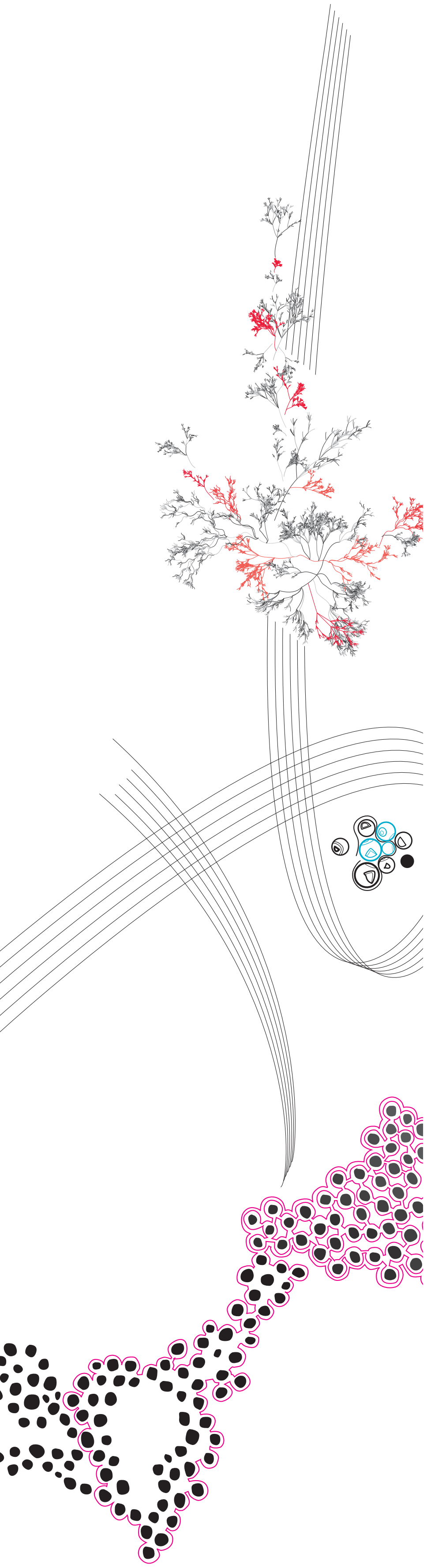
MSc Computer Science
Master's Thesis



Leveraging Earable Sensors for Lightweight Gait-Based User Recognition

Gergana Georgieva

Supervisor: Özlem Durmaz-Incel, Faiza Allah Bukhsh

February, 2025

Department of Computer Science
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Contents

Abstract

Wearable devices are evolving rapidly, with earables' popularity being on the rise as they gain new and sophisticated sensing capabilities. Their growing complexity, however, also poses heightened security risks. As these devices lack the interface to support traditional input-based authentication such as PIN or lock patterns, there is a call for new methods to provide reliable user verification. Gait-based behavioral biometrics, particularly in the context of leveraging IMU data, remain greatly underexplored for earables. This work investigates the feasibility of gait-based user recognition using IMU-equipped earable devices. The authors collect a new dataset consisting of 30 participants performing several different gait-related exercises at varying intensities. Traditional ML models (Random Forest, SVM, kNN, MLP) and a CNN-LSTM hybrid are benchmarked on authentication and identification tasks, in within-activity and across-activity scenarios, averaging at an EER below 2%. Feature selection and post-training quantization are shown to significantly reduce model size and inference cost without sacrificing accuracy. These findings confirm that gait-based user recognition using IMU-equipped earables is both feasible and practical, offering secure, unobtrusive verification on resource-constrained devices.

Keywords: Earables, Gait Sensing, Behavioral Biometrics, Inertial Sensing, Deep Learning, Edge Computing, Gait-based Recognition, Resource Optimization Techniques, Lightweight Models

Chapter 1

Introduction

Wireless earbuds have emerged as one of today's most widely adopted smart wearable devices, with the global Earphones and Headphones market projected to reach USD 193.23 billion by 2032 [1]. While they were initially designed for basic audio playback, modern earables now incorporate advanced sensors - such as inertial measurement units (IMUs) and optical sensors — that allow for an extensive range of functionalities, from real-time activity recognition [41] to health monitoring [19] and gesture elicitation [60].

The increasing complexity of earables required for advanced sensing not only places them at the forefront of ubiquitous computing, but also introduces heightened security and privacy risks. Earables often rely on continuous Bluetooth connections to stream audio and exchange data, creating potential entry points for eavesdropping and man-in-the-middle attacks (unauthorized real-time interception of a private communication) if communication channels are not adequately encrypted [95]. Malicious actors can exploit poorly secured links to intercept audio streams, sensor readings, or even device commands, posing a direct threat to user privacy and security. Furthermore, because earables frequently integrate with smartphones and cloud services, vulnerabilities in firmware or companion apps can be leveraged by malware to gain unauthorized access, remotely manipulate settings, or tamper with sensor data. Lastly, their portability also makes them susceptible to loss or theft.

Unlike other popular wearables such as smartphones or smartwatches that typically support input-based protection mechanisms like PINs or pattern locks, or incorporate biometric scanners, earables lack dedicated authentication interfaces, display screens, and direct user input controls. This absence of built-in security features makes them inherently more vulnerable to unauthorized access, device misuse, and data breaches. Paired with their compact size, limited on-device memory, and low-power processing capabilities, this further necessitates alternative authentication mechanisms that are robust, memory-efficient, and user-friendly.

To address some of these concerns, biometric authentication has gained traction as a secure and convenient alternative to traditional knowledge-based methods [101]. Among the various hard and soft biometric modalities, behavioral biometrics, particularly gait-based recognition, stand out due to their unobtrusive nature and high resistance to imitation attacks [74, 59], making it a method well-suited for continuous and implicit user verification.

Gait-based recognition can be captured through various sensing modalities, including vision-based methods, which analyze movement patterns from video footage, and floor sensor-based techniques, which detect footstep pressure and weight distribution [33]. While effective, these approaches often require specialized infrastructure, making them less practical for portable authentication. In contrast, the third type - the wearable sensor-based approach - leverages inertial measurement unit (IMU) sensors inherently embedded in consumer devices like smartwatches [6] and smartphones [40] for user recognition with promising accuracy and robustness [55, 128].

Despite their stable placement [16] and minimal susceptibility to motion artifacts [30] that make them equally compelling as a platform for motion sensing, earables remain comparatively under-examined for gait-based security applications, and present a promising yet untapped opportunity for IMU-based recognition.

To this end, this work aims to contribute to this field of research **by investigating the feasibility of IMU-equipped earables for gait-based user authentication and identification.** A central challenge herein is the lack of publicly available datasets to conduct and baseline such research, so **to overcome this limitation, the authors collect a new dataset** using the OpenEarable platform [88] encompassing multiple locomotion activities—such as walking, running, and stair climbing—at varying intensities. By openly releasing this dataset, we seek to establish a foundation for future research and foster collaboration in the field of earable-based biometric authentication.

Moreover, this study recognizes the importance of resource efficiency for on-device deployment on low-power, memory-constrained hardware. To address this, a range of machine learning (ML) and deep learning (DL) models are evaluated under a comprehensive framework that measures both authentication (via Equal Error Rate) and identification (via Weighted F1 Score) performance, along with resource utilization and efficiency. Hardware-aware optimization techniques such as quantization for deep architectures and feature selection for shallow ML algorithms are applied to make the said models viable for edge deployment.

To summarize, the key contributions of this work are as follows.

- First, it introduces a publicly accessible dataset that captures multiple locomotion activities (walking, running, stair climbing) across several intensities from 30 participants. This dataset addresses the scarcity of open-source resources for earable-based gait-based authentication, and lays a foundation for further research.
- The work proposes a comprehensive end-to-end pipeline for gait-based user recognition that includes data collection, preprocessing, feature extraction, feature selection, classification, and model optimization that leverages both shallow ML models (Random Forest, SVM, kNN, MLP) and a hybrid DL architecture (CNN_fixLSTM). The pipeline supports both activity-specific and across-activity gait-based identification and authentication tasks.
- The feasibility of earables as a platform for gait-based user recognition specific to an activity, and in an across-activity manner, is evaluated. Specifically, the combined use of accelerometer and gyroscope data is investigated along with the effectiveness of each locomotion type for user verification.
- Lastly, to address the computational limitations of earables, quantization to selected deep models, and feature selection techniques to shallow ML models, are applied, and their effictiveness is measured. Trade-offs between making the model lightweight and maintaining high accuracy are analyzed.

The remainder of this document is organized as follows. Chapter 2 surveys gait as a biometric trait, examines various sensing approaches for gait-based user recognition, explores earables as emerging sensing platforms, and reviews the state-of-the-art in authentication and identification with earables, identifying key gaps in the field. Chapter 3 outlines the methodology, including experimental design, data collection protocol, data preprocessing, classification models, and evaluation metrics. Chapte 4 presents a comprehensive analysis of the experimental results, discussing recognition identification performance alongside resource usage. Finally, Chapter 5 concludes the thesis by summarizing the findings, reflecting on limitations, and outlining future directions.

Chapter 2

Related Work

This chapter explores the existing body of work that informs this research, providing a foundation for understanding the advances and challenges in gait-based ear-worn recognition systems. It begins by discussing gait as a unique biometric trait, highlighting its distinct advantages for user identification and authentication. The main sensing approaches for gait-based user recognition, namely vision-based, floor-based, and wearable sensor-based methods, are examined. The role of machine learning in inertial gait-based authentication is explored, covering established methodologies and their applicability to resource-constrained environments. Additionally, publicly available datasets relevant to IMU-based gait-based recognition through earables are reviewed, identifying gaps that motivate the need for a new dataset tailored to gait-based recognition using earables, and comparing the proposed dataset to the existing ones. Finally, the challenges of deploying shallow and deep learning models on low-power devices are discussed, setting the foundation for the research contributions presented in later chapters.

By synthesizing previous research across these domains, this chapter contextualizes the work undertaken in this thesis, and establishes the rationale for the proposed methodology and assumed contributions.

