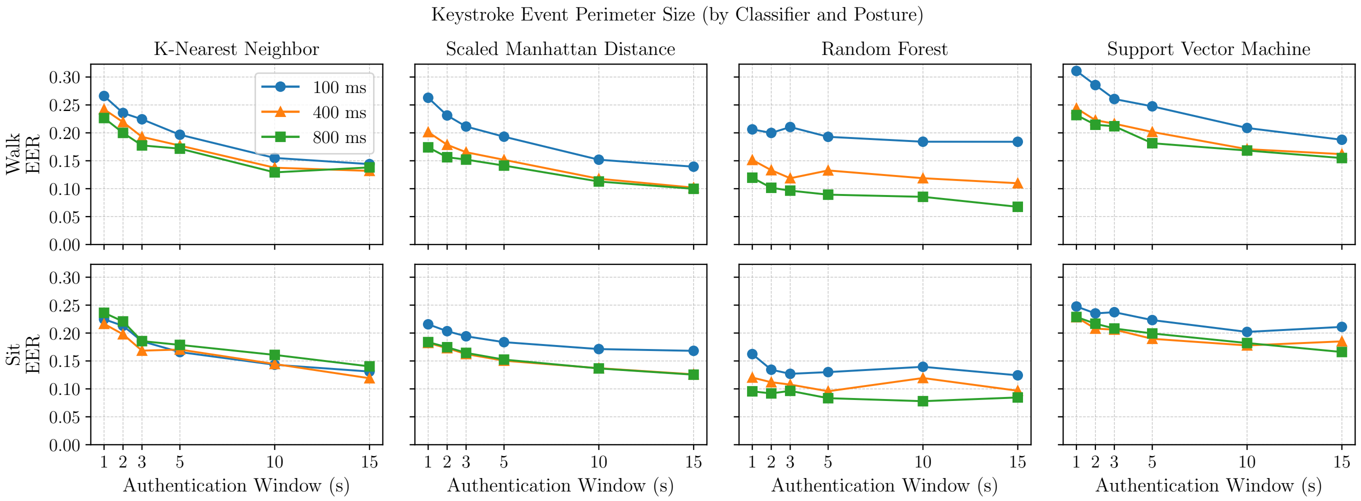


Fig. 5. EERs of the features extracted from the smartphone only across increasing authentication window sizes, comparing keystroke event perimeter for the walking and sitting modality using KNN, SM, RF, and SVM Classifiers.

