computers

MDPI

Article

Enhanced Security Access Control Using Statistical-Based Legitimate or Counterfeit Identification System

Aisha Edrah [1,2] and Abdelkader Ouda [1,*]

[1] Department of Electrical and Computer Engineering, Faculty of Engineering, Western University, London, ON N6A 5B9, Canada; aedrah@uwo.ca
[2] Faculty of Information Technology, Alasmarya Islamic University, Zliten 218521, Libya
* Correspondence: aouda@uwo.ca

Abstract: With our increasing reliance on technology, there is a growing demand for efficient and seamless access control systems. Smartphone-centric biometric methods offer a diverse range of potential solutions capable of verifying users and providing an additional layer of security to prevent unauthorized access. To ensure the security and accuracy of smartphone-centric biometric identification, it is crucial that the phone reliably identifies its legitimate owner. Once the legitimate holder has been successfully determined, the phone can effortlessly provide real-time identity verification for various applications. To achieve this, we introduce a novel smartphone-integrated detection and control system called Identification: Legitimate or Counterfeit (ILC), which utilizes gait cycle analysis. The ILC system employs the smartphone's accelerometer sensor, along with advanced statistical methods, to detect the user's gait pattern, enabling real-time identification of the smartphone owner. This approach relies on statistical analysis of measurements obtained from the accelerometer sensor, specifically, peaks extracted from the X-axis data. Subsequently, the derived feature's probability distribution function (PDF) is computed and compared to the known user's PDF. The calculated probability verifies the similarity between the distributions, and a decision is made with 92.18% accuracy based on a predetermined verification threshold.

Keywords: access control systems; identity verification; gait cycle analysis; smartphone sensors

check for updates

Citation: Edrah, A.; Ouda, A. Enhanced Security Access Control Using Statistical-Based Legitimate or Counterfeit Identification System. *Computers* **2024**, *13*, 159. https://doi.org/10.3390/computers13070159

Academic Editor: Lilatul Ferdouse

Received: 31 May 2024
Revised: 18 June 2024
Accepted: 19 June 2024
Published: 22 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the rise in the internet and digital technologies, people are using a wide array of online services and platforms for various purposes, such as email, social media, banking, shopping, entertainment, and productivity. Each of these services typically requires a separate account with its own set of login credentials. Security best practices recommend using unique passwords for each online account to mitigate the risk of credential reuse attacks [1]. This means that individuals need to create and manage a different password for every service they use, leading to a proliferation of passwords. In response, high-profile security breaches and data leaks have become increasingly common, leading to concerns about the security of sensitive personal information [2].

The proposed solution can enhance security and improve the user experience by facilitating passwordless and seamless access control methods for users [3,4]. This reduces the likelihood of users resorting to insecure practices, such as reusing passwords across multiple accounts or writing them down in insecure locations. Seamless access control methods, such as biometric identification [5,6], offer a frictionless user experience by eliminating the need for users to remember and input complex passwords repeatedly. This streamlined process improves user satisfaction and makes interactions with digital services more convenient.

Other seamless access control methods leverage biometric factors such as fingerprints [7], facial recognition [8], or keystroke dynamics [9]. These biometric authentication

techniques provide significantly enhanced security compared to traditional password-based identification systems. They offer a frictionless user experience by eliminating the need for users to remember and input complex passwords repeatedly. Biometric data is unique to everyone, making it more difficult for unauthorized users to access accounts or devices. In addition, seamless access control methods minimize system disruptions by reducing the frequency of identification-related issues and downtime. This is particularly important for mission-critical systems or applications, where any disruption can have significant operational impacts.

This research investigates the use of gait patterns as a biometric element for unique person recognition and identification [10]. Gait patterns, like fingerprints and facial traits, are unique to each person [11]. This is because gait includes a variety of elements such as stride length, cadence, posture, arm swing, and body weight distribution, all of which contribute to a unique walking pattern that may be utilized for identification. While other biometric identifiers may change over time due to factors like aging or injury, an individual's gait pattern tends to remain relatively consistent and stable over the long term. This consistency makes gait a reliable biometric identifier for verification purposes. Most importantly, gait pattern recognition can be performed from a distance without requiring physical contact or active participation from the individual being identified. This non-intrusive nature makes it suitable for access control applications where privacy and convenience are important considerations.

To consolidate the use of gait pattern recognition, we propose a novel security feature integrated into a smartphone app named Identification: Legitimate, or Counterfeit (ILC). ILC is a system of identification intended for detection and control. It employs a smartphone accelerometer sensor along with advanced statistical methods to detect the walking pattern and then deliver a response that identifies the smartphone owner. The ILC system normally maintains accelerometer data regarding the phone owner's walking pattern to help detect whether another person is carrying the phone.

The following scenario demonstrates the relevance of the ILC system as a strong security mechanism in access control systems, preventing unauthorized persons from using stolen devices to obtain access to sensitive resources or locations.

Sarah, a resident of a modern apartment complex, is returning home after a long day at work. As she walks closer to the gate, the Bluetooth Low Energy (BLE) beacon installed near the gate detects her smartphone's presence. The beacon sends a signal to the gate's ILC system, which is integrated with Sarah's mobile app. The management of the apartment complex has provided this app so that residents can use their smartphones to control various amenities, including the main gate. Upon receiving the signal, the gate's ILC system was able to recognize Sarah's walking steps and compare them with Sarah's walking pattern stored in the ILC system to verify her authorization. Once confirmed as a legitimate phone holder, the gate's ILC responds to initiate the opening sequence.

With a soft hum, the gate's motor activates, and the gate smoothly slides open, allowing Sarah to enter the complex without any physical interaction. Sarah appreciates the convenience of using her smartphone to access the gate, especially since she does not have to fumble for keys or access cards.

Let us now assume that, unbeknownst to Sarah, her smartphone was stolen earlier in the day while she was at work. The thief, now in possession of Sarah's smartphone, approaches the main gate of the complex. As the thief walks closer to the gate, the Bluetooth Low Energy (BLE) beacon detects the presence of the smartphone. The beacon sends a signal to the gate's ILC system and is waiting for the response. However, the gate's ILC system is designed to identify the legitimate from the counterfeit phone holder; thereby, ILC detects that the smartphone is being used by a thief and hence will not respond to the beacon, and it does not initiate the opening sequence. The thief quickly realizes that he or she cannot use the stolen smartphone to gain unauthorized access after making an unsuccessful attempt to enter the complex. Meanwhile, during these unauthorized access

attempts, the security administrators receive instant alerts on their monitoring dashboard, enabling them to take immediate action and maintain the integrity of the secured premises.

This ILC system not only offers convenience to residents like Sarah but also enhances security by ensuring that only authorized individuals with the corresponding mobile app and permissions can enter the apartment complex.

The primary contribution of this work is the integration of access control systems with smart lock technology. This integration allows businesses to achieve comprehensive security, streamline operations, and provide a unified experience for residents and employees. The integration offers a centralized system for managing access, remote functionalities, adherence to regulatory requirements, and smooth connectivity within intelligent building ecosystems.

The paper's structure is arranged in the following manner: Section 2 delves into the extant literature on gait cycle recognition, offering a comprehensive survey of the pertinent research carried out in this domain. Section 3 presents the proposed ILC system, including the assumptions and design goals, its constituent elements, and the methodology used for system development. Section 4 covers system analysis and validation, Section 5 discusses the experiments and results, and in Section 6, we draw our final conclusion and discussion and offer perspectives on future work.

2. Literature Review

The field of biometric identification based on gait analysis has evolved through intense efforts that have established it as a central area of academic research. This section examines previous work in this field to determine how our contribution stacks up against their accomplishments and whether there may be areas of overlap in goals before we can fully describe the significance of ILC.

N. Maulana et al. [12] developed a Role-Based Access Control (RBAC) system that ensures employees can access areas relevant to their roles by capturing fingerprints with BioLite N2, distributed at each door entry in the building. The system was able to respond and make decisions according to the correct access data for the employee role when using a single device. However, the RBAC system is constrained by its inability to clearly define roles for accurate access control across heterogeneous devices, the mapping of these roles according to their corresponding operations, and the restrictions related to time and location, which limit access to areas that are not predetermined or within time frames that have not been previously set. Also, another flaw in fingerprint systems is the need for the user to physically touch a scanner. Some people could find this uncomfortable [13].

S. M. R et al. [14] utilize a multi-modal framework to overcome the flaws of single-trait biometric systems. This approach improves user identity verification and security while maintaining processing speed using facial and iris traits. The findings show that combining facial and iris traits is more effective than using each individually. The authors' proposed integration requires significant storage, advanced algorithms, and processing power for both measures. However, the objective is consistently to devise a streamlined and lightweight solutions aimed at minimizing the risk of data intrusion and addressing the issues of resource intensity and privacy concerns.

V. Priyanka and G. K. [15] propose a method that employs a Dual Channel Convolutional Neural Network (DC-CNNPAD) to improve the accuracy of detection in the iris presentation attack. This system uses CNNs to address the challenge of effectively detecting fake iris presentations. The method demonstrated its reliability in distinguishing between real and artificial irises. However, limitations include the dependency on specific datasets, the potential variability in performance between different iris recognition systems, and the challenges of simulating real-world scenarios for presentation attack detection. Additionally, computational complexity and resource demands may restrict its use in low-resource environments or devices.

Cao et al. [16], Y. Wang et al. [17], X. Zeng et al. [18], and Q. Zou et al. [19] utilize Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) algorithms

for mobile gait recognition using smartphone sensors. These systems collect data using an accelerometer [16], or an accelerometer and a gyroscope [17–19].

In [16] the High Definition Learning Network (HDLN) system significantly improves feature-learning performance. This study shows that CNN combined with LSTM does not perform as effectively as CNN alone. Furthermore, it is complex and requires extensive training.