2.1 Gait as a Unique Biometric Trait

User authentication is a critical component of personal computing systems that provides secure user access to various devices. Traditional Knowledge-Based Authentication (KBA) is defined as "a form of user authentication that requires verification of identity claimed by matching one or more pieces of information presented by the user or claimant against the information sources associated with the claimant" [12] (e.g., passwords and swipe patterns) and, while commonly used, is shown to be increasingly vulnerable to various attacks such as automated dictionary and brute-force attacks, and can be visually observed (shoulder surfing) or captured using keyloggers (malware that records keystrokes) [3]. As a result, biometric authentication, defined as "establishing identity based on the physical and behavioral characteristics of an individual such as face, fingerprint, hand geometry, iris, keystroke, signature, voice, etc. [51]", has emerged as a more reliable and essential solution to identity verification that aims to bridge the gap between secure authentication and usability [81].

Biometric traits are classified as physical or behavioral (and can further be divided into hard and soft ones), with the former relying on static physical attributes such as fingerprints and iris, typically used for one-time verification, and the latter - behavioral biometrics - being based on behavior patterns of an individual including gait, keystroke dynamics, and voice recognition [51]. Behavioral biometrics are particularly suitable for continuous authentication (CA) wherein

the system continuously verifies the user identity throughout a user session.
Among said biometrics, gait-based CA systems have been an attractive research field for years as walking is inherently a repeated task that can be measured unobtrusively and continuously, and is also hard to imitate [81].

Gait is formally defined as the periodic and dynamic motion of the human body during locomotion, involving the displacement of the center of mass and the coordination of lower limbs and the trunk to move from one position to the other [94]. Previous research has shown that human gait, as it is influenced by individual biomechanics, physiological factors, and optimization principles, is a uniquely identifiable and powerful biometric trait highly resistant to various forms of impersonation [112, 72, 74, 76]. It is also a feature that can be applied in long-distance and uncontrolled scenarios without explicit cooperation by subjects [127], making it suitable for remote identification and authentication tasks.

For instance, a study was conducted in which attackers were trained to imitate the gait of a preselected victim under controlled conditions. Despite significant efforts, the attackers exhibited no meaningful improvement in their ability to mimic the victim's gait, eventually hitting a "plateau", or an inherent boundary that limited the attackers' performance irrespective of training time. [74]. This underscores how gait involves complex motor patterns highly dependent on the physiology and biomechanics of an individual. Similarly, another work investigated the security of gait-based authentication systems on smartphones under both zero-effort and minimal-effort impersonation attacks. Using smartphone accelerometer data, the study reported a 0% false match rate (FMR) under impersonation scenarios, even when attackers were professional actors trained in mimicking body motions. Furthermore, attackers were shown to lose regularity in their own gait in 29% of impersonation attempts [76]. This disruption in natural walking patterns highlights the difficulty of mimicking gait without degrading the biometric signature.

Few works have demonstrated that the possibility for a spoofing attack does exist with relatively high success rates (26%) [58] by adjusting gait characteristics such as speed, step length, step width, and thigh lift yet such attacks have only been performed with the help of an off-the-shelf treadmill and with assumed access to the target's biometric samples which in itself introduces additional steps to spoofing.

These findings establish a strong foundation for leveraging gait as a biometric for secure and unobtrusive user recognition.

2.2 Sensing for Gait

Gait-based recognition is the process of identifying or authenticating individuals based on the unique patterns they generate while e.g., walking or runnign. These patterns can include visual features, dynamic weight and pressure distributions, and acceleration data captured by wearable sensors. Gait-based recognition is thus conducted through three main approaches [33]: Machine Vision (MV) based, Floor Sensor (FS) based, and Wearable Sensor (WS) based.

In the **MV-based approach**, gait is captured from a distance using a camera. As said data can be collected without physical interaction with the system, the approach is considered non-invasive. It does, however, often require high levels of preprocessing to extract the gait information from the recorded data as varying illumination, occlusion, pose, view angle variations, appearance changes due to different clothing, occlusion due to multiple people walking in a group, and objects carried by the subject, among others, can pose challenges to (direct) successful recognition [97].

Similarly, in the **FS-based approach**, researchers rely on pressure and weight measurements from sensor-equipped floors to quantify physical user attributes. While the MV-based approach is the most widely established, FS-based gait-based recognition—or footstep recognition—offers several advantages. It is less affected by lighting changes, background movement, and viewing angles, making it more robust in varied conditions. Additionally, FS-based methods can reduce the user discomfort often associated with video collection [61]. Even so, as the approach requires special equipment, and given the objective of this study to leverage widely accessible and portable devices with minimal reliance on additional infrastructure, the focus will be directed toward the third mentioned approach: WS-based approaches. Consequently, MV- and FS-based methods will not be considered further in this research.

The **WS-based method** relies on devices equipped with motion sensors worn or attached to various locations on the body to capture motion dynamics. Unlike the MV and FS approaches, WS-based systems do not require any environmental setup, making them portable, compact, and suitable for continuous and ubiquitous gait-based recognition. Furthermore, devices for everyday use like earbuds, smartwatches, and phones typically come with a built-in accelerometer and gyroscope so additional external hardware is often not required. Building on this advantage, a range of wearable sensors, including accelerometers, gyroscopes, and magnetometers, and a fusion of such [98, 91, 65], are shown to be able to capture distinctive gait patterns, with the accelerometer being effectively one of the most suitable modalities to be used as a single metric that achieves high recognition accuracy [34, 64, 69, 23, 85, 16].

Accelerometers measure how much an object moves or shifts from its static position, in three directions: `X, Y, Z`. At any given moment, the accelerometer provides three values corresponding to these directions. Among the most important features an accelerometer provides is its linearity, meaning the readings are proportional to the actual physical acceleration, and the bandwidth, or sampling rate, which refers to how often the sensor can detect changes, measured in Hertz (Hz). For gait-based recognition, these two characteristics are the most relevant ones [72]. In this context, the gait produces a pattern of accelerations over the three axes that are best captured by sensors integral to the body parts. The produced pattern is characteristic of each individual. In modern devices, accelerometers are often integrated into single-component micro-electro-mechanical systems (MEMS) alongside a gyroscope, and, occasionally, a magnetometer. While magnetometers are not commonly employed in gait-based recognition, the effect on performance when combining accelerometer and gyroscope readings has shown mixed results [93, 72], with some works reporting only a marginal increase in performance after sensor fusion [73], and others - a more substantial improvement [60].

The unobtrusiveness of wearable-based gait tracking and authentication makes it particularly appealing as a method, as users are not required to perform explicit actions to verify their identity. Instead, their gait patterns, captured as they move naturally while wearing the device, serve as a reliable biometric for continuous authentication. Building upon the strengths of wearable sensor-based approaches, **the scope of this research narrows to a specific class of wearable devices — earables — as they offer great advantages for gait-based authentication by leveraging unique positioning and high-level accessibility.**

The capabilities and applications of earables are explored in Section 2.3.

2.3 Sensing with Earables

Earables are "devices that attach in, on, or in the immediate vicinity of the ear to offer functionalities beyond basic audio in- and output" [87]. While originally intended for audio playback, with the help of the rapid advancements in Artificial Intelligence and the Internet of Things in recent years, these ear-worn devices have become a one-of-a-kind platform for ubiquitous computing equipped with advanced sensing capabilities.

Until recently, much of the research on wearable-based sensing focused on smartphones [40, 8] and smartwatches [6, 121] as the current mass-market wearables. However, earables have by now evolved into versatile smartphone-like devices that increasingly integrate more and more sensors that expand their functions, with the global earphone and headphone market share estimated to grow continuously, namely with a Compound Annual Growth Rate between 10-12%, or with nearly USD 24 Billion between 2024 and 2028 [37, 104].

The predicted growth of the global ear-worn device market has prompted researchers to explore the various sensing capabilities earables offer. These span several areas, including skeletal movements such as (real-time) gait detection [79, 24], facial expressions detection [110, 67]; nervous system monitoring such as brain activity analysis through Electrocardiography (EEG)-equipped earables to, among others, detect epileptic seizures [38], and identify attention states [53]; cardiovascular metrics measuring blood pressure [15] and generic health monitoring abilities [19]; respiratory patterns including breathing mode detection [43]; energy expenditure [9, 36]; reactions to music [63]; digestive activities like tracking snacking behaviors [14]; and various authentication methods using e.g., the ear as an echo chamber to amplify internal body sounds [35, 18], and in-ear Photoplethysmography (PPG) signals [20] as novel biometrics.

What makes earables suitable means to sense said things are, non-exclusively, the position of the ears on the head that allows for various muscle activation tracking [10]; the ear being a secure attachment point for capturing movements of the body and head [16] that is relatively stable and thus less susceptible to motion artifacts and noise [31]; furthermore, ear-worn devices stays consistently in the same location, are firmly attached to the user's body, and are not directly influenced by external disturbances during movement that e.g., smartwatches endure, like carrying objects; their shape typically allows for them to be worn seamlessly and continuously throughout the day.

This diversity in detection functions highlights the potential of earables to revolutionize various fields, seamlessly integrating health monitoring, activity recognition, and biometric authentication, among others, into everyday life. It also highlights the wide variety of sensors an earable can be equipped with, including PPG, EEG, IMU, pressure sensor, thermistor, infrared thermometry, microphone, electromyography (EMG), proximity sensor, hygrometer, piezoelectric sensor, impedance sensor, and haptics. To group these applications systematically, in their work [87], Roddiger et al. propose four main areas of phenomena that can be sensed with earables: (i) physiological monitoring and health, (ii) movement and activity, (iii) interaction, and (iv) authentication and identification. Out of the enlisted, **this research focuses on the authentication and identification category.**

Section 2.4 will delve deeper into the types of authentication enabled through earables, presenting the potential of gait biometrics as a secure and effective motion-based method for user verification.

2.4 Authentication and Identification with Earables

Taking into account recent work to raise earables as stand-alone devices [84], and their widespread integration with AI assistants like Siri and Alexa, this piece of technology continuously gets granted access to sensitive multi-source information. This makes the need for an authentication accelerator for earables evident. Traditional authentication methods such as fingerprint and face recognition require installing additional hardware on these already densely packed with sensors devices, lowering ease of use and increasing costs.

As earables advance in sensing capabilities, they offer an opportunity to explore biometric authentication methods that rely on their existing hardware. Voice authentication is one such means of providing device and information access using the in-built microphone, yet it is vulnerable to various spoofing and replay attacks [126, 62], and is also more susceptible to noise. Recent research has focused on other (novel) physiological and behavioral biometrics signals that can be captured on earables, by either combining existing ones, repurposing them, or discovering new ways of sensing them, as addressed below.