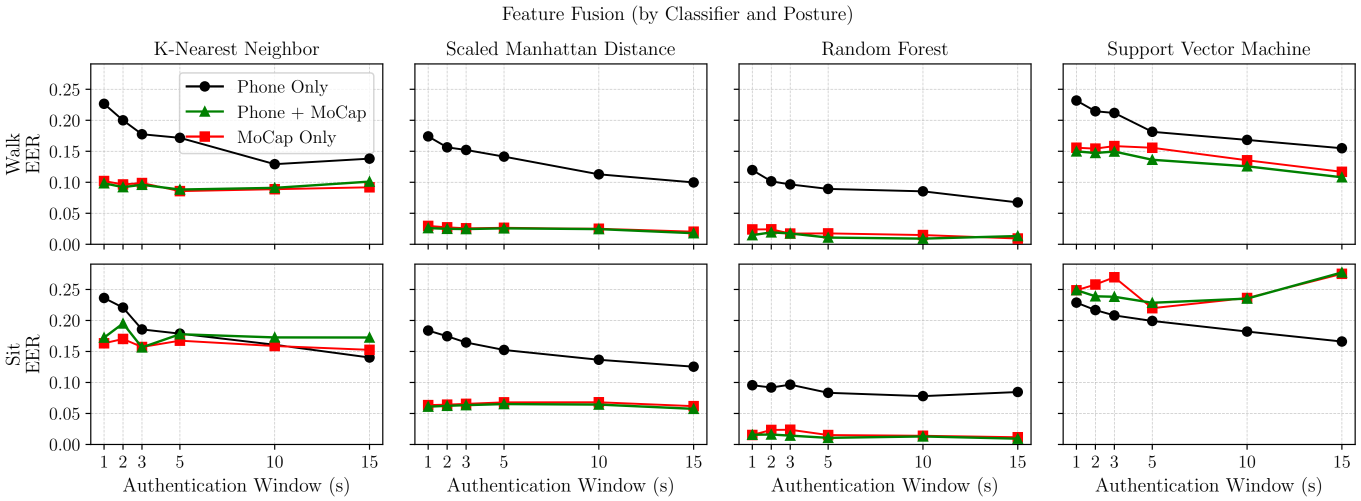


Fig. 6. EERs of smartphone and motion capture features, evaluated separately and together, for the walking and sitting modality using KNN, SM, RF, and SVM Classifiers.

We observed that, when walking, the center region combined with phone motion features yielded the lowest region-specific authentication error rates, with an EER of 2.94% at 1-second. In the same setting, the worst-performing region was the lower region, with an EER of 5.28% EER at 1-second. This is likely due to the amount of motion-induced noise recorded by motion sensors while the user is walking. When sitting, the center region combined with the phone motion features yields an EER of 3.05% EER at 1-second, while the worst performing region is the upper region with an EER of 6.49% EER at 1-second. This is likely due to the static nature of these features while the user is sitting.

Across both walking and sitting, we saw the best performance when all regions are combined, with Random Forest, and we achieved EERs of 1.47% while walking and 1.57% while sitting in 1-second. These error rates are substantially lower than those reported in similar settings in previous work, including [7], [11], [14]. This is particularly remarkable given that the previously reported EERs correspond to authentication windows that are longer than one second. These accuracy improvements likely stem from two factors: first, the individual distinctiveness of body motion features [11]; and second, the enhancement offered by extracting features from event perimeters rather than from individual samples. By capturing localized motion patterns instead of isolated “snapshots”, we better preserve the user’s unique behavioral signatures over time. Using these features, we obtained very low EERs between 1.07%–1.92% for 1–5 second windows, with the lowest EER (0.91%) with 10-second windows. These results confirm that our features are meaningful in identifying behavioral characteristics of users while typing.

The integration of motion capture data with phone features demonstrated a consistent improvement in performance across all authentication window lengths—the most significant being

TABLE IV
EER (%) FOR PHONE AND MOTION CAPTURE FEATURES BY BODY REGION, POSTURE, AND AUTHENTICATION LATENCY USING RANDOM FOREST.

Posture	Region	1s	2s	3s	5s	10s	15s
Walking	Upper	3.57	3.43	2.88	3.38	3.57	3.42
	Center	2.94	3.28	3.43	2.40	2.75	1.12
	Lower	5.28	7.74	7.81	5.99	6.10	5.20
	All Regions	1.47	1.92	1.74	1.08	0.91	1.33
Sitting	Upper	6.49	4.68	3.55	3.65	3.65	3.72
	Center	3.05	3.62	2.71	2.27	2.01	2.22
	Lower	5.34	4.98	3.19	3.53	3.11	3.83
	All Regions	1.57	1.61	1.43	1.07	1.30	0.93

the shortest window length of just 1 second, where our features achieved very low EERs. This is a strong indicator that the features extracted from the motion capture data are meaningful and provide a strong performance boost to the overall system.