The study by [17] presents a recognition framework based on grayscale images to protect user-sensitive information using smartphones and tablets. The performance of user-recognition models on multiple devices is superior to that on single devices. However, this system struggles to adapt to the proposed continuous user identification system across a diverse range of smart devices with varying sensor configurations. The Edge Device Integration Architecture (EDIA) system [18] leverages edge computing for gait biometric identification and achieves good recognition performance, but it is only suitable for devices with low energy consumption and a limited dataset.

The paper by [19] employs a hybrid CNN and LSTM model for gait biometrics, using data collected in an unconstrained environment, to robustly extract walking features across pace and time domains. However, due to the high computational requirements for training, applying this technique to large datasets poses challenges.

The study by F. S. Chen et al. [20] introduces the PQRST Complex for user recognition, which employs an accelerometer to capture foot movement in three dimensions (X, Y, and Z axes). This method utilizes nine features divided into amplitude and intervals, enhancing the performance of the identification system. The study focuses on the swing phase and applies two ML classifiers. However, focusing on the swing phase may overlook additional insight from the stance phase, which could improve gait analysis for identification. The impact of large datasets on power consumption and computational demands is also highlighted. Applying this system to real-world scenarios, such as security systems, could validate the practical effectiveness of its performance.

S. Kumari et al. [21] introduce a smartphone-based identification framework that captures patterns of daily activity using an accelerometer, magnetometer, and gyroscope. The model achieves high performance using various machine learning (ML) algorithms, aiming to replace traditional verification methods (e.g., passwords, PINs) in banking and email, and to reduce computational time. Although the two classifiers achieve high performance, due to the challenges of computational cost and system processing time, their framework may not be the most pragmatic smartphone-based identification framework.

M.S. Axente et al. [22] propose a smartphone gait recognition solution utilizing built-in sensors on Android devices. The system achieves significant performance by employing histogram similarity with accelerometer data, defined by setting a bin to identify the minimum score between histogram recordings. This performance is further enhanced with gyroscope inputs for activity recognition. Nevertheless, the method's reliance on continuous data collection and processing, particularly when using multiple sensors, could strain a smartphone's battery resources.

P. Delgado-Santos et al. [23] introduce GaitPrivacyON, a mobile gait recognition method for user identification that merges CNNs and recurrent neural networks (RNNs), enhancing gait verification security while protecting sensitive data. Despite its success, the study highlights the challenge of simultaneously enhancing both identification and privacy.

Unlike previous investigations [16–19,21,23], which required more power and faced challenges in full training, our approach provides an efficient solution that rapidly identifies users without extensive training, maintaining low energy consumption during battery power identification. It effectively handles the complexity of full training scenarios, making it highly suitable for smartphones due to its higher performance with smaller datasets. In contrast to [22], our method considers the possible impact of handling sensitive information in large datasets, given that our model involves 12,000 samples. The suggested method emphasizes X-accelerometer peaks, improves user identification, and speeds up recognition.

A. Muro-De-La-Herran et al. [24] introduce the gait evaluation method using wearable sensors, capturing patterns from observed individuals using devices attached on their lower legs using a three-acceleration axis. The study provides a valuable insights but is limited by controlled settings and lack of long-term follow-up. Larger datasets and more reliable algorithms that can handle pattern variations are suggested. This method relies on wearable sensors worn on the body, with limitations, including restricted data range and discomfort caused by constant wear, as well as environmental constraints.

A summary of the aforementioned related work is presented in Table 1. The table clearly shows that each method has specific limitations that make it unsuitable for seamless access control system. Therefore, the goal of this work is to provide the ILC solution that uses a smartphone to address the aforementioned restrictions without requiring the user to wear any sensors. ILC makes it possible to implement seamless access control systems, providing a safe and unobtrusive method of controlling access.

Table 1. Literature review summary.

Method	Characteristics	Advantages	Limitations
RBAC System by N. Maulana et al. [12]	Uses BioLite N2 fingerprint scanners at each door entry to ensure access based on employee roles.	Ensures access control based on roles, simplifies management for single devices.	Inability to define roles accurately across heterogeneous devices, time and location restrictions, physical discomfort from touching scanners.
Multi-Modal Framework by S. M. R et al. [14]	Combines facial and iris traits for identity verification.	Improved security and processing speed, more effective than single-trait systems.	Requires significant storage, advanced algorithms, and processing power; resource-intensive.
DC-CNNPAD by V. Priyanka and G. K. [15]	Uses Dual Channel Convolutional Neural Network for iris presentation attack detection.	High accuracy in detecting fake iris presentations, suitable for smartphones, low energy consumption.	Challenges in enhancing both identification and privacy, dependency on specific datasets, computational complexity.
Gait Evaluation by A. Muro-De-La-Herran et al. [24]	Uses wearable sensors on lower legs to capture gait patterns.	Valuable insights into gait patterns, potential for clinical applications.	Limited by controlled settings, lack of long-term follow-up, discomfort from constant wear, environmental constraints.
HDLN System by Cao et al. [16]	Uses CNN and LSTM algorithms for mobile gait recognition with smartphone sensors.	Improved feature-learning performance, good recognition performance.	Complex, requires extensive training, high computational requirements, not suitable for large datasets.
PQRST Complex by F. S. Chen et al. [20]	Uses accelerometer to capture foot movement in three dimensions for user recognition.	Enhances identification system performance, focuses on swing phase.	May overlook insights from stance phase, high power consumption and computational demands.
Smartphone-Based Framework by S. Kumari et al. [21]	Captures daily activity patterns using accelerometer, magnetometer, and gyroscope.	High performance with various ML algorithms, aims to replace traditional verification methods.	High computational cost and system processing time, not the most pragmatic for smartphones.
Gait Recognition by M.S. Axente et al. [22]	Utilizes built-in sensors on Android devices for gait recognition.	Significant performance with histogram similarity, enhanced with gyroscope inputs.	Continuous data collection and processing strain battery resources.
GaitPrivacyON by P. Delgado-Santos et al. [23]	Merges CNN and RNN for mobile gait recognition.	Privacy-preserving, effective user identification.	High computational requirements, challenges in real-world application.

3. Identification: Legitimate or Counterfeit (ILC) System

Identification: Legitimate or Counterfeit (ILC), is designed to enhance the biometric-based access control system. ILC leverages the smartphone's capabilities and the advantages of the probability distribution function [25] to differentiate between the genuine owner and possible risks, such as counterfeit phone holders. By distinguishing between legitimate and counterfeit in real-time, ILC ensures that only authorized personnel gain access to restricted resources, thereby enhancing overall security and mitigating potential security threats. Subsections will outline the design objectives of the ILC system, its primary components, and the approach used to achieve the established goals.

3.1. ILC Assumptions and Design Goals

To ensure the completeness of ILC design, we have made the following assumptions and enlisted design goals:

3.1.1. Assumptions

- Assumption 1—Registration: The smartphone accelerometer sensor is activated and records data on the gait pattern of the phone's user.
- Assumption 2—Smartphone positioning: The smartphone is in the individual's pocket as they are walking towards the designated destination.
- Assumption 3—Sensing time frame: The individual is walking for at least 90 s, but not exceeding two minutes, traversing a predetermined distance of 100 m in a straight line.

3.1.2. Design Goals

- Goal 1—Seamless identification: ILC should be able to differentiate between a legitimate and counterfeit phone owner without requiring direct user involvement.
- Goal 2—Real-time user verification: ILC should process the sensing of the individual walk pattern, validate it with the stored data, and provide the outcome in real-time.
- Goal 3—High detecting accuracy: ILC should be able to achieve a detection accuracy of 90% or above in identifying counterfeit phone owners, according to the confusion matrix performance metrics [26].
- Goal 4—Lightweight system: ILC system should be designed for minimal memory usage, storing only less critical values on the owner's device.

3.2. System Components

The ILC system comprises two main components: the Know Your Owner (KYO) component and the Detecting and Control (DAC) component. The KYO component is responsible for registering the smartphone owner's gait data within the device. It also sets and updates the verification threshold used to verify the owner's identity. These data and verification thresholds can be updated automatically or manually at the owner's request if circumstances change. DAC is responsible for detecting the walking gait pattern of the phone holder and providing a real-time response based on whether the detected gait data belong to the owner of the smartphone or to someone else who is carrying the phone. Figure 1 illustrates the specific actions involved in the KYO and DAC components. The implemented approach to these actions is described next in the system methodology section.

Before exploring the system methodology, we will first provide an overview of gait cycle technologies. We will cover the essential aspects of the gait cycle, focusing on the stance and swing phases. Also, we explore the concept of the gait cycle, highlighting its length and the distinct patterns that vary among individuals. Moreover, we will explore the application of the probability density function (PDF) and key performance evaluation metrics in assessing the ILC system.

1. Gait Cycle Phases: key measurements in biometric identification, providing a valuable procedure in the gait of a person. The gait cycle comprises two main phases: the stance phase and the swing phase. Each phase is further divided into subphases [27–29] as shown in Figure 2. In this section, we provide an overview of the significance of these components in biometric identification.
    - Stance Phase: the period of time in which one foot makes contact with the ground and ends when the same foot contacts the ground again. This phase is divided into subphases, such as the loading response, mid stance, and terminal stance [30].
    - Swing Phase: the period of time in which one foot is flying and is not in contact with the ground. It also consists of subphases, such as initial, mid swing, and terminal swing [30,31].