Multiple works utilize inherently available earbud sensors such as in-ear facing microphones. Examples of such projects are EarGate [30] which performs gait-based identification from the sounds induced by walking that get propagated through the musculoskeletal system in the body, and EarSlide [117] that captures distinctive features of acoustic fingerprints when users slide their fingers on the face. Hu et al. [45] leverage the difference between in-ear and out-ear sounds, and estimate an individual-specific Occluded Ear Canal Transfer Function (OECTF) to distinguish between users.

Additionally, authentication may be performed by analyzing the dynamic structural changes of the ear canal while e.g., talking [113], or as a combination of ear canal deformations and tooth print acoustics [116]. Such authentication schemes usually require users to perform chewing or speaking actively, to emit a one-time sound, or require the joint of transceiver sensors (e.g., speaker and microphone). On the other hand, some works, while still making use of the in-ear microphone, facilitate passive user authentication instead like BreathSign [39] that is based on biometrics in bone-conducted breathing sounds, and HeartPrint [18] which exploits bone-conducted heart sounds.

Despite the potential of audio-based solutions, they come with notable challenges. Firstly, similar to voice authentication, they present privacy concerns as such systems often rely on always-on microphones, raising questions about unauthorized data collection and misuse. Furthermore, the high sampling rates required for accurate audio processing demand substantial computational resources, leading to increased energy consumption and possible strain on device performance. Moreover, these systems remain susceptible to (environmental) noise, with some works reporting that e.g., the frequency of sound playing in earphones and the frequency of breathing-induced body sounds overlap greatly [39]. Filtering for noise artifacts appears as an additional preprocessing step to perform authentication. These limitations underscore the need for alternative approaches that address the said shortcomings.

Examples of authentication based on more advanced sensors are EarPass [66] which utilizes PPG sensors to capture in-ear PPG signals with high accuracy yet requires an additional integration of PPG sensors onto commercial devices; and Jawthenticate [100] which utilizes a custom-built earable prototype with two IMUs - one on the temporomandibular joints, and one on the temporal bone, to allow for authentication through distinctive speech motions without the need of a microphone. These and other methods of similar advancement either require extensive hardware modifications or rely on custom prototypes not readily available in consumer devices. In contrast, standard IMU sensors are readily available in commercial earbuds (e.g., in Apple Airpods Pro, Samsung Galaxy Buds Pro, Sony WF-1000XM4), and present a promising alternative for user recognition due to their accessibility and ability to capture motion data with minimal computational overhead. Several works utilize said type of sensor in varying ways; for instance, BudsAuth [114] leverages IMU data from vibration signals induced by on-face touching interactions with the earbuds. MandiPass [68] extracts a novel biometric from the vibration signal of the user's mandible picked up by IMU sensors; the mentioned model is not tested further due to limitations in obtaining raw IMU data from commercialized wireless earbuds that are not open-source. Lastly, LR-Auth [44] utilizes the modulation of sound frequencies by the user's ear canal occlusion, generating user-specific templates through linear correlations between two audio streams. Table 2.1 provides an overview of the existing earbud recognition systems in related research, highlighting the novelty of our proposed system. Performance is only reported for authentication as the majority of the enlisted works do not report on identification scenarios.

Table 2.1: A comparison of existing earbuds authentication systems in terms of authentication performance.

Paper	**Sensor**	**Biometrics**	**Backbone**	**Performance**
EarGate [30]	Microphone	Gait	SVM	92.5% (BAC)
HeartPrint [18]	Microphone	Heartbeat pattern	CNN	1.6% (FAR), 1.8% (FRR)
EarSlide [117]	Microphone	Fingerprint	SiameseNN	96.86% (ACC)
TeethPass [120]	Microphone	Teeth structure	SiameseNN	96.8% (ACC)
ToothSonic [116]	Microphone	Teeth structure	DNN	92.9% (ACC)
BreathSign [39]	Microphone	Breathing pattern	Triplet network	95.22% (ACC)
Hu et al. [45]	Microphone	Ear canal geometry	Linear Regression	4.84% (BER)
EarEcho [35]	Microphone	Ear canal geometry	SVM	94.52% (BAC)
LR-Auth [44]	Microphone	Ear canal geometry	Cosine similarity	99.8% (BAC)
EarPrint [129]	Microphone	Behavioral & Physiological Acoustics	DML	4.23% (EER)
EarPass [66]	PPG	Heartbeat pattern	SVM	98.7% (ACC)
EarPPG [20]	PPG	Ear canal geometry	ReGRU	94.2% (ACC)
BudsAuth [114]	IMU	Tissue composition	DAL-CNN	5% (EER)
MandiPass [68]	IMU	Mandible structure	CNN	1.28% (EER)
Jawthenticate [100]	IMU	Jaw structure	SVM	92% (BAC)
This paper	**IMU**	**Gait**	CNN+LSTM	1.96% (EER)

Despite the established uniqueness and robustness of gait as a biometric trait, as discussed in Section 2.1, the demonstrated effectiveness of gait for IMU-based authentication in other wearables [125, 102, 121], and the high accuracy achieved in gait detection and classification using earables [54, 16, 25, 122], to the best of the authors' knowledge, **no research has yet proposed a gait-based authentication system using IMU data from earables**, **nor has the feasibility of such systems been systematically evaluated for real-world deployment.** This, along with the lack of publicly available datasets collected for this purpose, creates a significant gap in assessing earables' viability for secure user authentication.

This research addresses this gap by evaluating the effectiveness of gait-based authentication using IMU-equipped earables, considering both algorithmic performance and practical deployment constraints. By investigating the trade-offs between model complexity, resource efficiency, and authentication and identification performance, this work contributes to the development of lightweight biometric solutions tailored for real-world use on commercial ear-worn devices. Although similar works specific to earables do not currently exist, this research draws on established methods for IMU-based gait-based recognition in other wearable devices, adapting and optimizing these techniques to fit the unique constraints and advantages of earables.

Building on this foundation, Section 2.5 explores established machine learning techniques for inertial sensor-based gait-based recognition and evaluates their effectiveness in the context of biometric authentication.

2.5 Gait-Based User Recognition and Machine Learning

There are two training components gait-based user recognition consists of:

An **Identification Component** that has the form of a multi-class classification problem wherein the aim is to automatically identify a target subject given their gait information. Assuming M target subjects, given an unknown gait segment, a gait-based identification system provides the corresponding subject identity, ϕ, where $\phi \in \{1, \ldots, M\}$. In a closed-set scenario, which is common in controlled datasets, the system assumes that every test sample belongs to one of the M known individuals, and the model must minimize misclassification errors between them.

An **Authentication Component**, or a two-class classification problem, aims to verify whether a given gait segment belongs to a target subject (genuine user) or not (impostor). In literature, creating impostor feature vectors is commonly done by labeling the users in the dataset other than the genuine user in the Test dataset as impostors [55, 59, 58]. For authentication, the system evaluates each participant individually within each activity category, resulting in M binary classification problems per activity. Given an unknown gait segment, the system outputs either a positive classification, indicating the segment belongs to the genuine user, or a negative classification, identifying the segment as originating from an impostor. To address the inherent class imbalance in this setup, where the genuine user's samples are significantly outnumbered by those of the impostors (e.g. herein, 1:29), a data balancing approach is commonly adopted. This ensures that both classes are represented equally in the training set, preventing the model from being biased toward the majority class.

The effectiveness of gait-based recognition using IMU sensors heavily depends on the ability to extract meaningful patterns from raw motion data. Given the complexity and variability of human gait, machine learning is a crucial tool for analyzing IMU signals, enabling the identification of distinctive movement signatures between individuals. These methods leverage statistical patterns, temporal dependencies, and feature representations to differentiate users based on their gait. The process consists of several key steps, namely data segmentation, feature extraction and selection, and model training, as enlisted in Subsections 2.5.1, 2.5.2, 2.5.3, subsequently.

2.5.1 Data Segmentation

Before distinctive features can be extracted from IMU signals, the raw time-series data must be segmented into smaller, structured units that capture relevant gait cycles. Segmentation is a crucial preprocessing step, as gait is a continuous motion that varies in speed and rhythm across individuals and, without proper segmentation, the extracted features may fail to capture the periodicity and distinctive motion patterns inherent in gait. Signal segmentation is done either by extracting gait cycles [123, 26, 76, 81] or fixed-length frames [92, 58, 4, 57].

Gait cycle-based segmentation involves identifying repetitive walking patterns, defined as the time interval between consecutive heel strikes of the same foot that serve as markers to segment data into individual gait cycles. Said gait cycles are extracted, and an average cycle template is generated for classification. A point-wise comparison of train and test samples is carried out using classical distance measures (Euclidean, Manhattan) and Dynamic Time Warping (DTW) [123, 28, 77].

Frame-based gait segmentation divides inertial signals into multiple (overlapping or non-overlapping) frames or windows of the same length and, for each window, a list of features is computed. The fixed-length frame-based approach combined with machine learning is shown to outperform cycle-based approaches in the context of authentication, as per multiple previously conducted surveys [106, 7, 58], **and is therefore the segmentation method of choice for this paper.**

Frame-based approaches, especially in the case of small and unbalanced datasets, are often paired with a fixed-size sliding window (FSW) approach. In FSW, the raw data is segmented into overlapping frames of fixed length wherein the length (in seconds) is chosen in a way that each frame contains information for at least one gait cycle so that the complete features exist in each period [99, 96]. Overlapping ensures that a portion of the data from one window is included in the next one, allowing for smoother transitions between frames. In the case of time-series data, using overlapping windows also helps capture transitions or events that might otherwise be split across two adjacent non-overlapping windows, and, ultimately, provides more data samples to feed into the chosen models. The common overlap used in literature is 50% [16, 13, 58], while the window length varies per work, ranging between 1 and 12 seconds. According to best practices, the number of valid windows considered per subject per feature set should be at least 10 times [111] and ideally 20 times [32] more than the number of features that are inputted into the model.

2.5.2 Feature Extraction and Selection

Once the IMU data has been segmented, the next crucial step in gait-based recognition is feature extraction, where meaningful characteristics are derived from each window. The choice of features directly impacts the classification accuracy, computational load, and robustness of the authentication system [90]. Feature extraction methods can be categorized into statistical (e.g., mean, variance, standard deviation, skewness, and kurtosis), frequency domain (power spectral density, entropy), and time-series features (peak-to-peak intervals, step regularity), each capturing different aspects of the gait signal. A detailed summary of the most commonly used accelerometer and gyroscope features can be found in [113], which also serves as the basis for the feature extraction process of this work.
The exact features this work extracts are addressed in Subsection 3, and enlisted in Table 3.1.