C. Classifier Performance Evaluation

When we complemented motion capture features with smartphone motion features aligned to keystroke events, the best performing classifiers were Random Forest and Scaled Manhattan Distance. Random Forest performed best in both walking and sitting modalities, maintaining a low EER of 1.57% when sitting and 1.47% when walking at a 1-second authentication window. Scaled Manhattan Distance performed similarly when walking, with an EER of 3.31% at 1-second, though it performed worse when sitting, with an EER of 6.24% at 1-second. K-Nearest Neighbor showed improvement relative to the phone-only baseline but remained less effective overall, producing EERs of 17.87% when sitting and 9.93% when walking at 1-second. One-Class Support Vector Machine was our worst performing classifier and struggled to provide meaningful results, with error rates exceeding 25% when sitting and 16.9% when walking. The full classifier performance can also be seen in Figure 6.

D. Ablation of Motion Features and Keystroke Alignment

The experiments presented in Sections V-A and V-B are representative of our ablation analysis. Section V-A isolates the contribution of keystroke alignment by comparing keystroke-triggered feature extraction against fixed-interval and randomized sampling (Figure 3). Section V-B isolates the contribution of each feature modality by evaluating phone-only, MoCap-only, and fused configurations (Figure 6). Together, these results establish that both keystroke alignment and multimodal fusion are necessary components: alignment ensures features capture typing-relevant motion patterns, while fusion provides complementary behavioral signatures that enable reliable authentication within 1 to 2 second windows.

VI. Discussion

The successful adaptation of body motion integration from swipe-based to keystroke-based smartphone authentication represents more than a straightforward extension of existing techniques. Our research reveals fundamental insights into the nature of multimodal biometric authentication and demonstrates that the benefits of motion enhancement can be realized across different interaction modalities.

A. Keystroke vs. Swipe Authentication Performance

One of the most significant findings of our research is the superior authentication performance achieved by keystroke-aligned motion features compared to our previous swipe-based results. The achievement of 1.5% Equal Error Rate for keystroke authentication compared to 6.4% EER for swipe authentication under similar conditions suggests fundamental differences in how body motion features interact with different types of smartphone interactions.

The superior performance of keystroke authentication likely stems from several factors related to the temporal and biomechanical characteristics of typing versus swiping activities. Keystroke events provide discrete, precisely timed temporal anchors that enable more accurate alignment between interaction events and body motion patterns. The repetitive nature of keystroke events within typing sequences also provides multiple authentication opportunities within short interactions.

B. Physical Interpretation of Discriminative Features

Our results are consistent with the biomechanical rationale outlined in Section III: center-body features, particularly clavicle-to-elbow distances and forearm-wrist angles, appear to provide strong discriminative signals for authentication (Table IV). We hypothesize that these features capture how users position their arms relative to their torso while typing in a characteristic “typing posture” shaped by individual body geometry and habituated motor learning. The forearm-wrist angles may additionally encode fine motor control patterns during keystroke execution, which could be difficult to consciously modify and thus may contribute to robustness against imitation. The consistent performance across sitting and walking conditions (EER difference of only 0.1%) suggests that our features may capture relatively stable aspects of each user’s movement coordination rather than posture-specific artifacts, though further investigation across additional conditions would be needed to strengthen this conclusion.

These findings also provide preliminary support for the ripple effect hypothesis introduced in Section I: the discrete impact of each keystroke propagates through the user’s kinetic chain, and we posit that the manner in which different users absorb and stabilize this disturbance may be influenced by factors such as muscle tone, joint stiffness, and habituated postural strategies. This could create user-specific signatures captured within our keystroke event perimeters, though the precise biomechanical mechanisms warrant further study.

C. Short-Window Authentication Scenarios

Our demonstration that reliable authentication can be achieved within 1-second windows using keystroke-aligned motion features addresses a critical gap in smartphone authentication coverage. Many common smartphone interactions, including URL entry, query input, and brief message composition, occur within timeframes that have been inaccessible to existing behavioral biometric authentication approaches.

The extension of authentication coverage to these brief interaction scenarios transforms the fundamental character of continuous authentication from an occasional security check based on extended usage sessions to a pervasive security capability that can provide ongoing identity verification throughout typical smartphone usage patterns.