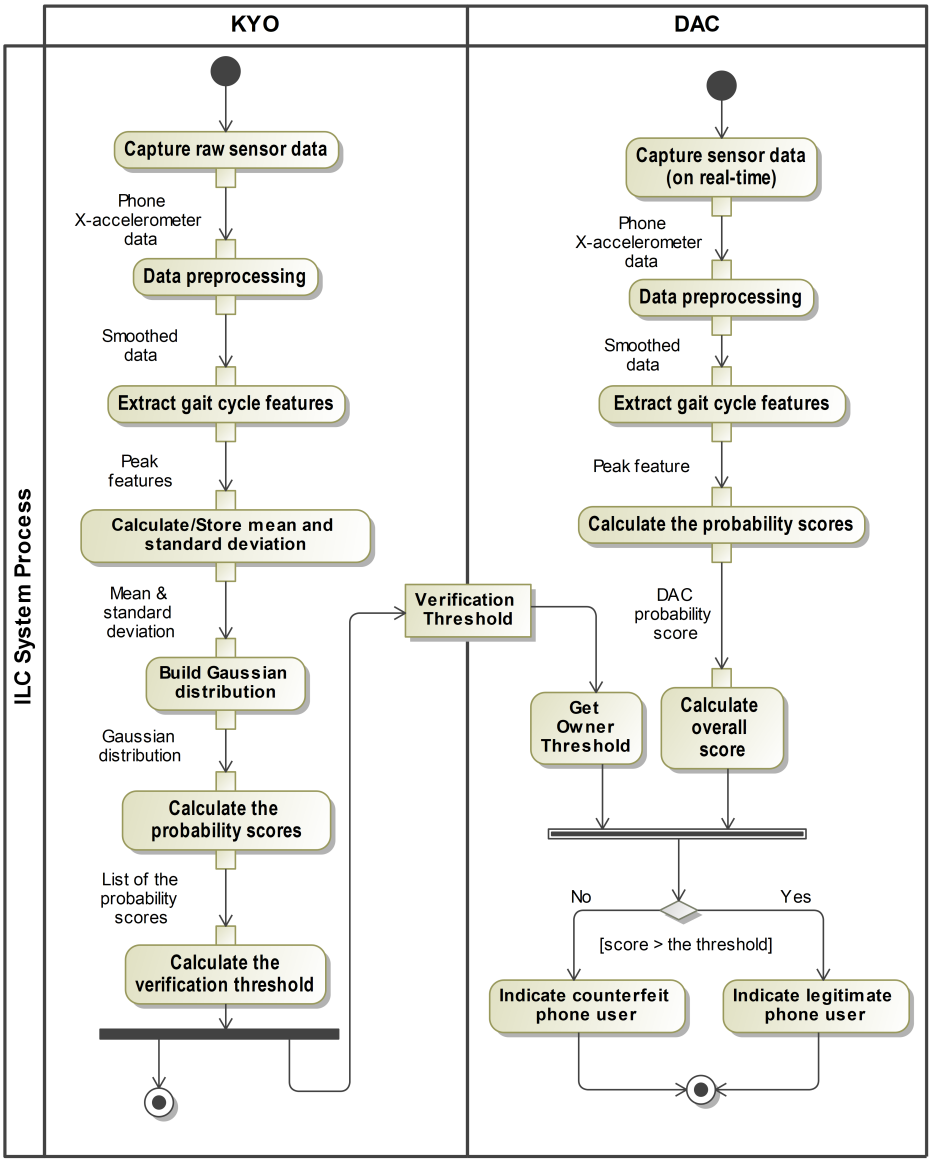


Figure 1. Overview of the (ILC) system process: the KYO lane presents the gait registration, and the DAC lane presents the gait detection.

The diagram depicted in Figure 2 takes the perspective of the right foot for the gait cycle, illustrating a percentage scale from 0% to 100%. This scale represents the full gait cycle, starting from the initial contact of one foot to the next initial contact of the same foot. This entire cycle is also known as a stride. The stance phase, highlighted in orange, constitutes approximately 60% of the gait cycle, while the swing phase, indicated in blue, comprises the remaining 40%. Each full cycle, or stride, consists of two steps—one with each foot. The stride is delineated by the blue line double-ended arrow, which marks the span from one initial foot contact to the next. The green bar in Figure 2 indicates the full extent of the gait cycle of the right foot.

In addition, Figure 3 represents the gait cycle percentages for both the stance phase and the swing phase for user A, with each phase containing several subphases. The plot shows the magnitude of X, Y, and Z acceleration data (m/s^2) and the percentage of a single gait cycle, where the stance phase comprises 60% and the swing phase comprises 40%. The graph illustrates that the magnitude of acceleration is higher at the beginning and end of the gait cycle, corresponding to heel strike events. During the stance phase, multiple peaks

indicate subphases such as loading response and mid stance, while the swing phase shows a relatively smoother curve, indicating the movement of the foot on the fly.

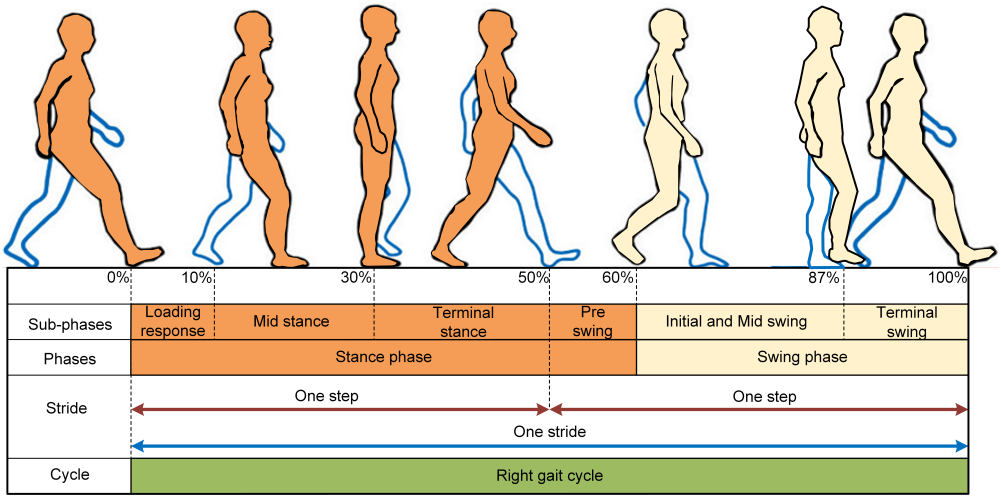


Figure 2. Phases and subphases of the gait cycle, showing the stance and swing phases of walking with their corresponding percentages of completion, along with the transitions between steps and stride for one complete right gait cycle.

While the essential elements of the gait cycle, such as the stance and swing phases, are consistent for each user, unique variations in walking lead to distinct gait patterns for each individual.

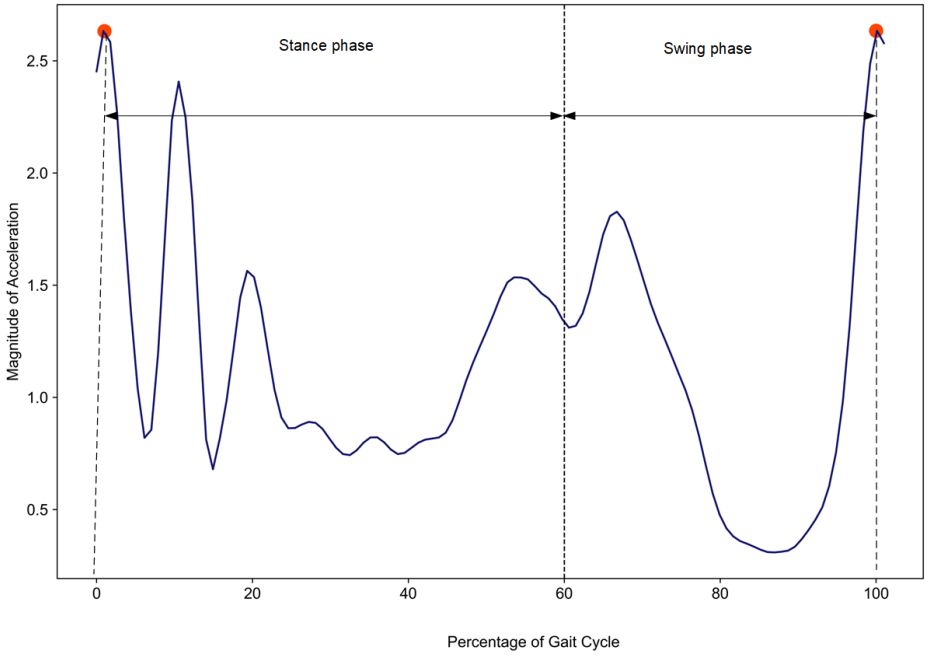


Figure 3. Visualizing a single gait cycle's percentage for user A. The percentage represents the completion of one gait cycle from 0% (heel strike) to 100% (next heel strike). The magnitude represents the resultant acceleration measured by the accelerometer.

2. The probability density function (PDF) is a statistical function that defines the probability distribution of a random variable. It is used to measure the similarity score between the registered data and the real-time outcome data. This score indicates the likelihood of an event occurring within a specific range of values [25].

3. Evaluation metrics refer to quantitative measures used to evaluate the performance and effectiveness of models, algorithms, or analyses. These metrics provide insights to evaluate and compare the effectiveness, accuracy and predictive capabilities of various statistical models or algorithms as shown in Table 2.

Accuracy is a measure of the proportion of accurately classified predictions for the phone owner and counterfeits in relation to the overall number of input samples [32]:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

- True Positive (TP): Refers to situations in which both the actual and expected classes are positive. This occurs when the system correctly identifies the legitimate phone owner.
- False Positive (FP): Occurs when the actual class is negative, but the anticipated class is positive. Essentially, the system incorrectly identifies someone who is not the legitimate owner, labeling them as legitimate.
- False Negative (FN): Refers to cases where the actual class is positive but the predicted class is negative. In this scenario, the system incorrectly identifies the legitimate phone owner as counterfeit.
- True Negative (TN): Occurs when both the actual and predicted classes are negative, indicating that the system correctly identifies a counterfeit user as not the legitimate owner.