2.5.3 Machine Learning Models

Machine learning models play a crucial role in gait-based recognition, with traditional machine learning (ML) classifiers and deep learning (DL) models both demonstrating effectiveness in different scenarios.

While no studies have yet explored training ML and DL models on earable IMU data collected during gait for recognition purposes, extensive research has been conducted on gait-based recognition using IMU data from other wearable devices. The most commonly used ML algorithms for this purpose include Random Forest (RF) [59], Support Vector Machine (SVM) [70, 118, 55, 81, 86, 111, 5, 124], k-Nearest Neighbours (kNN) [108, 107], Logistic Regression [58], and Multi-layer Perceptron (MLP) [29, 4], with SVM being a common choice due to not being computationally expensive, and to being able to work well with unstructured data and in the presence of a small sample size [89]. SVM demonstrated an accuracy above 90%, with peaks of 96%, in different experimental scenarios [72]. SVM, RF and kNN are generally indicated as suitable for dealing with datasets of small scale [108], while MLP is known for having the lowest model size, making it suitable for deployment on resource-constrained devices [4]. The enlisted models are shown to consistently score high in the task of gait-based identification, often exceeding 85% [72, 22, 108]. **Given their proven effectiveness in similar contexts, these ML algorithms are selected as part of the evaluation for this work.**

However, while conventional ML models provide strong performance, they depend on handcrafted features that often require domain expertise for optimal selection, which may limit their adaptability to complex gait variations. Deep Learning (DL) methods offer the advantage of automatically learning and extracting discriminative patterns from IMU signals, making them effective for modeling the data [56]. Various architectures have been explored for both gait-based authentication (verifying whether a gait sample belongs to a known user) and gait-based identification (classifying a gait sample among multiple users). The most widely used DL methods in the literature include Convolutional Neural Networks (CNNs) [11, 46] followed by Recurrent Neural Networks (RNNs), more specifically Long short-term Memory (LSTM) [82, 109]. Research also suggests that combining CNN and LSTM outperforms competing deep non-recurrent networks [42]. Said CNN-LSTM hybrid models have shown superior performance compared to standalone CNN or LSTM models in both authentication and identification tasks, as confirmed by [27, 48, 17], scoring as high as 96% recognition accuracy. This combination ensures that both short-term variations and long-term gait patterns are effectively modeled, improving robustness against noise and individual gait fluctuations, **making CNN-LSTM a compelling choice for this work.** Given their success across multiple wearable-based gait studies, their application to earables remains an unexplored yet promising direction for effective user authentication and identification. This work addresses this gap **by investigating whether gait-based recognition using ear-worn IMU sensors can achieve similar levels of reliability as other wearable-based approaches,** advancing the understanding of gait biometrics in the context of earable devices. The specific model implementation this work follows can be found in [17], with its adaption for this work detailed in Subsection 3.3.2.

To evaluate gait-based authentication methods on earable data, it is essential to consider existing datasets that capture relevant motion patterns for the use case. Subsection 2.6 explores publicly available datasets that contain IMU data from earables and assesses their suitability for this research.

2.6 Publicly Available Datasets

This subsection reviews the three (to the best of the authors' knowledge at the time of writing this work) publicly available datasets relevant to the field of gait-based recognition using earables, at least in terms of utilized sensors and recorded activities.

Comparing existing works in the literature is challenging due to the scarcity of publicly available benchmarks for IMU data collected from the ear. While the potential of earables for gait-based authentication is rather evident, the development and evaluation of such systems relies heavily on the availability of suitable datasets. A major challenge not only for gait-based authentication using earable-collected IMU signals but also for behavioral biometrics research as a whole is the limited number and accessibility of real-world datasets [101]. Additionally, the fairly recent emergence of earbuds with sensing capabilities further contributes to the lack of well-curated, open-source datasets specific to this technology [21].

While no publicly available datasets have been designed for gait-based authentication through ear-collected IMU, some existing datasets could be used for that purpose given they contain accelerometer data from ear-mounted or in-ear devices (and optionally - gyroscope and magnetometer data) during gait-related movements such as walking, running, jogging, and stair climbing. The said datasets, while not originally intended for authentication tasks, can still be leveraged to derive gait patterns. Table 2.2 provides an overview of their key characteristics, with emphasis on activities relevant to gait-based recognition. For comparison, the dataset collected for this research is included in the discussion.

The **WEEE** dataset is part of the work "A Multi-Device and Multi-Modal Dataset for Wearable Human Energy Expenditure Estimation" [36] and is designed for energy expenditure estimation across multiple devices and body locations, including the ear, wrist, chest, and head. It collects a wide variety of data, with these being oxygen consumption, heart rate, gyroscope, accelerometer, PPG, ECG, EDA, EEG skin temperature. This is done through the use of 7 different devices placed across the human body. The following continuous physical activities are recorded: sitting, standing, cycling, and running. This is accompanied with questionnaires, participant demographics, and participant body composition information. The device used to collect the IMU data is a Nokia Bell Labs earbud. The dataset itself is hosted on Zenodo [49] and involves recordings of 17 participants. The provided repository comes with a study information file that contains a column per participant for errors that occurred during their recordings, such as “Stopped the treadmill at 11:30 - Earbud fell off at 11:39” (P16), “Interruption at 18:41” (P4), and similar. These have not been properly documented and require individual inspection of each recording in order to manually determine when precisely the interruption or error occurred, and when and if it was fixed again as no timestamps have been provided for the end of the issue,

Dataset	Year	Earable Type	Device Position	Sensor(s)	Sampling Rate	Relevant Activities	Participants	Recording Length per Activity	Setting	Additional Notes
WEEE [36]	2022	Nokia Bell Labs eSense Earbuds	Right Ear	Accelerometer, Gyroscope, PPG	100 Hz	Running (Speed 1, Speed 2)	17 (12M, 5F)	10 minutes	On treadmill; self-determined pace	Recording issues reported but not time-stamped
Auritus [90]	2022	Nokia Bell Labs eSense Earbuds	Both Ears	Accelerometer, Gyroscope	100 Hz	Walking, Jogging	45 (29M, 16F)	23-25 seconds	On ground; self-determined pace	Built-in Butterworth filter with 5 Hz cutoff applied
EarSet [75]	2023	Custom Earbuds	Both Ears	Accelerometer, Gyroscope, PPG	100 Hz	Walking, Running	30 (18M, 12F)	2 minutes	On treadmill; preset speed	-
Our dataset	2024	OpenErable Earbuds	Right Ear	Accelerometer, Gyroscope, Magnetometer	50 Hz	Walking, Running, Stair Climbing (Speed 1, Speed 2)	30 (21M, 11F)	4 minutes	On treadmill, on stair master; self-determined pace	-

Table 2.2: Summary of Freely Available Datasets Collected Through Earables

only the start of it. As there is no supporting text that addresses this in the research paper, it is not clear how to interpret the reported errors. Some participants also lack earbud recordings from left ear, and/or include an additional file named “PX_EARBUDS_Final.csv” while neither the paper nor the GitHub repository indicate any information about said inconsistencies. No data preprocessing is applied, recommendations are given to convert the raw accelerometer and gyroscope data to a more usable format (i.e., milli-g and milli-dps), and to remove the direct current offset from the gyroscope data by applying a Butterworth band-pass filter; The lack of clear documentation regarding said interruptions, and the presence of unaddressed files add further challenges to the dataset’s reliability.
This dataset is not pursued further in this work due to the researcher not being able to verify whether said issues have properly been handled, and relying only on manual (i.e. visual) inspections; and due to the lack of variability in locomotion activities (no walking is included).

The **Auritus** dataset is part of the work "Auritus: An Open-Source Optimization Toolkit for Training and Development of Human Movement Models and Filters Using Earables" [90], and one of its goals is to provide a comprehensive dataset designed for earable-based activity recognition and head-pose estimation. It leverages the Nokia Bell Labs earbuds. The dataset addresses the scarcity of open-source earable data by offering nine distinct activities of daily living (ADL), including walking, jogging, jumping, sitting, and lying. The work collects data from 45 participants, and includes calibration processes for the IMU sensors to mitigate sensor biases and improve accuracy. A custom graphical data-labeling tool is developed to facilitate the segmentation and labeling of the data. Albeit the wide range of activities, each activity is relatively short, with namely 23-25 seconds IMU trace per participant.
The dataset is indicated to be hosted on GitHub [103], yet said repository contains further links redirecting to a Google Drive location, which, at the date of writing this work, point to non-existent files. This renders the dataset inaccessible and ultimately unusable, as there is no alternate way provided to access the data.

The **EarSet** dataset is part of the work "EarSet: A Multi-Modal Dataset for Studying the Impact of Head and Facial Movements on In-Ear PPG Signals" [75] and is tailored for exploring how subtle head and face motions affect in-ear IMU and PPG signals. Data collection is performed using custom earbuds equipped with a 3-channel PPG sensor and a co-located 6-axis IMU (accelerometer, gyroscope). These devices were placed in both ears to simultaneously record 18 data streams at a sampling rate of 100 Hz. The study includes data from 30 participants, with a diverse representation of skin tones. The participants perform 16 activities, including head motions (nodding, shaking, tilting), facial motions (brow raiser, lip puller, mouth stretch, etc.), and full-body activities (walking at 5 km/h and running at 8 km/h on a treadmill). The raw data is hosted on Zenodo [2]. The dataset includes a total of around 17 hours of recordings, with each activity session lasting 2 minutes. During these sessions, the PPG sensor configurations are varied every 30 seconds to explore the trade-off between power consumption and signal quality. Ground-truth physiological measurements, including heart rate, heart rate variability, and respiration rate, are collected for validation purpose, and also included on Zenodo. Additionally, the dataset contains an additional file, Demographics.csv, with demographics and skin tone of each participant in an anonymous format. No preprocessing is implemented; the conversion recommendations are the same as in WEEE. As the relevant to this work activities are performed at set pace, the dataset does not capture intra- and inter-subject variability in gait (e.g., self-paced walking, varying speeds per activity, etc.) which limits its applicability for user authentication.