D. Practicality via Wearable Device Sensors

We conducted a feature importance analysis that is available as supplemental material. This analysis demonstrates that arm positioning and movement patterns (particularly clavicle-elbow relationships and forearm-wrist angles) provide substantial biometric discrimination capability across both sitting and walking conditions. These features fall directly within the sensing capabilities of wrist-worn devices that many smartphone users already carry.

E. Computational Cost and Real-Time Feasibility

Deployment of this continuous authentication strategy requires consideration of computational overhead, energy consumption, and latency constraints on mobile devices. Our deliberate selection of lightweight classifiers (Random Forest, Scaled Manhattan Distance, KNN, and One-Class SVM) addresses these concerns directly.

The statistical features we extract (mean, standard deviation, min, max, kurtosis) require $O(n)$ computation per feature window, where n is the number of samples in the keystroke event perimeter. For our 800ms windows at typical smartphone IMU sampling rates (100Hz), this represents approximately 80 samples per window. The FFT-based frequency features require $O(n \log n)$ computation. These operations are well within the capabilities of modern smartphone processors and can be computed in under 1ms per keystroke event. [43]

When considering classification latency, for Random Forest (our best-performing classifier), inference time scales with the number of trees and tree depth [31]. With our configuration (100–1000 trees), classification of a single feature vector typically completes in 1–5ms on modern mobile processors [27]. SM requires only a single pass through the feature vector, completing in sub-millisecond time. These latencies are negligible compared to the 1–2 second authentication windows.

In the context of energy consumption, prior work on smartphone-based behavioral biometrics has demonstrated that similar feature extraction and classification pipelines consume approximately 2–5% additional battery over baseline usage when authentication occurs every 1–2 seconds [14], [28]. The computational simplicity of our statistical features and lightweight classifiers suggests comparable or lower energy overhead.

For real-world deployment, we envision an event-driven architecture where keystroke events trigger feature extraction from a rolling buffer of IMU data. This approach avoids continuous processing overhead while ensuring features are available immediately upon keystroke detection. The entire pipeline, from keystroke event to authentication decision, would likely complete in under 10ms, enabling truly real-time continuous authentication without perceptible delay.

VII. Conclusion

This proof-of-concept study demonstrates that body motion features when temporally aligned with keystroke events, provide substantial discriminative value for continuous smartphone authentication in very short (1–2 second) windows. Using controlled motion capture experiments, we establish which body regions and feature types hold the most promise for authentication—a necessary foundational step before pursuing real-world implementation with commodity sensors.

Our key findings include:

- Body motion integration reduces EER from 9.5–11.9% (keystroke-only) to 1.5–1.6% at 1-second windows—the first approach to achieve reliable authentication during brief typing interactions.
- The novel concept of *elastic temporal windowing* through keystroke event perimeters successfully captures the "ripple effect" of body motion around discrete keystroke events.
- Center-body features (clavicle-elbow distances, forearm-wrist angles) provide the strongest discriminative signals, features increasingly capturable via smartphone pose estimation and wrist-worn wearables.
- Performance remains consistent across sitting and walking postures (0.1% EER difference), demonstrating robustness to postural variation.

This work bridges the gap between laboratory-established potential and practical deployment by identifying which features warrant further development using commodity hardware. The convergence of advancing smartphone pose estimation capabilities, the widespread availability of wearable devices, and the lightweight computational requirements of our approach creates favorable conditions for translating these findings into deployable authentication systems. Our publicly available dataset and methodology provide a foundation for continued research in this direction.

Acknowledgment

We leveraged Opus 4.1, Opus 4.5, and Sonnet 4.5 by Anthropic (https://platform.claude.com/docs/en/about-claude/models/overview) and GPT-5 and GPT-5.2 by OpenAI (https://platform.openai.com/docs/models) for editing and grammar enhancement for some of the content in this paper. Figures 1 and 2 was created using Nano Banana Pro by Google DeepMind (https://gemini.google).