Table 2. Classification of verification decisions for legitimate users and counterfeit attempts.

Decision/Status	Legitimate	Counterfeit
Legitimate	TP	FP
Counterfeit	FN	TN

3.3. System Methodology

This section outlines the approach utilized to achieve the objectives of the ILC goals mentioned in Section 3.1.2 above. We will describe the steps involved in implementing the ILC system, focusing on the actions that include the KYO and DAC components as illustrated in Figure 1.

3.3.1. Know Your Owner (KYO)

This component involves a series of sequential actions aimed at establishing and registering the owner's gait cycle pattern on the smartphone. We will explain how the verification threshold is set to verify the legitimate phone holder. This component involves the following actions:

Capture raw sensor data: The process begins by capturing the gait measurements of the owner from the x-axis accelerometer sensor data.

Data preprocessing: The moving average filter [33], which we denote as the moving average process, is used. This filtering methodology operates by convolving the input signal with a rectangular window function, thereby effecting a smoothing operation on the signal samples. The convolution integral finds the mathematical mean of the signal values within a certain window span. It does this by reducing the high-frequency parts of the signal while keeping the low-frequency parts and step transitions. This approach mitigates the destructive effects of additive noise and enhances the perceptibility of the underlying walking signal. This method calculates the average of data samples within a specified window size of three that moves across the samples.

Extract gait cycle feature: While various features can be extracted from the gait cycle, our focus is on a single feature, local maximum peak samples. These peak samples are calculated using the following algorithm (Algorithm 1):

Algorithm 1 Gait cycle feature extraction algorithm.

Input: $X = \{x_1, x_2, \ldots, x_N\}$ ▷ `Raw Accel-X data, (N: data size)`
Output: Gait cycle feature—list of the local maximum peak samples
Begin:
$X' \leftarrow \text{movingAverageProcess}(X)$ ▷ `Smoothing data with window size(s), store it in` X'
$\mu \leftarrow \frac{1}{N}\sum_{i=1}^{N} X'[i]$ ▷ `Calculate the mean of` X'`, store it in` μ
$\sigma \leftarrow \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X'[i]-\mu)^2}$ ▷ `Calculate the standard deviation of` X'`, store it in` σ
$P \leftarrow []$ ▷ P `is a list to store the peak values`
$maxP \leftarrow (\mu + k\sigma)$ ▷ `maxP: local maximum peak threshold, k is a constant equal to 2`
for $i = 1$ to N **do** ▷ `Check the smoothed data`
  if $((X'[i] > maxP) \wedge (X'[i] > X'[i-1]) \wedge (X'[i] > X'[i+1]))$ **then**
    Append $X'[i]$ to P ▷ `Detect Peaks`
  end if
end for
Return: P
End

Calculate the mean and standard deviation: Afterward, the mean and standard deviation as shown in Equations (2) and (3), respectively, are computed from the phone owner's peak samples, which represent the usual walking pattern. The mean μ_o is calculated as:

$$\mu_o = \frac{1}{N}\sum_{i=1}^{N} P_o(i) \tag{2}$$

where o represents ownership (specific to the phone owner's device); $P_o(i)$ denotes the peak acceleration values in each sample i, $\{p_1, p_2, \ldots, p_N\}$; and N is the total number of used samples.

The standard deviation σ_o is calculated as:

$$\sigma_o = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_o(i) - \mu_o)^2} \tag{3}$$

where μ_o is the mean calculated as per Equation (2).

Store mean and standard deviation: The computed mean and standard deviation are then stored on the phone owner's device for future identity verification purposes.

Build a Gaussian distribution: We model the legitimate phone owner's walking pattern using a Gaussian distribution, visually representing the PDF [34]. This density quantifies how probable it is to observe each peak X-acceleration value, providing a clearer picture of data distribution around the mean. The distribution illustrates peak X-acceleration values during gait cycles as depicted in Figure 4. This normal distribution, characterized by the calculated mean (μ) and standard deviation (σ), provides a statistical summary of the owner's typical walking pattern, which is essential for establishing a personalized baseline for future verification.

As shown in Figure 4, the majority of acceleration measurements cluster around a mean of 0.2 and standard deviation of 0.4, indicating the typical behavior during the owner's gait cycle. The distribution is centered around this mean, with most data points concentrated within one standard deviation. This range, from $\mu - \sigma$ to $\mu + \sigma$, encapsulates approximately 68% of the data, reflecting the common variability in the owner's walking pattern.

Central peak (mean, μ): The peak of the distribution at μ signifies the most frequent acceleration value, reflecting the average peak acceleration during the walking cycle.

Standard deviation (σ): The marks $\mu + \sigma$ and $\mu - \sigma$ indicate where the majority of peak acceleration values fall, demonstrating the typical variability in the gait cycle. This helps in understanding the breadth of normal movement variation, critical for identifying deviations in new data.

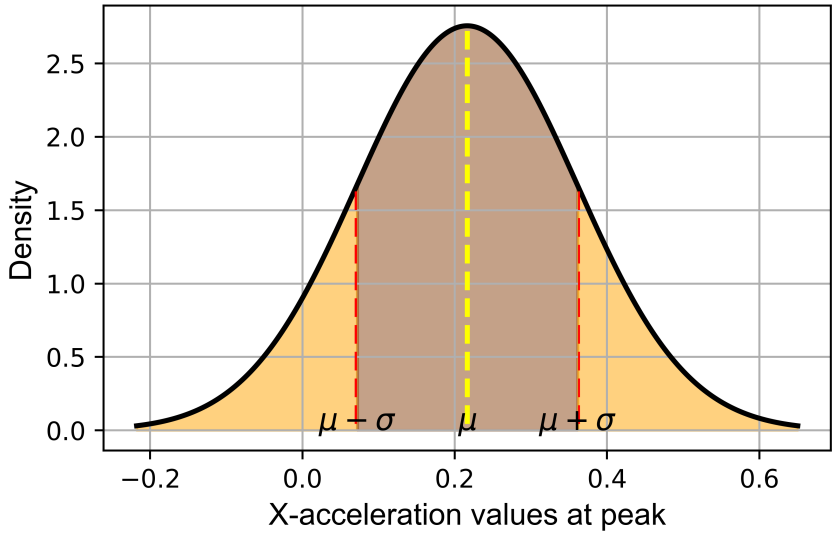


Figure 4. PDF for peak X-acceleration values of the phone owner, delineated with the mean (μ) and the range of one standard deviation ($\mu \pm \sigma$). This visualization highlights the central tendency and variability of the legitimate walking pattern, serving as a reference for assessing new observations.

Using a Gaussian distribution based on the phone owner's walking data, deviations in the mean and standard deviation can be effectively identified when someone else uses the phone. This approach ensures that significant variations from the established pattern of peak x-acceleration values, which are expected to cluster around a mean of 0.2, indicate potential unauthorized or counterfeit usage. Thus, this statistical model serves as a critical measure for securing personal devices against unauthorized access. Additionally, data points that significantly deviate from the curve—especially those far beyond the expected standard deviations—are often flagged as potential unauthorized actions.

Calculate probability scores: The probability score for each peak sample is calculated using the PDF as shown in Equation (4), considering the owner's mean and standard deviation. These probability scores are then used to calculate the verification threshold:

$$PD(p_i) = \frac{1}{\sigma_o \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{p_i - \mu_o}{\sigma_o}\right)^2} \tag{4}$$

where $PD(p_i)$ represents the probability score for each phone owner's peak sample; p_i represents each peak sample $\{p_1, p_2, \ldots, p_N\}$ derived from the owner's x-accelerometer samples; μ_o and σ_o indicate the calculated owner's mean and standard deviation; and N is the total number of peak samples used for thresholding.

Calculate the verification threshold: We now proceed to define the threshold score that will ensure an accurate identification of the owner. The threshold score is set for each peak sample as shown in Equation (5) after establishing probability scores as demonstrated in Equation (4). These probability scores, which reflect the typical gait pattern of the phone owner, are averaged to determine the verification threshold as illustrated by the following equation:

$$\text{Verification Threshold} = \frac{1}{N} \sum_{i=1}^{N} PD(p_i) \tag{5}$$

The steps used to calculate the verification threshold are outlined in the following algorithm (Algorithm 2): After the verification threshold is defined and stored on the owner's device, it becomes a benchmark for comparing new motion data. Thus, any new data collected, such as when the phone is picked up, are immediately analyzed against the established the verification threshold.

Algorithm 2 Verification threshold calculation algorithm.

Input: $P = \{p_1, p_2, \ldots, p_n\}$ ▷ P is a set of owner's peak values retrieved from Algorithm 1, (n:data size)
Output: verification Threshold
Begin:
$\mu_o \leftarrow \frac{1}{n}\sum_{i=1}^{n} P[i]$ ▷ Calculate the mean of P, store it in μ_o
$\sigma_o \leftarrow \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P'[i]-\mu)^2}$ ▷ Calculate the standard deviation of P, store it in σ_o
$PDList \leftarrow []$ ▷ $PDList$ is a list to hold the probability distribution scores
for $i = 1$ to n **do** ▷ n: number of peak samples to calculate the verification threshold
    $p_i \leftarrow \text{probabilityScore}(P[i], \mu_o, \sigma_o)$ ▷ Calculate probability using PDF
    Append P_i to $PDList$ ▷ Add calculated probability distribution score to the list
end for
Store μ_o, σ_o ▷ Store mean and standard deviation on the owner's device
$Threshold \leftarrow \frac{1}{n}\sum_{i=1}^{n} PDList[i]$ ▷ Calculate verification threshold using the mean of the probability scores
Return: *verification Threshold*
End

3.3.2. Detection and Control (DAC)

Once the phone is held, the DAC is activated for verification, indicating that the holder may be either the legitimate owner or an unauthorized phone holder. This involves the following process:

Capture sensor data (real-time): This action involves the real-time monitoring and analysis of the current phone holder's gait pattern. Similarly to the KYO process, the DAC component begins with capturing accelerometer data along the x-axis from the current phone holder.