The lack of publicly available datasets for gait-based user authentication using earables is a significant barrier to research in this field. Existing datasets, namely WEEE, Auritus, and EarSet, are not suitable due to various constraints, e.g. incomplete or inaccessible data and limited locomotion variability. To address the identified gap, **this work introduces a new dataset designed for gait-based recognition with earables**. Said dataset is intended to be made publicly accessible in order to stimulate further research, and to enable the development and evaluation of novel methodologies in the field. Summarized details of the dataset are found in Table 2.2, while the data collection process is described in Chapter 3.

2.7 Deployment on Resource-Constrained Devices

The deployment of ML and DL models on resource-constrained devices, herein on **Arduino Nano 33 BLE Sense**, is challenging due to their limited computational power, memory, and energy efficiency. To ensure the feasibility of deploying participant recognition models on the OpenEarable earbud used in this work, several model optimization techniques commonly used in research have been explored in this work.

In **DL**, models are commonly trained and deployed using 32-bit floating-point (FP32) arithmetic. However, this can be inefficient for deployment on memory-constrained devices. To tackle this, **quantization** is adopted to reduce numeric precision (e.g., to 8-bit integers) making the model size smaller, and speeding up inference.
Two primary forms of quantization are explored, namely Post-Training Quantization (PTQ) and Quantization Aware Training (QAT). Post-Training Quantization (PTQ) is a technique applied after a model has been fully trained in floating-point. It converts the model's weights from 32-bit floats to a lower bit precision, herein 8-bit integers. The idea is to reduce model size and enable faster inference without having to retrain the model from scratch. Quantization Aware Training (QAT), on the other hand, involves simulating the quantization effects during the training phase itself. While the model still stores weights in floating point during training, the forward pass injects fake quantization operations so the network learns to be robust to 8-bit precision. This helps preserve accuracy after quantization at the cost of additional training time and complexity. Prior work has demonstrated that PTQ and QAT are effective in reducing the computational footprint of models for TinyML applications [78], particularly in human activity recognition on low-power devices. However, these techniques remain underexplored for participant recognition, an area this work aims to contribute to.

For **ML** models, feature selection and hyperparameter tuning are explored. When working with a large set of features, there is a higher chance of overfitting as the ML classifier risks learning from the specific dataset rather than uncovering broader patterns, and loses on generalization capabilities [108]. Meanwhile, computation overhead is also increased. To counter this, classification algorithms such as Random Forest [108], Sequential Backward Selection (SBS), Recursive Feature Elimination (RFE) [107], Principal Component Analysis (PCA) [112], and Mutual Information [58] can be applied to reduce the large set of initial features to relevant gait variables, and to lower inference time. This is often paired with correlation filtering to avoid multicollinearity [108]. To improve the generalization performance of ML models, grid search is employed to find the optimal set of hyperparameters. By systematically evaluating different combinations, this process ensures that the final models achieve high accuracy with minimal resource usage. Details on how this is implemented in this work can be found in Chapter 3.

Chapter 3

Methodology

Given the scarcity of publicly available datasets that meet the requirements for gait-based user authentication using earables, this study designs and collects its own dataset. The dataset aims to address the shortcomings identified in Section 2.6 by ensuring comprehensive locomotion diversity, high data quality, and a structured experimental protocol. This section outlines the methodology used to collect, preprocess, and analyze the data, and the feature extraction and selection process. Then, the experimental protocol is described, along with the pipelines for ML and DL-based user authenication and identification. Finally, to ensure feasibility for real-world deployment on edge devices, the most suitable models are further optimized to reduce their computational complexity and memory footprint while maintaining classification accuracy. This methodology ensures a structured and reproducible approach to gait-based user recognition using earables.

3.1 Device and Experimental Setup

For this purpose of collecting the target dataset, a controlled experiment is designed wherein participants are asked to perform a set of predefined tasks in the form of continuous gait-based movements in a gym setting. The choice for a controlled experiment allows for running a detailed analysis of the data and data collection process and supports a higher level of replicability of the experiment. The experiment is conducted following the ethical regulations of the University of Twente, more specifically those of the Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS), and the Computer And Information Sciences (CIS) committee has granted approval to this study's design and methodology. All participants have signed an informed consent form and have agreed to have their anonymized data used for research purposes.

The dataset is planned to be made publicly available. This section provides an overview of the technologies relevant to this report and outlines how said dataset is collected.

3.1.1 Device

The device used for conducting the experiment is the earpiece from OpenEarable [105] - an open-source web-based dashboard connected to the earbud via Bluetooth Low Energy (BLE) - that visualizes the data being collected in real-time, and allows for configuring sensor settings and starting and ending the recordings remotely [88]. It also includes a mobile smartphone app. As OpenEarable is a fully open-source platform with firmware and hardware developed using free tools, its use aligns with the goals of this research by providing a customizable solution for sensor configuration and data collection, enabling reproducibility and collaboration within the field. The said device is furthermore chosen because it functions as regular binaural Bluetooth in-ear earphones, making it closer in similarity to everyday use devices, while still providing the sensing capabilities necessary for this experiment [88]. Its design and sensor performance have been validated in various studies comparing its accuracy and reliability relative to established benchmark measurements [105]. The device has previously also been paired with mobile applications for posture tracking, jump height testing, jump rope counter, microsleep monitoring, and a tightness meter, among others, as showcased on the official website [105].

In terms of technology, the earbud itself is a multi-sensory device equipped with a 9-axis inertial measurement unit (IMU) consisting of an accelerometer, a gyroscope, and a magnetometer; an ear canal pressure sensor, a temperature sensor, an inward-facing ultrasound microphone as well as a speaker, a push button, and a controllable RGB color LED. The highest sampling rate available for IMU data is 50 Hz, which is also the frequency used for this work.
The device has **Arduino Nano 33 BLE Sense** as its microcontroller - a board suitable for wearable applications that make use of Bluetooth connectivity, integrate multiple sensors, and are required to have low power consumption [80]. The memory capacity of the Arduino Nano 33 BLE Sense is as follows:

- 1 MB of flash memory for storing the program code.
- 256 KB of SRAM for running the application and temporary data storage.

So far, no datasets collected with the said device have been made available to the public to the knowledge of the authors at the time of writing this work. This work addresses this by providing the first publicly accessible dataset collected on an OpenEarable earbud, in an attempt to stimulate further research with said device. The data collection procedure is described in Subsection 3.1.2.

3.1.2 Experimental Setup

A total of **30** subjects took part in the experiment (11F, 19M, aged between 18 and 30). They were instructed to wear comfortable attire. No compensation for participating was offered. The tests took place in several fitness centers in Enschede, Rotterdam, and Utrecht, the Netherlands. As a preparation for each experiment, the device was fully charged. The device was mounted behind the right ear of the participant. It was ensured that there were no distractions during activity data collection, and the process was restarted when the earable device became loose or whenever any unforeseen circumstances arose.

Participants were asked to walk, run, and climb stairs at their normal pace (average and paced), using a treadmill and a stair master. The exercises were demonstrated to the participants beforehand to ensure familiarity with each exercise. Each exercise lasted a total of 4 minutes, wherein the participant performed each movement for 2 minutes at low intensity, and 2 minutes at high intensity. The total duration of the experiment consisted of 12 minutes of active work, yet in-between exercises participants were able to decide how to rest and for how long, with the sensor staying mounted to the ear but not collecting data. For this reason, the total duration of the experiment varied per person. The exercise session length was chosen to be long enough to yield good quality signal recordings while not being harmful to the participant. The intensities at which each exercise was performed were selected by the participants so that they could remain representative of their habits. The transition from walking to jogging was achieved by the subject increasing the speed 0.5 km/h at a time until they felt more comfortable running than walking, done similarly to [16]. The protocol was run from low to high intensity per exercise to avoid the impact of high-intensity activities on low-intensity ones. For consistency, the treadmill incline was kept the same across all participants. Figure 3.1 summarizes the data collection protocol.

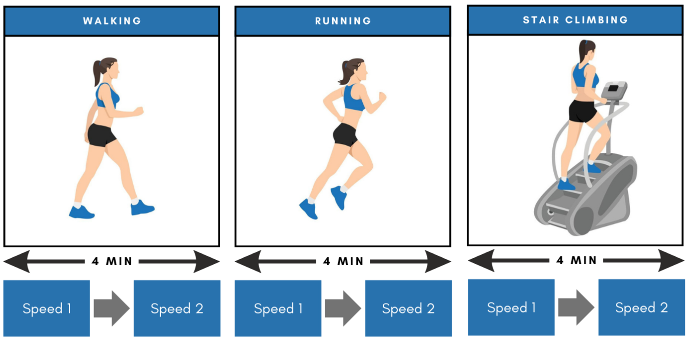


FIGURE 3.1: Data Collection Protocol.

3.1.3 Data Management

To record the data, the sensor needs to be connected via Bluetooth to the desktop dashboard. Starting and ending recordings is also controlled through the dashboard, and the raw data is ready for download from the internal server. No personally identifiable information such as name and age was collected from the participants. Each recording session was initially anonymized with a combination of an alphanumeric format of P#, herein Participant Identifier P1-P30, the name of the exercise, and the intensity (e.g. `P1-stairs-slow.csv`). Later on, recordings were grouped per activity folder. In each folder, each participant's raw data obtained from the earbud during data collection is kept. Each .csv file contains the following columns and content:

- **time**: Timestamp in UNIX format, measured in milliseconds
- **sensor_accX[g]**: X-axis of the accelerometer sensor in g-force units
- **sensor_accY[g]**: Y-axis of the accelerometer sensor in g-force units
- **sensor_accZ[g]**: Z-axis of the accelerometer sensor in g-force units
- **sensor_gyroX[°/s]**: X-axis of the gyroscope sensor in degrees per second
- **sensor_gyroY[°/s]**: Y-axis of the gyroscope sensor in degrees per second
- **sensor_gyroZ[°/s]**: Z-axis of the gyroscope sensor in degrees per second
- **sensor_magX[μT]**, **sensor_magY[μT]**, **sensor_magZ[μT]**: X-axis, Y-axis, Z-axis magnetometer sensor data in microteslas, which will not be used in this study but will be made publicly available along with the rest of the data.

Figure 3.2 displays an example accelerometer and gyroscope raw sensor segment recorded during a test run at high intensity of the same participant. To avoid ambiguity and difficulties in identifying the start and end frame of each exercise, the recordings were started only once the participant had begun performing the exercise and had confirmed that they had found their comfortable speed per intensity.