References

[1] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *2009 IEEE/IFIP International Conference on Dependable Systems & Networks*, 2009, pp. 125–134.

[2] S. Dutta, S. Roy, and U. Roy, "Advanced keystroke dynamics for secure smartphone authentication," in *Biologically Inspired Techniques in Many Criteria Decision-Making*. Cham: Springer Nature Switzerland, 2025, pp. 144–152.

[3] A. Ray-Dowling, D. Hou, and S. Schuckers, "Stationary mobile behavioral biometrics: A survey," *Computers & Security*, 2023.

[4] E. Maiorana and P. Campisi, *Keystroke Dynamics*. Cham: Springer Nature Switzerland, 2025, pp. 1370–1375.

[5] J. Chao, M. S. Hossain, and L. Lancor, "Swipe gestures for user authentication in smartphones," *Journal of Information Security and Applications*, vol. 74, p. 103450, 2023.

[6] R. Kumar, V. V. Phoha, and A. Serwadda, "Continuous authentication of smartphone users by fusing typing, swiping, and phone movement patterns," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Sep. 2016, pp. 1–8.

[7] Y. Li, H. Hu, and G. Zhou, "Using data augmentation in continuous authentication on smartphones," *IEEE IoT Journal*, vol. 6, no. 1, pp. 628–640, Feb 2019.

[8] C. Shen, Y. Zhang, X. Guan, and R. A. Maxion, "Performance analysis of touch-interaction behavior for active smartphone authentication," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 498–513, March 2016.

[9] C. Shen, Y. Li, Y. Chen, X. Guan, and R. A. Maxion, "Performance analysis of multi-motion sensor behavior for active smartphone authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 1, pp. 48–62, Jan 2018.

[10] G. Stragapede, R. Vera-Rodriguez, R. Tolosana, A. Morales, A. Acien, and G. Le Lan, "Mobile behavioral biometrics for passive authentication," *Pattern Recognition Letters*, vol. 157, pp. 35–41, 2022.

[11] N. Cariello, R. Eslinger, R. Gallagher, I. Kurtzer, P. Gasti, and K. S. Balagani, "Posture and body movement effects on behavioral biometrics for continuous smartphone authentication," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 7, no. 1, pp. 3–15, 2025.

[12] S. Roy, U. Roy, D. Sinha, and R. K. Pal, "Advancing smartphone sensor-based keystroke dynamics for implicit and active authentication: Addressing challenges and enhancing usability control," in *Applied Computing for Software and Smart Systems*. Springer Nature Singapore, 2025, pp. 63–96.

[13] D. Senarath, S. Tharinda, M. Vishvajith, S. Rasnayaka, S. Wickramanayake, and D. Meedeniya, "Re-evaluating keystroke dynamics for continuous authentication," in *Proceedings of the 3rd International Conference on Advanced Research in Computing (ICARC)*. Belihuloya, Sri Lanka: IEEE, 2023, pp. 202–207.

[14] Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani, "Hmog: New behavioral biometric features for continuous authentication of smartphone users," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 877–892, May 2016.

[15] A. Ray, D. Hou, S. Schuckers, and A. Barbir, "Continuous authentication based on hand micro-movement during smartphone form filling by seated human subjects," in *Proceedings of the 7th International Conference on Information Systems Security and Privacy - ICISSP*, INSTICC. SciTePress, 2021, pp. 424–431.

[16] D. Senarath, S. Tharinda, M. Vishvajith, S. Rasnayaka, S. Wickramanayake, and D. Meedeniya, "Behaveformer: A framework with spatio-temporal dual attention transformers for imu-enhanced keystroke dynamics," in *IEEE International Joint Conference on Biometrics (IJCB)*. Ljubljana, Slovenia: IEEE, 2023, pp. 1–9.