Data preprocessing: This action involves preprocessing the X-axis accelerometer data using the same process as in KYO. By applying the moving average process, smoother samples are obtained.

Extract gait cycle feature: Peak sample features are extracted from X-accelerometer sensor samples as shown in Algorithm 1.

Calculate probability scores: Subsequently, the PDF is established for each peak sample taking into account the owner's mean and standard deviation. The probability score for each peak of the observed gait pattern is calculated using the following formula:

$$PD(p_i) = \frac{1}{\sigma_o\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{p_i-\mu_o}{\sigma_o}\right)^2} \tag{6}$$

where $PD(p_i)$ represents the probability density score for each peak sample derived in real-time from the X-accelerometer data of the phone holder, which can be either from the phone owner or a counterfeit holder; p_i represents each peak sample $\{p_1, p_2, \ldots, p_N\}$ of the phone holder's x-accelerometer samples; μ_o and σ_o indicate the calculated owner's mean and standard deviation; and N represents the total number of peak samples used for identity verification.

Calculate the overall score: The overall probability score is calculated first by determining the total number of peak samples that have close probability values P_{close}, and then dividing this P_{close} by the total number of peak samples P_{total} used for identity verification.

This process effectively computes the proportion (or frequency) of each specific probability as shown in Equation (7). The DAC probability score is then calculated using the following equation:

$$\text{DAC probability score} = \frac{P_{\text{close}}}{P_{\text{total}}} \tag{7}$$

where P_{close} represents the number of peak samples with close probability values, and P_{total} indicates the total number of peak samples considered for identity verification.

This probability score is used to assess the legitimacy of the observed pattern and determine how closely the current peak sample matches the owner's peak sample. This

comparison is made by evaluating this score against a predefined threshold during the verification process.

Verification Threshold: The verification threshold, established in the KYO process, is a benchmark score derived from the owner's gait data. This threshold is stored on the device and used in the DAC component to verify the legitimacy of the phone user.

Get Owner Threshold: Retrieve the stored verification threshold from the owner's device.

The final step in the ILC system involves making a decision based on the comparison between the DAC probability score and the predefined verification threshold. This crucial action determines the legitimacy of the phone holder's walking pattern.

- Indicate Legitimate Phone Owner: If the DAC probability score exceeds the verification threshold, the system concludes that the phone is being used by the legitimate owner.
- Indicate Counterfeit User: If the DAC probability score falls below the verification threshold, the system identifies the phone user as a potential counterfeit holder.

This decision-making process significantly enhances security by detecting unauthorized use, ensuring that deviations from the expected gait patterns are accurately flagged. This robust verification method is pivotal in safeguarding the device against unauthorized access. More details will be provided in the results in Section 5.2.3.

4. System Analysis and Validation

This section explores methodologies for extracting and analyzing key gait cycle features from accelerometer data. We will discuss the process of extracting local maximum peak features from the X-axis as outlined in Algorithm 1, followed by a detailed calculation of three specific gait cycle features across three accelerometer axes: the number of strides, average stride time, and total time. This method is validated through detailed algorithmic implementation, revealing unique distinguishing values for each user, ensuring accurate biometric identification.

4.1. Feature Extraction

Feature extraction is a crucial step in processing accelerometer data to identify unique walking patterns. The method used focuses mainly on the X-axis, which has proven to be more informative than analyzing features along both the X and Y axes separately. In addition, this approach surpasses the effectiveness of analyzing other sensor data, such as gyroscope readings, in three different directions. Furthermore, to identify these unique patterns within accelerometer sensor data, we have implemented a peak detection algorithm, detailed in Algorithm 1, which focuses on data from the X-direction. This algorithm is specifically designed to identify local maxima—peak points based on a predetermined threshold. The presence of these peak points, indicative of significant changes in acceleration, is particularly effective in distinguishing among user patterns. In the preprocessing, the raw X-axis accelerometer data are initially smoothed using a moving average process to eliminate noise and ensure consistent sampling intervals.

One key element of feature extraction is the gait cycle analysis based on peak data detection. This process involves identifying significant points in a signal that correspond to specific movements or events during a user's gait cycle, utilizing accelerometer data from the X-axis. This method helps to determine a unique representation of the user's walking pattern. Figure 5 illustrates the extraction of distinctive patterns of the gait cycle from accelerometer data. Smoothed peak points represented in the X-axis acceleration serve as key features, while the magnitude of acceleration is presented on the Y-axis against the peak points index on the X-axis, revealing each user's unique gait characteristics.

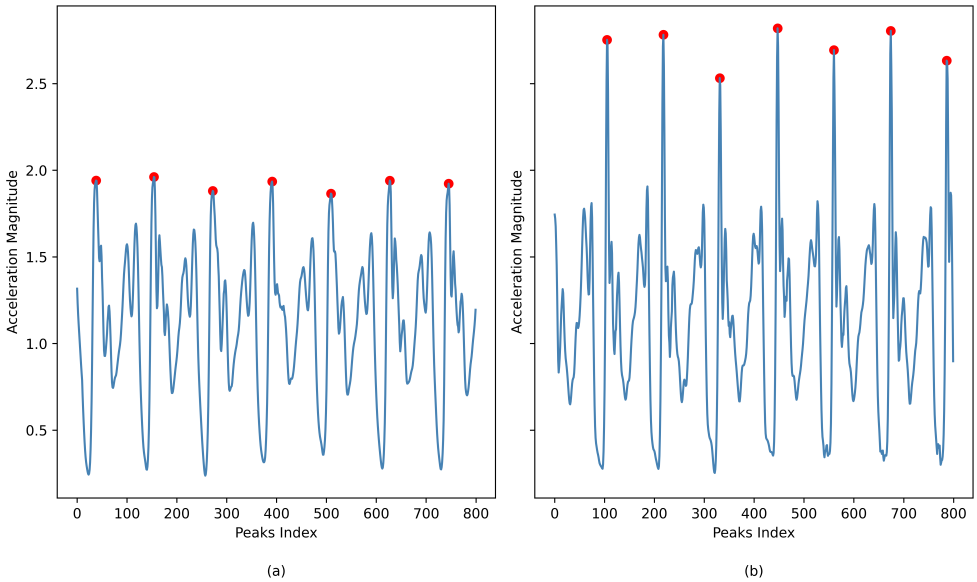


Figure 5. (**a**) Gait cycle pattern for user A. (**b**) Gait cycle pattern for user B.

The comparison between two arbitrary users, as depicted in Figure 5, effectively showcases the algorithm's proficiency in distinguishing unique walking patterns through peak detection (the red dots) in the accelerometer data. This comparative analysis is crucial for evaluating the algorithm's performance, particularly in applications necessitating individual gait characterization, such as biometric identification systems that verify the identity of authorized phone users. The gait cycle pattern of user A, shown in Figure 5a, presents consistent peak magnitudes at regular intervals, indicative of a stable and uniform walking pattern. In contrast, the gait cycle pattern of user B, depicted in Figure 5b, exhibits greater variability in both the height and distribution of the peaks. This variability could reflect a more complex gait, potential irregularities, or differences in walking speed. The distinctions observed between the walking styles of the two users not only illustrate their uniqueness but also validate the algorithm's capability to accurately depict and capture these differences.

4.2. Gait Cycle Features Extraction

In this section, we outline the techniques used to analyze three key gait cycle features from accelerometer sensor data for each user. Specifically, we extract three features: the number of strides, average stride time, and total time. These features are derived from the acceleration magnitude across the three dimensions (X, Y, and Z). The acceleration magnitude at a timestamp t is calculated by taking the square root of the sum of the squares of the individual components of the acceleration vector (X, Y, and Z axes), yielding a single value that represents the overall acceleration magnitude. In the following, we describe the process used to calculate these features for each user, outlining the detailed steps and methodologies involved. Additionally, as illustrated in Algorithm 3, we document all the steps involved in calculating the aforementioned gait cycle features.

1. Acceleration Magnitude Calculation: Utilizing the accelerometer sensor, measure the acceleration magnitude on the X, Y, and Z axes.
2. Data Smoothing: Apply a moving average technique to the magnitude variable, using a window size of 5, to reduce noise and ensure data consistency.
3. Mean and Standard Deviation Calculations: Compute the mean of the acceleration data. Then, compute the first and second standard deviations to establish thresholds for identifying significant local maximum peaks.

4. Local Maximum Peaks Detection: Identify the local maximum peaks by applying a threshold of mean + 2 standard deviations, which assumes that approximately 95% of the data falls within this range under a normal distribution.

Algorithm 3 Gait cycle features calculation algorithm.