3.2 Data Preprocessing

Since the data obtained by the OpenEarable device is in raw signal form, it is necessary to perform several types of preprocessing to improve its quality. These include frame-based gait segmentation, feature extraction, feature selection, data normalization, and (optionally) data

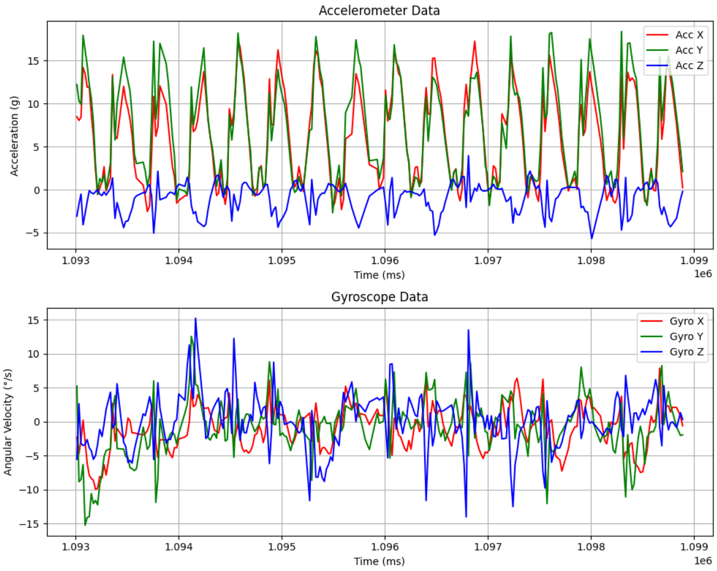


FIGURE 3.2: Acc and Gyro data segment from a high-intensity run of a test user.

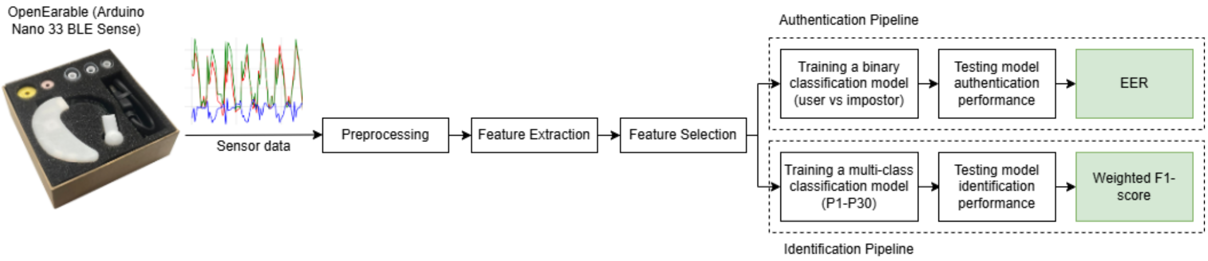

OpenEarable (Arduino Nano 33 BLE Sense)
Sensor data
Preprocessing
Feature Extraction
Feature Selection
Authentication Pipeline
Training a binary classification model (user vs impostor)
Testing model authentication performance
EER
Training a multi-class classification model (P1-P30)
Testing model identification performance
Weighted F1-score
Identification Pipeline

FIGURE 3.3: Architecture of the proposed recognition system

balancing. This section details the methodology applied to preprocess and segment the raw IMU data, extract meaningful features, and select the most relevant ones for ML-based classification. For DL-based classification, no manual feature selection is executed. Each step is critical in transforming the raw sensor signals into a feature representation suitable for accurate and efficient identification and authentication. Figure 3.3 depicts the high-level process.

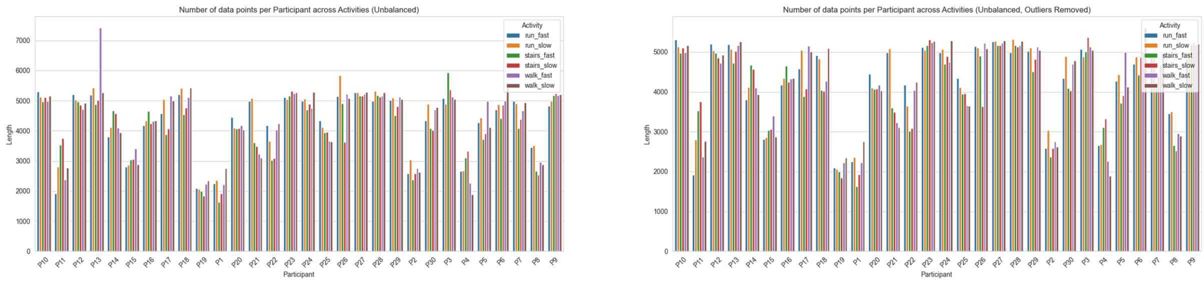
3.2.1 Data Segmentation and Outlier Removal

An authentication system should achieve high accuracy while minimizing recognition latency and having low computational complexity. To achieve this, in this work, sensor readings are segmented into fixed-length frames. The authors believe that, as gait cycle-based approaches are sensitive to gait cycle starting point detection precision, choosing a frame-based classification ensures robustness to variations in cycle start positions, and reduces bias.

This work utilizes **frames of fixed length of 2 seconds**, with the size chosen in a way that ensures that each frame segment contains at least one gait cycle so that the complete set of features exist in each period. Furthermore, a fixed-size sliding window (FSW) approach is adopted, wherein **a 50% overlap** is applied consistently across all windows. At 50Hz frequency, each window contains 100 time steps, or 2 seconds worth of recording, and the dataset has a mean of 83.94 windows per participant.

Upon post-recording inspection, it was discovered that some participants have consistently fewer data points than the mean number of data points across all six exercises and speed combinations. This is seen in Figure 3.4a, namely in `P11, P15, P19, P1, P2, P4, P6`. The observed discrepancy in the number of data points between participants appears to possibly stem from the recordings of said individuals been captured at a sampling rate lower than 50 Hz sampling rate to hardware or software limitations not evident at the time of recording that fall outside the scope of this work. Despite this, upon visual inspection, it seemed the essential gait patterns remain well-preserved, allowing for meaningful feature extraction. The reduced data points primarily affect the dataset's size rather than its quality.

Another observation in Figure 3.4a is the presence of evident outliers, where some participants' recordings are significantly longer than the mean number of data points for that exercise (e.g., `P13 walk_fast` recording, `P3 stairs_fast` recording). Although recording was initiated only after the participant began with performing the exercise, and continued for a fixed duration of two minutes per exercise, some recordings managed to capture additional data points both before the exercise started and after it ended. To address this, thresholds based on low variability in data points (indicating inactivity) were applied to identify and remove these sections at the start and end of the recordings. As a result, Figure 3.4b shows a dataset that is slightly more balanced across activities. No further adjustments are performed on the raw data.



(A) Dataset (Imbalanced) without Outlier Removal

(B) Dataset (Imbalanced) with Outlier Removal

FIGURE 3.4: Impact of Outlier Removal on the Dataset.

3.2.2 Feature Extraction

The next step consisted of extracting features from each frame that describe the gait cycle in a way that allows comparison between two windows. Each sensor used for data collection (accelerometer, gyroscope) provided three components: x, y, and z (i.e., three time-series signals). Next to that, a fourth signal, i.e., the magnitude, was computed. The magnitude vector v_i for each instant of time i is defined in 3.1, namely:

$$v_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{3.1}$$

where x_i, y_i, and z_i are the acceleration values along the three axes at time i. It is a representation of the 3-axes and its main advantage is its invariance with respect to the sensor orientation [72]. In this work, it is considered the fourth component.

There are a variety of feature vectors that can be extracted from time and frequency domains. In this work, features are selected according to best practices in frame-based feature representation methods for gait-based recognition, namely by extracting the most widely used accelerometer and gyroscope feature categories for this purpose, inspired by previous works [16, 58, 72], and thoroughly summarized in [112], `Table 6`. The specific features include arithmetic mean, standard deviation, range, energy, spectral entropy, bandpower, median frequency, Inter Quartile Range (IQR), 16-bin histogram; mutual information, correlation (Pearson correlation between the axes). Pair-wise Dynamic Time Warping (DTW) distance was also among the most commonly derived features according to previous work yet it can be expensive if done for every window, and as the purpose of this algorithm is also efficiency, it was skipped. The selected features encompass a wide range of statistical, spectral, and relational properties of both the accelerometer and gyroscope signals, and have been chosen for their low computational cost and high discriminatory power [52].

The complete list of features along with their definitions is included in Table 3.1.

TABLE 3.1: Accelerometer and Gyroscope Features

No.	Feature	Feature Description
1	Mean	The arithmetic mean of the x, y, and z components, as well as the magnitude, over a given time window.
2	Standard deviation	The standard deviation within x, y, z-axis, and m within a time window.
3	Range	The absolute difference between the maximum and minimum recorded values within a time window for each axis and the magnitude.
4	Energy	The total energy of the signal computed for each of the x, y, and z axes, as well as the magnitude. It captures the intensity of motion.
5	Spectral entropy	Represents the complexity of the signal by analyzing the power distribution across the frequency spectrum. It is based on Shannon's entropy and power spectral density.
6	Bandpower	Computes the average power within a specific frequency range (from 0 to Fs/2), giving insight into the signal's energy distribution. It depends on the signal's sampling rate.
7	Median frequency	The frequency that divides the power spectrum into two equal halves, indicating the balance point of energy distribution. It depends on the signal's sampling rate.
8	Inter Quartile Range (IQR)	The statistical spread of the signal, calculated as the difference between the third quartile (Q3) and the first quartile (Q1).
9	Histogram (16 bins)	Divides the signal values into 16 discrete bins and records the number of occurrences within each, capturing the signal's distribution.
10	Mutual information	Quantifies the dependency between pairs of signals, reflecting how much information one variable shares with another.
11	Correlation	Computes the Pearson correlation coefficient between pairs of signals to measure their linear relationship.

The feature selection process resulted in extracting a total of **210** features per window of data, combining features from both the accelerometer and gyroscope signals in the X, Y, and Z axes, plus their magnitudes.

After extracting the features and before any preprocessing steps, the dataset was partitioned into training and testing subsets using an `80:20` split. The 80% training set was used for further ML-related preprocessing and model training, while the 20% test set was kept completely separate to provide an unbiased evaluation of the final model. Importantly, all subsequent preprocessing steps, such as normalization, feature selection, and feature reduction, were performed exclusively on the training data to avoid data leakage and to ensure the test set remained a true representation of unseen data.

3.2.3 Feature Selection

It is important to note that the following feature selection process is only relevant to the machine learning (ML) models described in this study.