[17] K.-N. Nguyen, S. Rasnayaka, S. Wickramanayake, D. Meedeniya, S. Saha, and T. Sim, "Spatio-temporal dual-attention transformer for time-series behavioral biometrics," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 4, pp. 591–601, Oct 2024.

[18] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2010, pp. 306–311.

[19] N. Cariello, R. Gallagher, K. Balagani, I. Kurtzer, and P. Gasti, "New York Institute of Technology Smartphone Activity Dataset," 2025. [Online]. Available: https://www.nyit-lamp.com/dataset/dataset4/

[20] D. Pathirage, D. De Silva, S. Wickramanayake, D. Meedeniya, and S. Rasnayaka, "Tezarnet: Temporal zero-shot activity recognition network," in *Neural Information Processing, Communications in Computer and Information Science, 30th International Conference on Neural Information Processing (ICONIP)*. Springer, 2023, pp. 444–455.

[21] M. Chandirakumar, T. Kanagarajah, N. Kalanantharasan, S. Wickramanayake, and D. Meedeniya, "Human activity recognition using spatio-temporal dual attention with cross-sensor attention," in *2025 International Joint Conference on Neural Networks (IJCNN)*. Rome, Italy: IEEE, 2025, pp. 1–8.

[22] D. Y. De Silva, S. Wickramanayake, D. Meedeniya, and S. Rasnayaka, "Sez-harn: Self-explainable zero-shot human activity recognition network," *arXiv preprint*, 2025.

[23] "Vicon, oxford metrics, oxford, uk - full body modeling with plug-in gait," https://help.vicon.com/space/Nexus212/11249602/, accessed: 2025-08-03.

[24] C. L. Colby and M. E. Goldberg, "Action-oriented spatial reference frames in cortex," *Neuron*, vol. 20, no. 1, pp. 15–24, 1998.

[25] F. Lacquaniti, M. Le Taillanter, L. Lopiano, and C. Maioli, "Coordinate transformations in the control of cat posture," *Journal of Neurophysiology*, vol. 74, no. 3, pp. 1225–1235, 1995.

[26] L. M. McGuire and P. N. Sabes, "Spatial representations in the human brain for reaching movements," *Experimental Brain Research*, vol. 195, no. 2, pp. 215–226, 2009.

[27] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahn, "Estimation of energy consumption in machine learning," *Journal of Parallel and Distributed Computing*, vol. 134, pp. 75–88, 2019.

[28] A. Mahfouz, H. Mostafa, T. M. Mahmoud, and A. Sharaf Eldin, "M2auth: A multimodal behavioral biometric authentication using feature-level fusion," *Neural Computing and Applications*, vol. 36, no. 34, pp. 21 781–21 799, 2024.

[29] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," in *2009 IEEE/IFIP Intl. Conference on Dependable Systems and Networks*, Lisbon, 2009, pp. 125–134.

[30] L. Araújo, L. Sucupira, M. Lizarraga, L. Ling, and J. Yabu-uti, "User authentication through typing biometrics features," in *Intl. Conference Biometric Authentication*, vol. 3072, 01 2004, pp. 694–700.

[31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001.

[32] R. Vert, J.-P. Vert, and B. Schölkopf, "Consistency and convergence rates of one-class svms and related algorithms." *Journal of Machine Learning Research*, vol. 7, no. 5, 2006.

[33] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Annals of translational medicine*, vol. 4, no. 11, 2016.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[35] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*. IEEE, 2008, pp. 413–422.

[36] M. E. Tipping and C. M. Bishop, "Mixtures of Probabilistic Principal Component Analyzers," *Neural Computation*, vol. 11, no. 2, pp. 443–482, 02 1999.

[37] P. Gasti, J. Sedenka, Q. Yang, G. Zhou, and K. S. Balagani, "Secure, fast, and energy-efficient outsourced authentication for smartphones," *Trans. Info. For. Sec.*, vol. 11, no. 11, p. 2556–2571, Nov. 2016.