Input: ▷ Raw Accel-X, Accel-Y, Accel-Z data
$X = \{x_1, x_2, \ldots, x_m\}$, ▷ m = data size
$Y = \{y_1, y_2, \ldots, y_m\}$,
$Z = \{z_1, z_2, \ldots, z_m\}$
Output: Number of strides, average stride time, total time.
Begin:
magnitudeList $\leftarrow$ [] ▷ Initialize magnitude list
for timestamp $t = 1$ to m **do**
    $X_t \leftarrow$ value of acceleration in the X-axis at timestamp t
    $Y_t \leftarrow$ value of acceleration in the Y-axis at timestamp t
    $Z_t \leftarrow$ value of acceleration in the Z-axis at timestamp t
    $X'_t, Y'_t, Z'_t \leftarrow movingAverage(X_t, Y_t, Z_t, s)$ ▷ Smooth data with consistent window size(s)
    $mag_t \leftarrow \sqrt{X_t'^2 + Y_t'^2 + Z_t'^2}$ ▷ Calculate the magnitude
    Append mag_t to magnitudeList
end for
$\mu \leftarrow$ mean(magnitudeList) ▷ Calculate mean of magnitudeList
$\sigma \leftarrow$ std dev(magnitudeList) ▷ Calculate standard deviation of magnitudeList
peakList $\leftarrow$ [] ▷ Initialize list for local maximum peaks
$maxP \leftarrow \mu + k\sigma$ ▷ maxP: local maximum peak threshold, k is a constant equals to 2.
for $i = 1$ **to** length(magnitudeList) **do**
    if $\text{mag}_t[i] > maxP \wedge \text{mag}_t[i] > \text{mag}_t[i-1] \wedge \text{mag}_t[i] > \text{mag}_t[i+1]$ **then**
        Append $\text{mag}_t[i]$ to peakList ▷ Detect peaks
    end if
end for
strideNumber $\leftarrow$ length(peakList) $-$ 1 ▷ Number of strides
strideTimeList $\leftarrow$ [] ▷ Initialize list for times between peaks
for $i = 1$ to length(peakList) $-$ 1 **do**
    strideTime $\leftarrow$ peakList$[i+1]$ $-$ peakList$[i]$
    Append strideTime to strideTimeList
end for
averageStrideTime $\leftarrow$ mean(strideTimeList) ▷ Average stride time
totalTime $\leftarrow$ (end_time $-$ start_time) ▷ Total experiment time
$X_{\text{peaks}} \leftarrow$ DetectPeaks(X) ▷ X-peaks as defined in Peak Algorithm 1
Return: strideNumber, averageStrideTime, totalTime,
End

The calculated gait cycle features are described as follows:

- Number of Strides: Accurately determine the total number of strides by subtracting the initial peak from the count of local maximum peaks. This correction accounts for the initial peak, which may not be associated with a preceding stride and is therefore excluded from the stride count.
- Average Stride Time: Calculate the time duration between two consecutive local maximum peaks, then average these durations to determine the average stride time.
- Total Experiment Time: Determine the exact data points (timestamps) that mark the start and end of the walking activity.

5. Experiment and Result

This section provides detailed insights into the setup of our study and its results. It outlines the critical factors, tools, and steps involved in collecting, preprocessing, and analyzing users' motion data. Following this, we present the results of the proposed system along with its analysis and then evaluate the performance of the ILC system.

5.1. Experiment

Data Acquisition and Processing: This outlines the procedures for capturing walking data and details the steps involved in processing the collected data. The experiment is designed to meet the specific requirements of our research.

1. Data Acquisition Team: A diverse group of 10 people (users), carefully selected to include a (3:7) distribution ratio across gender's representation and age groups, facilitated the data collection process. This team is comprised of male and female participants; their age demographics are strategically segmented into two distinct categories, the first encompassing those within the 22–25 year age bracket, and the second encompassing individuals ranging from 26 to 30 years of age. This deliberate selection criteria ensured a well-rounded perspective, capturing the variations that may arise from the interplay of gender and generational differences within the targeted sample.
   Users were instructed to walk in a straight line for 100 m on the Western University Football Field, with data acquisition occurring at a frequency of 100 Hz. This process yielded a dataset comprising a total of 12,000 samples obtained from the accelerometer sensor on the X-axis, generated in CSV format.
2. Data Acquisition Tools: using the SensorLog app, version 5.3.1., installed on an iPhone 11. The app was configured as shown in Table 3 to mirror the sensor settings described in [17,35].
3. Data preprocessing: We performed a series of steps to refine the collected data. Initially, we removed the initial records gathered during preparation and the data captured when the users' tasks ended and the recording ceased. This step aimed to eliminate potential noise in the data, ensuring that the retained data accurately reflect the intended measurements. Additionally, we employed the moving average process, which involves calculating the average of a small group of neighboring data points using a window size of five. This approach resulted in smoother data, enhancing the ability to identify meaningful user patterns during the analysis.
4. Experiment setup: The data for this experiment consist of 12,000 samples. Specifically, 7500 samples represent data from the phone owner, while the remaining 4500 samples are from unauthorized phone holders. The phone owner's samples are divided into three distinct segments: 3000 samples are allocated for constructing the phone owner's normal distribution, 2000 samples are used to determine the threshold value, and the remaining 2500 samples, along with the 4500 samples from unauthorized phone holders, are used for real-time verification.

To generalize the ILC system, we repeated the process ten times, each iteration involving a different user acting as the phone owner. We adopted this approach to ensure the system's effectiveness across various users. This diverse sampling strategy significantly enhances the system's robustness and adaptability, ensuring high performance across a broad range of user behaviors anywhere it may be deployed, with unlimited numbers of users in different regions.

Table 3. Data collection overview.

Experimental Setup Item	Settings
Number of users	10
Activity	Walking with phones in pocket
Session length	Approx: 90 s–less than 2 min
Age range	22 years to 30 years
Sex	3 male & 7 female
Time line	2 weeks
Device	iPhone 11
Data sampling	100 Hz

5.2. Results

This section presents the significant findings that support our methodology to achieve the ILC objectives. First, we compute three gait cycle features that capture distinguishing patterns specific to each individual; second, we calculate the average accuracy of each legitimate phone owner by comparing it against unauthorized attempts, assuming that

each user is the correct phone owner. Additionally, we demonstrate how the ILC system's real-time verification process confirms the legitimacy of the phone holder. Finally, we explore the variability and distributional properties of the data.

5.2.1. Gait Cycle Features Analysis

We perform an extensive gait cycle feature analysis for user identification by examining different individuals' walking patterns. Specifically, we concentrate on three crucial gait cycle features extracted for each phone owner: the number of strides, the average stride time, and the total time. As detailed in Table 4, the data demonstrate variation in these features across individuals, emphasizing the uniqueness of their gait patterns. The uniqueness of these gait characteristics is instrumental in improving the integrated ILC system's identification capabilities, thereby improving user differentiation accuracy.

Table 4. Gait cycle features analysis for each user.

User	Number of Strides	Average Stride Time (s)	Total Time (s)
1	92	0.984	89.389
2	63	1.097	67.936
3	52	1.155	59.301
4	96	0.658	62.037
5	81	0.741	58.997
6	66	1.052	64.243
7	91	0.726	63.943
8	57	1.043	58.373
9	104	0.579	60.108
10	87	0.796	69.211

Table 4 presents varied gait cycle characteristics among ten distinct users, providing insights into their individual walking patterns. Significantly, users 4 and 9 have the most strides, with 96 and 104, respectively, indicating a greater number of steps taken during the walking. In contrast, user 3 has the lowest stride count at 52, which may imply a slower gait with each stride covering a longer distance. Although users 4 and 9 have the highest stride number, they do not have the longest total walking times, suggesting that their walking during the experiment was marked by a brisk pace.

Additionally, these users exhibit the shortest average stride times, at 0.658 and 0.579 s, respectively, representing a quicker pace. In contrast, user 3 has the longest average stride time, at 1.155 s, potentially reflecting a slower pace. Interestingly, user 1 has the longest total time, which, despite a moderate stride count and average stride time, might indicate a longer duration of data collection rather than an increased pace of walking.

The analysis reveals no definitive correlation between the number of strides and total time, indicating varying walking durations during the experiment, paces, and levels of energy across the users. These variations and distinctions between users underscore the complexity of using gait analysis for individual identification, as each user exhibits a unique combination of gait features.

We demonstrate that the peak extracted from the acceleration samples is unique because they are based on substantial differences in the gait cycle data , which exhibit significant variance between users. This uniqueness of the peaks is founded on the premise that each peak arises from distinct features—number of strides, average strides time, and total walking time—each contributing to the overall variability observed in user gait patterns. These features inherently contain peak points, as they represent critical moments of change or intensity in the gait cycle, making them vary from one user to another. Peak samples serve as a reliable metric for user identification because they highlight differences in the number of strides, average strides time, and the total walking time, reflecting the information in Table 4.

Therefore, the peak samples calculation is based on these foundational characteristics, ensuring that the ILC system leverages the most informative and distinctive aspects of gait data. By focusing on these variances, the system can effectively distinguish and identify users based on their individual walking patterns.

5.2.2. Overall Accuracy Evaluation

Table 5 displays the individual accuracy rates for each legitimate phone owner compared to attempts made by counterfeit users. Additionally, it presents the overall average accuracy rate, indicating the general performance of the identification system across all evaluated phone owners. We utilize peak points extracted from smartphone accelerometer data along the X-axis as the sole feature. This approach yields the best results compared to data from the Y and Z axes. This is because as users walk, motion predominantly occurs in the X-direction, which contains more informative data because of the natural walking movements. In contrast, movements along the Z and Y axes are almost constant and do not change significantly, implying that including these axes would introduce redundancy into the feature set. Additionally, the X-axis accelerometer data outperform those from all gyroscope directions (X, Y, and Z) after comparative analysis. This finding is drawn after testing each one individually, which confirms that the X-axis accelerometer is the most effective at capturing relevant motion information. After conducting extensive experiments with a diverse group of 10 users, the overall findings reveal that the ILC system identifies unauthorized phone holders with an accuracy rate of 92.18%.

Table 5. Average accuracy across all users as phone owners.