The deep learning (DL) models, particularly the CNN-LSTM architecture used in this work and adopted from [17] (called **CNN_fixLSTM** for the remainder of this work), learn feature representations automatically from raw sensor data and do not rely on manually selected feature sets. Therefore, feature selection is applied exclusively to the feature vectors used in the ML-based pipelines, ensuring dimensionality reduction and improved computational efficiency for these models. Data segmentation in overlapping windows is applied to both.

Before applying any feature selection, a MinMaxScaler is used as a normalization technique to scale all features in the training set to a range of $[-1, 1]$, similarly to [119]. The said range is preferred for datasets where the original features are both positive and negative and transforms the values to be centered around zero, which can help some ML models converge faster and perform better. The already-fitted scaler is later applied to the test set. This step is crucial before applying a correlation filter because scaling ensures that all features contribute equally to the correlation computation, preventing features with larger numeric ranges from disproportionately influencing the results.

The initial reduction on the 210-feature set was based on excluding highly intercorrelated variables by applying a correlation threshold filter. This was done activity-wise instead of dataset-wise, as this ensures that the retained features are tailored to the unique characteristics of each movement and prevents important features from being overlooked due to their low relevance in other activities (e.g., fast-paced running may rely more on high-frequency accelerations and consistent rhythmic patterns, while stair climbing may highlight variability in step timing). A correlation filter with a threshold of $|r| > 0.9$ was applied to remove highly collinear features, ensuring that the retained features are not near-linearly dependent on each other. Unlike more aggressive collinearity removal approaches like $|r| > 0.5$ in [108], this threshold aims to strike a balance between feature reduction and information preservation in the absence of medical expertise or extensive domain knowledge.

Once the features have been filtered for correlation, Mutual Information (MI) [58] was used as the final step in the feature selection process to retain the most informative features for classification. Mutual Information-based selection quantifies the dependency between each feature and the class labels, ensuring that only features with the highest discriminative power are retained. Unlike tree-based methods, MI-based selection does not assume linearity or feature independence, making it well-suited for diverse feature distributions.

By selecting the top 20 features within-activity, the method achieves a balance between retaining discriminative information and reducing computational complexity. Mutual Information is inherently robust to feature redundancy and provides an interpretable ranking of features based on their relevance to the classification task. This ensures that the retained feature set captures the most distinguishing characteristics of gait patterns without introducing unnecessary complexity.

Several features, such as standard deviation and mean of the accelerometer axes, are consistently identified as important across all activities. These features suggest that variations and averages in accelerometer data (primarily in the X and Z axes) play a significant role in distinguishing between activities and participants. Gyroscope histogram features, particularly those associated with the Z-axis (e.g., Histogram Bin 7, Histogram Bin 15), frequently appear among the top features. Lastly, the prominence of Mean (Acc Y) in stairs-related activities indicates the importance of vertical motion in differentiating this task from walking and running where the frequent selection of Mean (Acc X) shows forward motion is a dominant factor.

The reduced training dataset, now consisting of the top 20 features, is used to train multiple machine learning classifiers wherein each model is evaluated on the unseen test set using the same scaling and feature subset. The implementation details are found in Subsection 3.3.1. Additionally, the models are trained with the full set of features to measure and compare possible effects on size, computation time, and performance.

3.3 Classifiers, Validation Methods, Performance Metrics

This section details the Machine Learning and Deep Learning models employed for gait-based recognition, and the metrics used to measure their performance.

For the **Identification Component**, the labels are the ID numbers of the users, and a model is ran within activity category, resulting in six models per configuration, and across-activity, resulting in one model per configuration.
For validation, each model undergoes a single 80:20 train-test split, ensuring that participant data is distributed consistently across training and testing subsets. This split is applied once per activity for within-activity identification and once for the across-activity scenario. The two main **identification** scenarios explored in this work are as follows:

1. **Within-activity identification:** Training and evaluate each model separately on individual activities (e.g., `walk_slow, stairs_slow`) to test how well users can be recognized when performing the *same* type of activity during training and inference.

2. **Across-activity identification:** Combine data from multiple activities into one dataset, so the model learns a more general representation of a user's gait across different movement types.

For the **Authentication Component**, the labels are the ID numbers of the users, and a model is ran 30 times within activity category as each of the 30 users is treated as the genuine user in turn, while the remaining 29 participants serve as impostors. This process is repeated for each of the six activity categories, meaning that 30 authentication models are trained per activity. This results in a total of 180 models per configuration. For across-activity authentication, each of the 30 users is evaluated separately, resulting in an additional 30 authentication models.
Because the dataset is naturally imbalanced, with only one genuine user versus 29 impostors per training iteration, the Synthetic Minority Over-sampling Technique (SMOTE) is applied to generate additional artificial genuine-user (train) samples. displayed in tab creates these synthetic samples by selecting existing data points from the minority genuine user class and generating new points along the line segments joining them with their k-nearest neighbors. This is done to mitigate bias towards the impostor classifications.

The two main **authentication** scenarios explored in this work are as follows:

1. **Within-activity authentication:** Train and evaluate the model separately on individual activities (e.g., `walk_slow` or `stairs_slow`). This scenario tests how reliably the model can authenticate a user when the enrollment (training) and verification (inference) processes involve the *same* type of activity.

2. **Across-activity authentication:** Combine data from multiple activities into a single dataset, allowing the model to learn a more generalized representation of a user's gait that is robust to different movement types. In this scenario, the model is trained on a mixture of activities, and authentication attempts can involve any activity from that set, testing how well the system handles cross-activity variability.

The same set of models are employed for both components; however, their evaluation metrics are tailored to align with the specific objectives of each task. The subsequent Subsections 3.3.1 and 3.3.2 describe the training pipeline and evaluation metrics for each component and model type combination.

3.3.1 Machine Learning Models

Two main pipelines are implemented to address the tasks of gait-based identification and gait-based authentication. Both pipelines rely on a shared set of supervised machine learning (ML) models—Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), and Multi-Layer Perceptron (MLP)—together with systematic hyperparameter tuning.

After handcrafted features are extracted from raw sensor data, highly correlated features (`Pearson correlation` > 0.9) are removed to reduce redundancy. A Min-Max scaler then normalizes remaining features within a $[-1, 1]$ range. Mutual information is used to select the top 20 most informative features to the target classification (i.e., user identity or genuine-impostor). Each of the four classifiers (RF, SVM, kNN, MLP) undergoes a **Grid Search** over a pre-defined hyperparameter range (in Table 3.2). Three-fold cross-validation (`CV=3`) is employed to prevent overfitting and to identify optimal parameter configurations. Once the best hyperparameters are found, the final model is retrained on the full available training data. A time limit (`100 ms`) is enforced to maintain practical performance for real-time resource-constrained settings. Performance metrics are computed on a held-out test set to measure each model's effectiveness. The top-performing models are serialized (`pickle` format) for potential deployment.

Table 3.2: Hyperparameter grid search ranges for each classifier

Classifier	**Hyperparameters**

Random Forest (RF)	Number of trees (`n_estimators`): {10, 50, 100}
	Maximum tree depth (`max_depth`): {5, 10, None}
	Minimum samples to split (`min_samples_split`): {2, 5}
	Minimum samples per leaf (`min_samples_leaf`): {1, 2}

Support Vector Machine (SVM)	Regularization parameter (`C`): {0.1, 1, 10}
	Kernel type: {linear, rbf}
	Kernel coefficient (`gamma`): {scale, auto}

k-Nearest Neighbors (kNN)	Number of neighbors (`n_neighbors`): {3, 5, 7, 9}
	Weight function: {uniform, distance}
	Distance metric: {euclidean, manhattan}

Multi-Layer Perceptron (MLP)	Hidden layer sizes (`hidden_layer_sizes`): {(10,), (20,), (50,)}
	Activation function: {relu, tanh}
	Regularization term (`alpha`): {1e-4, 1e-3, 1e-2}
	Solver for weight optimization: {adam, sgd}

For **gait-based user identification**, the system must recognize each participant as one of the known classes (*multi-class classification*). Two settings, as mentioned previously, are evaluated: within-activity identification and across-activity identificaiton. In each case, the dataset is split using an 80:20 train-test ratio, ensuring class-stratified sampling. Feature preprocessing and selection follow the common pipeline mentioned. RF, SVM, kNN, and MLP classifiers are then trained using Grid Search CV. Final evaluation is based on accuracy, weighted F1 score, and confusion matrices, and training time and model size are recorded.

In **gait-based user authentication**, the goal is to verify a claimed user identity (*genuine* vs. *impostor*). Each participant is treated as the genuine class, while all others form the impostor class. Two settings are examined: within-activity authentication and across-activity authentication. Following the same feature processing, displayed in tab is applied to balance the genuine vs. impostor classes where necessary. Model hyperparameters are optimized via the same Grid Search CV strategy. The Equal Error Rate (EER) is reported to measure authentication performance, along with training time and model size for each classifier–activity combination. Final models are similarly stored in `pickle` format.

This approach enables direct comparisons of model performance across the different activities, tasks, and classification algorithms.

3.3.2 Deep Learning Models

This section describes the deep learning-based pipelines used to perform both **gait-based identification** and **gait-based authentication**. Two primary architectures, **LSTMOnly** and **CNN_fixLSTM**, are evaluated under both within-activity and across-activity scenarios.

The architecture of the LSTM and CNN_fixLSTM models is adapted from the work of Cao et al. [17], who introduced a hybrid deep learning approach combining CNN and LSTM for gait-based user authentication using smartphone-collected IMU signals. LSTM is used as a feature extractor for temporal gait patterns. A distinguishing feature of this implementation is the freezing of the LSTM parameters during training. This design choice is made to leverage the pre-trained sequential feature extraction capability of the LSTM while allowing the CNN branch to adapt to domain-specific characteristics in the dataset. Freezing LSTM also means fewer parameters are updated, leading to faster training and lower memory usage, which aligns with the goals of this work. Further details about the design choices surrounding the said models can be found in the referred paper.

In this work's implementation, the key architectural components are retained while several modifications are made in terms of CNN depth to align the model with the size of the collected dataset and computational constraints, and the model is adapted to identification tasks as well. Notably, while the original paper makes use of a random grid hyperparameter selection followed by a hand-tuning method to determine the final hyperparameters for the LSTM part, no hyperparameter tuning is performed in this work due to resource constraints and lack of domain knowledge. This limitation can be addressed in future work.