[38] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, "An overview of privacy-enhancing technologies in biometric recognition," *ACM Comput. Surv.*, vol. 56, no. 12, Oct. 2024.

[39] K. Ahuja, S. Mayer, M. Goel, and C. Harrison, "Pose-on-the-go: Approximating user pose with smartphone sensor fusion and inverse kinematics," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–12.

[40] C. Liang, C. Yu, X. Wei, X. Xu, Y. Hu, Y. Wang, and Y. Shi, "Auth+ track: Enabling authentication free interaction on smartphone by continuous user tracking," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021, pp. 1–16.

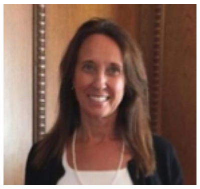
[41] D. Kim, K. Park, and G. Lee, "Oddeyecam: A sensing technique for body-centric peephole interaction using wfov rgb and nfov depth cameras," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020, pp. 85–97.

[42] V. Xu, C. Gao, H. Hoffmann, and K. Ahuja, "Mobileposer: Real-time full-body pose estimation and 3d human translation from imus in mobile consumer devices," in *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '24. New York, NY, USA: Association for Computing Machinery, 2024.

[43] A. M. Khan, M. H. Siddiqi, and S.-W. Lee, "Exploratory data analysis of acceleration signals to select light-weight and accurate features for real-time activity recognition on smartphones," *Sensors*, vol. 13, no. 10, pp. 13 099–13 122, 2013. [Online]. Available: https://www.mdpi.com/1424-8220/13/10/13099

Kiran S. Balagani is currently a Professor of Computer Science at New York Institute of Technology. His research interests are in anomaly detection, behavioral biometrics, and privacy-preserving biometrics. His research has been sponsored by the National Science Foundation and DARPA. He received the Ph.D. degree from Louisiana Tech University, USA.

Nicholas Cariello is currently a Ph.D. Candidate in Computer Science at the New York Institute of Technology. His research interests include behavioral biometrics, machine learning, deep learning, artificial intelligence, and cybersecurity. He expresses gratitude to his employer, IBM for their support of his doctoral research. He received his B.S. and M.S. degree in Computer Science from New York Institute of Technology, USA.



Rosemary Gallagher is a Professor at NYIT's Doctor of Physical Therapy Program. Gallagher graduated from Colorado State University in 1983 with a B.S. in Physical Education and has earned a B.S. in Physical Therapy in 1989 and a Clinical Doctorate in Physical Therapy in 2008 from Stony Brook University. She holds a Ph.D. from Rutgers University. She has over 30 years of clinical experience with a variety of patient populations including geriatric, neurologic, and orthopedic patients.

Paolo Gasti is a Professor of Computer Science at NYIT. His work focuses on behavioral biometrics, privacy-preserving biometric authentication, secure multi-party protocols, and network security. His research has been sponsored by the National Science Foundation and the DARPA. He received his B.S., M.S., and Ph.D. degrees from University of Genoa, Italy. Before joining NYIT, he worked as a research scholar at University of California, Irvine.

Isaac Kurtzer is currently an Associate Professor in the Biomedical Science Department of New York Institute of Technology – College of Osteopathic Medicine. He received his PhD in Neuroscience from Brandeis University (Waltham, MA) and completed a post-doctoral fellowship at Queen's University (Kingston, ON). His research on sensori-motor control has addressed motor learning, stretch reflexes, motor cortical processing, and rapid decision-making. He also teaches human neuroanatomy and biostatistics to the university's medical students.

Lam Nguyen is a Ph.D. candidate in Computer Science at the New York Institute of Technology (NYIT), a Machine Learning Researcher at NYIT's LAMP Lab, and a Lecturer at NYIT. His research focuses on continuous user authentication using behavioral biometrics and robust machine learning for human-centered security. He received his M.S. in Business Analytics from Adelphi University and previously worked as a Software Engineer at Silicon Valley Bank.