Phone Owner	Accuracy (%)
1	94.95
2	87.13
3	87.02
4	86.70
5	84.71
6	94.97
7	98.35
8	98.30
9	92.63
10	97.06
Average accuracy	92.18

5.2.3. Real-Time User Verification

The process is initiated when the phone is picked up, and the device immediately records the peak value from the X-accelerometer. The real-time verification process of the ILC system consists of two essential steps to assess the legitimacy of the user. A detailed description of the verification process is as follows:

1. Calculate DAC Probability Score: The phone calculates the DAC probability score, quantifying the likelihood that the current user's gait matches the known pattern. This score is a numeric representation of similarity; higher scores indicate a closer match to the expected gait pattern, while lower scores suggest a significant deviation from the original owner's gait pattern.
2. Decision-Making: This step involves comparing the calculated DAC probability score against a set verification threshold, determined based on the mean and standard deviations from the phone owner's typical data. The system evaluates whether the DAC probability score exceeds the verification threshold derived from the owner's mean and variance. If it does, it implies that the new peak data closely match the owner's typical peak data as defined by the KYO probability distribution. This similarity indicates that the device is likely being used by the legitimate owner. Conversely, if

the DAC probability score falls below the verification threshold, it suggests that the current holder's peak data significantly deviate from the owner's typical walking patterns, suggesting that the phone is in the hands of an unauthorized user.

5.2.4. Analyzing Data Variability

We examine the variability of the data, particularly the distinctiveness of the study users. Histogram distribution plots and boxplots are used to gain insight into the distinguishing features of individuals. The experiment compares the characteristics of the phone owner (user A) and the nine counterfeiters. As shown in Figure 6, when creating two histogram plots, one for the phone owner and the other for counterfeiters, we ensure that the data distribution limits on both axes are the same for both histograms to make the comparison between them statistically accurate.

When comparing the two histograms, the phone owner and the counterfeiters, they seem to have a similar mean; however, they exhibit a different range, where the owner's range is a bit narrower, ranging from almost -0.37 to 0.98, while the counterfeiters overall show a wider range of almost -1.55 to 1.86. This variance indicates that every counterfeit might have a different mean, minimum, maximum, and standard deviation, which makes them viable features for user identification.

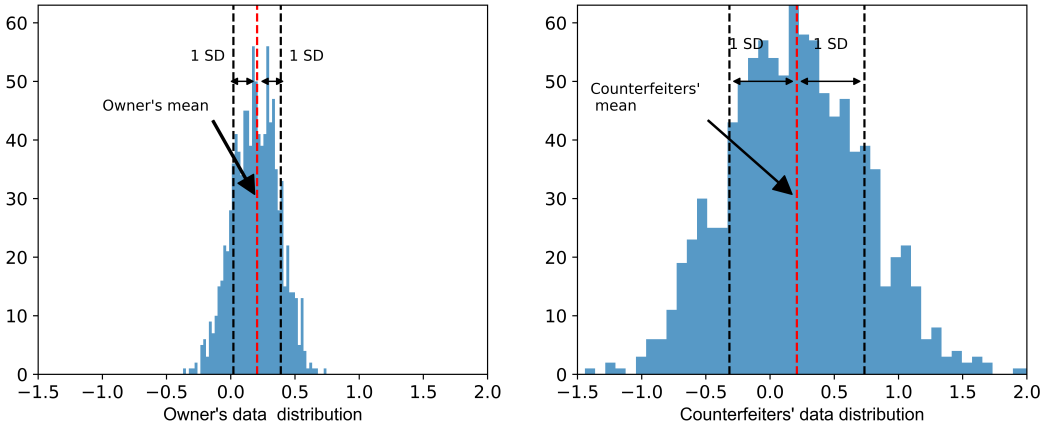


Figure 6. Probability distribution function comparing the phone's owner and the counterfeiters.

The boxplot in Figure 7 reveals that the owner has less data variability compared to counterfeiters. This is shown by the difference in the range of magnitude values for counterfeiters, which is wider than the owner. The wider interquartile range (IQR) indicates that counterfeiters have distinct minimum and maximum values, which contributes to the diversity of data. This is obvious from the boxplot, where the width of the (IQR) of the owner is clear compared to that of counterfeiters.

This variation between the phone owner and counterfeiters further emphasizes the potential use of these data for user verification. This highlights the importance of incorporating this application to improve identification reliability, as (IQR) improves user distinctiveness.

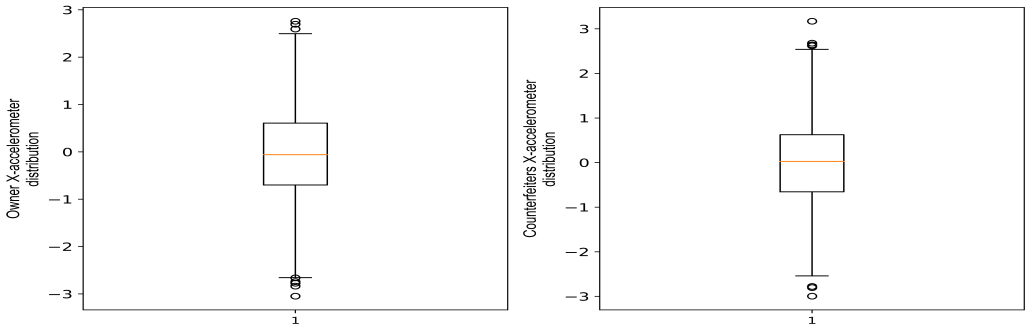


Figure 7. Data variability comparison between phone owner and counterfeiters.

In real-time user verification scenarios, the ILC system can compute the PDF for both the legitimate phone owner and counterfeit phone holders. The focus is on verifying the identity of the legitimate phone owner in comparison to counterfeit phone holders. This approach ensures the uniqueness of the gait patterns, thus enabling accurate recognition of the authorized phone owner as highlighted in Figures 6 and 7.

6. Discussion and Conclusion

6.1. Discussion

Table 6 presents a comprehensive performance evaluation of the ILC system compared to existing approaches. The ILC system achieves a high level of accuracy, identifying 92.18% counterfeit phone holders. In this way, the system outperforms statistical methods in [22] and [36] (see Table 6). Furthermore, the ILC has a low level of complexity and does not require training.

Table 6. Comparative performance evaluation: proposed method vs. other work.

Reference	Approach	Accuracy %	Complexity
Matei-Sorin et al. [22]	Statistical method	90.00	High
S. F. Auliya et al. [36]	Statistical method	88.00	High
ILC system	Statistical method	92.18	Low

The ILC system utilized a single peak of characteristics extracted from accelerometer data on the X-axis of smartphones and was able to recognize unauthorized phone holders with an accuracy of 92.18%. For the remaining 8%, the system either mistakenly authenticated incorrect phone holders or failed to recognize legitimate phone holders. This outcome demonstrates the overall effectiveness of the proposed system in detecting unauthorized attempts.

The ILC system is a distinctive approach to using accelerometer data in the X direction to enhance mobile security. Its efficiency is notable compared to the method of [22], which requires additional data from accelerometer and gyroscope sensors on multiple axes, leading to increased complexity and memory usage. The ILC system not only simplifies the process but also offers the potential for more streamlined and efficient mobile applications. The ILC system is less complex when compared with the system proposed in [22]. With a total of 1779 data peaks, ILC consumes less memory because it must store only three values on the owner's device: the mean, the standard deviation, and the threshold. The system in [22] will have 5625 features and characteristics.

Although the study by [36] achieved an accuracy of 88.00% in identifying user activity, it requires the analysis of numerous app usage metrics. This increases the data handling complexity and memory demands, which may prevent the system from operating efficiently. In contrast, the ILC system focuses on gait analysis using peak X-axis acceleration, simplifying data processing and reducing memory usage. This makes the ILC system more efficient and easier to implement in real-time applications.

6.2. Conclusions

In this paper, we have proposed a biometric authentication system called ILC that utilizes gait cycle analysis obtained using sensors on the smartphone to seamlessly identify the phone owner. This research highlights the importance of using data from mobile phone sensors to authenticate the legitimacy of the phone's holder. On the basis of the data examined, the research draws the following significant inferences:

1. The ILC system achieves an accuracy of 92.18% in recognizing unauthorized phone holders, illustrating the effectiveness of the proposed approach for user identification. Furthermore, the extracted gait cycle features expose variations among users, further accentuating the potential integration of this application to increase verification reliability.

2. A new solution for real-time behavior analysis, introduced as an identification method, seamlessly verifies the eligibility of legitimate phone holders.
3. The proposed method outperforms some other statistical methods with no need for training and low complexity.
4. This new approach focuses first on verifying the identity of the smartphone's owner. Once the phone verifies its rightful owner, it can be used to access a range of accounts and services seamlessly. This innovation eliminates the need to remember multiple passwords for various accounts, simplifying the verification process and enhancing user convenience.
5. Lightweight System: The memory usage of the ILC system is minimal, storing only three values on the owner's device: the mean, the standard deviation, and the verification threshold.

The ILC system's design goals and operational results provide a user-friendly solution that requires no direct involvement from the user, thus maintaining the seamlessness of the identification process. In this manner, it fulfills the first goal as represented in Section 3.1.2. The ability of the ILC system to provide real-time user verification aligns with the second goal, showcasing its capacity to process and validate individual walk patterns compared to stored data seamlessly and instantaneously. The ILC system achieves an accuracy rate of 92.18% in differentiating between legitimate phone owners and counterfeiters, which exceeds the initial high detection accuracy of 90%. This result not only demonstrates the system's efficacy in meeting the third goal but also highlights its potential to significantly enhance security measures in biometric-based identification access control systems. The success of the ILC system in achieving these goals highlights its innovative approach to security, blending efficiency, accuracy, and user convenience in a way that sets a new standard for biometric identification technologies.

To further strengthen the validity and generalizability of the system findings, additional statistical analyses, such as correlation analysis or hypothesis testing, can be considered as future directions. By employing these advanced statistical methods, researchers can better understand the relationships between different variables and assess the significance of the results. This would not only enhance the credibility of the study but also provide deeper insights into the system's performance, paving the way for more informed decision-making and potential refinements.

Future research directions could include long-term validation studies, integrating the ILC system with wearable technology, improving security features, conducting user experience studies, exploring privacy-preserving techniques, addressing scalability and deployment concerns, and investigating real-world applications to advance the field of biometric authentication based on gait analysis. These initiatives would aid in the system's refinement, assuring its robustness in a variety of situations, and broadening its application to a broader range of use cases, so cementing its place in the future of secure and convenient biometric authentication.

Author Contributions: Conceptualization, A.E. and A.O.; Methodology, A.E. and A.O.; Software, A.E.; Validation, A.E. and A.O.; Formal analysis, A.E. and A.O.; Investigation, A.E. and A.O.; Resources, A.E. and A.O.; Data curation, A.E.; Writing—original draft, A.E.; Writing—review and editing, A.E. and A.O.; Supervision, and A.O.; Project administration, A.O.; Funding acquisition, A.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article material; further inquiries can be directed to the corresponding author.

Acknowledgments: This work was supported by the Department of Electrical and Computer Engineering, Western University, Canada, and the Department of Computer Science, Faculty of Information Technology, Alasmarya Islamic University, Zliten, Libya. These supports are greatly appreciated.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hussain, H. Password Security: Best Practices and Management Strategies. 2022. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4136333 (accessed on 18 June 2024).
2. Sobers, R. 84 Must-Know Data Breach Statistics. 2023. Available online: https://www.varonis.com/blog/data-breach-statistics (accessed on 10 November 2023).
3. Okoli, K.; Bekeneva, Y. Balancing security and user experience in the evolving digital landscape. *E3S Web Conf.* **2024**, *471*, 04007. [CrossRef]
4. Okoli, K.; Joseph, I.; Chijioke, F.; Bekeneva, Y.; Chijioke, A.; Kodondo, I. Human-Machine Interaction in E-commerce: A Multi-Faceted Examination of CAPTCHA Effect. In Proceedings of the 2024 Conference of Young Researchers in Electrical and Electronic Engineering (ElCon), Saint Petersburg, Russia, 29–31 January 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 259–263.
5. Panchal, G.; Samanta, D.; Das, A.K.; Kumar, N.; Choo, K.K.R. Designing Secure and Efficient Biometric-Based Access Mechanism for Cloud Services. *IEEE Trans. Cloud Comput.* **2022**, *10*, 749–761. [CrossRef]
6. Salunke, S.V.; Ouda, A. Ensemble Learning to Enhance Continuous User Authentication for Real World Environments. In Proceedings of the 2023 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Istanbul, Turkiye, 4–7 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 102–108.
7. Liu, Y.; Zhou, T.; Yue, Z.; Liu, W.; Han, Y.; Li, Q.; Yang, X. Secure and Efficient Online Fingerprint Authentication Scheme Based On Cloud Computing. *IEEE Trans. Cloud Comput.* **2023**, *11*, 564–578. [CrossRef]
8. Galbally, J.; Marcel, S.; Fierrez, J. Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition. *IEEE Trans. Image Process.* **2014**, *23*, 710–724. [CrossRef]
9. Bansal, P.; Ouda, A. Continuous Authentication in the Digital Age: An Analysis of Reinforcement Learning and Behavioral Biometrics. *Computers* **2024**, *13*, 103. [CrossRef]
10. Huang, X.; Boulgouris, N.V. Unobtrusive Biometric Identification Based on Gait. In *Biometrics: Theory, Methods, and Applications*; Wiley: Hoboken, NJ, USA, 2010; pp. 539–558. [CrossRef]
11. Pataky, T.C.; Mu, T.; Bosch, K.; Rosenbaum, D.; Goulermas, J.Y. Gait recognition: Highly unique dynamic plantar pressure patterns among 104 individuals. *J. R. Soc. Interface* **2012**, *9*, 790–800. [CrossRef]
12. Maulana, N.; Istiqomah, F.; Priananda, C.W. Integration of Centralized Fingerprint Biometric Authentication to Prevent Room Access Violations Using RBAC. In Proceedings of the 2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA), Surabaya, Indonesia, 14–15 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 905–909.
13. Sharma, S.; Saini, A.; Chaudhury, S. A survey on biometric cryptosystems and their applications. *Comput. Secur.* **2023**, *134*, 103458. [CrossRef]
14. Bagwan, S.M.R.; Gupta, G.; Thigale, S.B. Robust Multi-Bio-Metric Authentication Framework in Face and Iris recognition. In Proceedings of the 2023 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, 3–5 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–10.
15. Priyanka, V.; Shyam, G.K. Enhancement of Accuracy for Iris Presentation Attack. In Proceedings of the 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, 1–2 September 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7.
16. Cao, Q.; Xu, F.; Li, H. User authentication by gait data from smartphone sensors using hybrid deep learning network. *Mathematics* **2022**, *10*, 2283. [CrossRef]
17. Wang, Y.; Zhang, X.; Hu, H. Continuous User Authentication on Multiple Smart Devices. *Information* **2023**, *14*, 274. [CrossRef]
18. Zeng, X.; Zhang, X.; Yang, S.; Shi, Z.; Chi, C. Gait-based implicit authentication using edge computing and deep learning for mobile devices. *Sensors* **2021**, *21*, 4592. [CrossRef]
19. Zou, Q.; Wang, Y.; Wang, Q.; Zhao, Y.; Li, Q. Deep learning-based gait recognition using smartphones in the wild. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 3197–3212. [CrossRef]
20. Chen, F.S.; Belman, A.K.; Phoha, V.V. Formalizing PQRST Complex in Accelerometer-based Gait Cycle for Authentication. *arXiv* **2022**, arXiv:2205.07108.
21. Kumari, S.; Singh, K.; Khan, T.; Ariffin, M.M.; Mohan, S.K.; Baleanu, D.; Ahmadian, A. A novel approach for continuous authentication of mobile users using Reduce Feature Elimination (RFE): A machine learning approach. *Mob. Netw. Appl.* **2023**, *28*, 767–781. [CrossRef]
22. Axente, M.S.; Dobre, C.; Ciobanu, R.I.; Purnichescu-Purtan, R. Gait recognition as an authentication method for mobile devices. *Sensors* **2020**, *20*, 4110. [CrossRef]
23. Delgado-Santos, P.; Tolosana, R.; Guest, R.; Vera-Rodriguez, R.; Deravi, F.; Morales, A. GaitPrivacyON: Privacy-preserving mobile gait biometrics using unsupervised learning. *Pattern Recognit. Lett.* **2022**, *161*, 30–37. [CrossRef]
24. Muro-De-La-Herran, A.; Garcia-Zapirain, B.; Mendez-Zorrilla, A. Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors* **2014**, *14*, 3362–3394. [CrossRef] [PubMed]
25. Nguyen-Trang, T.; Nguyen-Thoi, T.; Nguyen-Thi, K.N.; Vo-Van, T. Balance-driven automatic clustering for probability density functions using metaheuristic optimization. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 1063–1078. [CrossRef]
26. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]

27. Price, C.; Schmeltzpfenning, T.; Nester, C.J.; Brauner, T. Foot and footwear biomechanics and gait. In *Handbook of Footwear Design and Manufacture*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 79–103.
28. Zhao, H.; Xu, H.; Wang, Z.; Wang, L.; Qiu, S.; Peng, D.; Li, J.; Jiang, J. Analysis and evaluation of hemiplegic gait based on wearable sensor network. *Inf. Fusion* **2023**, *90*, 382–391. [CrossRef]
29. Núñez-Trull, A.; Álvarez-Medina, J.; Jaén-Carrillo, D.; Rubio-Peirotén, A.; Roche-Seruendo, L.E.; Gómez-Trullén, E.M. Influence of walking speed on gait spatiotemporal parameters and the functional rockers of the foot in healthy adults. *Med. Eng. Phys.* **2023**, 104002. [CrossRef]
30. Cho, H. Walking Speed Estimation and Gait Classification Using Plantar Pressure and On-Device Deep Learning. *IEEE Sens. J.* **2023**, *23*, 23336–23347. [CrossRef]
31. Di Gregorio, R.; Vocenas, L. Identification of gait-cycle phases for prosthesis control. *Biomimetics* **2021**, *6*, 22. [CrossRef] [PubMed]
32. Ruuska, S.; Hämäläinen, W.; Kajava, S.; Mughal, M.; Matilainen, P.; Mononen, J. Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behav. Process.* **2018**, *148*, 56–62. [CrossRef] [PubMed]
33. Han, D.; Bashar, S.K.; Lázaro, J.; Mohagheghian, F.; Peitzsch, A.; Nishita, N.; Ding, E.; Dickson, E.L.; DiMezza, D.; Scott, J.; et al. A real-time ppg peak detection method for accurate determination of heart rate during sinus rhythm and cardiac arrhythmia. *Biosensors* **2022**, *12*, 82. [CrossRef]
34. Chen, J. Normal Distribution: What It Is, Uses, and Formula. Investopedia. 2024. Available online: https://www.investopedia.com/terms/n/normaldistribution.asp (accessed on 13 March 2024).
35. Micucci, D.; Mobilio, M.; Napoletano, P. Unimib shar: A dataset for human activity recognition using acceleration data from smartphones. *Appl. Sci.* **2017**, *7*, 1101. [CrossRef]
36. Auliya, S.F.; Nugroho, L.E.; Setiawan, N.A. User identification using smartphone activity data. In Proceedings of the AIP Conference Proceedings, Yogyakarta, Indonesia, 7–8 September 2021; AIP Publishing: Melville, NY, USA, 2023; Volume 2654.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.