For each activity category (e.g., `walk_slow`, `stairs_slow`, ...), training and testing data are split (80:20), with each participant's data distributed appropriately (identification) or split according to genuine vs. impostor (authentication).
In the **identification** case, a separate model is trained per activity (6 total activities).
In the **authentication** case, the model is trained 30 times for each activity, once per user, treating that user's data as "genuine" and all others' as "impostor." Due to the resulting class imbalance (1 genuine vs. 29 impostors), SMOTE oversampling is applied to the genuine class in the training set.

The same across-activity and within-activity settings apply as in Subsection 3.3.1. The two architectures, summarized in Table 3.3 (**LSTMOnly**) and Table 3.4 (**CNN_fixLSTM**), are used for both identification and authentication tasks. These architectures follow a hybrid deep learning approach for gait-based user identification and authentication, inspired by the one utilized in [17], and adapted to the specific needs of this dataset.

Layer	**Details**	**Output Shape**
Input	–	$(B, 100, 6)$

`LSTM`	`hidden_size=64`
	`num_layers=2`
	`batch_first=True`
		$(B, 100, 64)$

Select last timestep	–	$(B, 64)$
`Linear`	64 → num_classes	$(B, \text{num\_classes})$

Table 3.3: Architecture of the `LSTMOnly` model. The input shape assumes a batch size B, a temporal dimension of 100 timesteps, and 6 sensor channels (3D accelerometer + 3D gyroscope).

The **LSTMOnly** model (Table 3.3) processes each window $\in R^{100\times 6}$ via a two-layer LSTM (`hidden_size` $= 64$). The hidden state from the final time step ($t = 100$) is passed to a linear layer that outputs either identification or authentication results. After training, the best LSTM checkpoint is saved for further use.

Layer	Details	Output Shape
	LSTM Branch (Frozen)	
Input (LSTM branch)	–	$(B, 100, 6)$

`LSTM`	`hidden_size=64`
	`num_layers=2`
	`batch_first=True`
	(frozen)
		$(B, 100, 64)$

Select last timestep	–	$(B, 64)$
	CNN Branch	
Input (CNN branch)	–	$(B, 6, 100)$
`Conv1d + ReLU`	$6 \to 32$, `kernel_size=8`	$(B, 32, 93)$
`MaxPool1d`	`kernel_size=2`	$(B, 32, 46)$
`Conv1d + ReLU`	$32 \to 64$, `kernel_size=5`	$(B, 64, 42)$
`MaxPool1d`	`kernel_size=2`	$(B, 64, 21)$
`Conv1d + ReLU`	$64 \to 64$, `kernel_size=3`	$(B, 64, 19)$
`AdaptiveAvgPool1d`	–	$(B, 64, 1)$
Squeeze	–	$(B, 64)$
	Fusion and Classifier	
Concat	*concatenate* (CNN branch, LSTM branch)	$(B, 128)$
`Linear`	128 → num_classes	$(B, \text{num\_classes})$

Table 3.4: Architecture of the `CNN_fixLSTM` model. The LSTM is identical to `LSTMOnly` but is frozen (no gradient updates). The last hidden state is extracted from the LSTM and gets concatenated with the final CNN features for classification.

The **CNN_fixLSTM** model (Table 3.4) enhances the LSTM by adding a parallel CNN branch:

1. A frozen LSTM branch (identical to LSTMOnly) extracts temporal dynamics. Its pre-trained weights (from the best LSTM checkpoint) are fixed (no gradient updates).

2. A CNN branch processes the same window in a `(channels, timesteps)` format $(B, 6, 100)$. A sequence of 1D convolutions, ReLU, and pooling layers yields a 64-dimensional latent feature via global average pooling.

3. The CNN feature vector is concatenated with the 64-dimensional LSTM hidden state, forming a 128-dimensional fused representation. Another linear layer then outputs the final logits (multi-class or binary).

During training, the CNN parameters and fusion layer are optimized, while the LSTM remains frozen. The training procedures are as follows:

1. **LSTM Pretraining.** The **LSTMOnly** network is first trained end-to-end for 10 epochs (cross-entropy loss, Adam optimizer, initial `lr` $= 10^{-3}$). For *within-activity* experiments, a separate LSTM is trained per activity (identification) or per user-activity pair (authentication), whereas in *across-activity* experiments, a single LSTM is trained on the combined dataset for each user or for all users (identification). The best checkpoint is chosen based on F1 (identification) or EER (authentication).

2. **CNN_fixLSTM Training.** Next, the pre-trained LSTM parameters are loaded and frozen; the CNN and final fusion layer are randomly initialized. Training proceeds with the same optimizer/hyperparameters. Since the LSTM is not updated, the CNN learns features complementary to the already extracted temporal embeddings. Best checkpoints are again selected using F1 or EER.

Each final model is saved in `pickle` format or via `torch.save` for potential deployment. Model size and training time are tracked to ensure compatibility with resource-constrained devices.

3.3.3 Performance Evaluation Metrics

For **user identifcation**, the performance metric used for performance evaluation is the F1 score. Precision measures the proportion of true positive predictions among all positive predictions made by the model. Recall measures the model's ability to identify all true positive samples correctly. The F1 score combines both of them into a single metric, representing their harmonic mean. The weighted F1 score is chosen as the primary metric as, since this work deals with a multi-class classifcation problem with the number of classes equal to the number of participants, for the classifer performance observing only a single F1-score suffices instead of having one for each participant class. The weighted F1 score is evaluated per activity, and across-activities. In addition to these performance metrics, training time is recorded to evaluate computational efficiency, and model size is measured to assess deployment feasibility for resource-constrained environments.

For **user authentication**, the performance metrics used for performance evaluation are the standard evaluation metrics in an authentication system [83, 115, 129], namely False Acceptance Rate (FAR), False Rejection Rate (FRR), and Equal Error Rate (EER). In literature, the naming conventions of FAR and FRR are interchangeable with FPR and FNR, respectively. Herein, FAR is the likelihood that an impostor is identified, or wrongly accepted, as the legitimate user. This reflects the system's security. On the other hand, FRR is the likelihood that the authentication system incorrectly rejects the legitimate user, marking them as an impostor. FAR and FRR are computed as follows:

$$FAR = \frac{FP}{FP + TN}, \quad FRR = \frac{FN}{FN + TP} \tag{3.2}$$

where TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively.

Finally, EER represents the point at which FAR and FRR are equal, and, similar to F1 for identification, is the primary metric used to compare the authentication models in this work. A smaller EER indicates higher overall authentication accuracy and a better balance between security and usability. In the described experiments, EER is reported while performing k-fold cross-validation as the test set contains observations from both negative and positive classes. Lastly, training time and model size are also measured.

It is noteworthy that authentication can also be viewed as impostor testing. As such, the environment is referred to as the Zero-effort attack environment, meaning the impostors do not intentionally try to mimic the Genuine user; and participants' regular gait patterns are used as impostor samples [58]. Testing for Imitators who make a deliberate attempt to copy the Genuine user falls outside of the scope of this work.

The outlined evaluation metrics provide a comprehensive framework to assess the identification and authentication systems' performance, and the models' efficiency and size.

3.3.4 Model Optimization

As the memory and computing capabilities of edge devices such as the one utilized in this work, namely the `Arduino Nano 33 BLE Sense`, are significantly limited compared to those of cloud and mobile devices, optimization techniques for making ML and DL models suitable for deployment on smaller devices are explored in this work. The 33BLE has limited RAM

(256KB) and flash storage (1MB). Throughout optimization, it is crucial to measure model size and ensure it fits within these constraints. Optimizations should also decrease inference time and power draw; yet, as the scope of this work does not entail real-time deployment on the actual device, said metrics cannot be tracked while running the models. The techniques this work utilizes are Feature Set Minimization (which is inherently embedded in the ML recognition pipelines used in this work) and two forms of quantization—**Post-Training Quantization (PTQ)** [50] and **Quantization-Aware Training (QAT)** [78]—for the DL models, as described in Subsection 2.7.

Quantized models are particularly beneficial on the BLE33, as its Cortex-M-based microcontroller supports efficient fixed-point operations and reduced-precision arithmetic [71], aligning well with int8 or int16 quantized networks. Each optimization step (feature set minimization, quantization) is followed by validation to ensure that classification accuracy remains within acceptable bounds. Slight accuracy loss herein is considered acceptable in exchange for significant gains in efficiency.

As already implemented, reducing the feature set the ML models utilize to the top 20 most discriminative features decreases computation and memory requirements as fewer input dimensions lead to smaller model sizes and faster inference. No further model optimization techniques beyond that and the afore-described grid search (3.2) are applied to the ML frameworks. The remaining techniques, namely the two types of quantization, are applied only to the DL within-activity and across-activity **identification** models, as noted in Subsection 4.3.
A single authentication model needs only to distinguish between one genuine user (the presumed device owner) and impostors. Therefore, it is inherently smaller in size than the models intended for identification, and, as it falls within reasonable size, will not be optimized in this work.

The optimization process for the **CNN_fixLSTM** model implements the following two techniques:

1) Post-Training Quantization (PTQ). Dynamic quantization is employed by converting weights in the **LSTM** layers (targeting `nn.Linear`, `nn.Conv1d`, and `nn.LSTM`) from 32-bit floating point to 8-bit integers (qint8). This reduces the model's memory footprint and accelerates inference on CPUs. CNN layers, however, mainly use convolutional operations, which dynamic quantization does not support well.

2) Quantization-Aware Training (QAT). Since CNN layers are not well-supported by dynamic PTQ, QAT is applied to the CNN portion of the **CNN_fixLSTM** model. During QAT, the LSTM branch remains frozen (keeping its pre-trained float32 weights), while the CNN layers are trained with simulated quantization effects to adapt them to lower precision operations. This ensures that CNN feature extraction remains effective despite reduced numerical precision.

Finally, the optimized models are exported to ONNX format to ensure compatibility with the environment Edge Impulse for future deployment [47].

Optimizing the CNN_fixLSTM architecture, despite its complexity compared to simpler shallow models, is justified by its capacity to capture both spatial patterns through convolutional layers and temporal dynamics via LSTM components inherent in sequential gait data, without requiring manual feature selection—providing a more adaptable product for deployment. Although the training time of shallow ML models is mostly relatively shorter than that of the DL architecture, the time to extract, prepare, and filter all necessary features to feed into said ML architectures should be accounted for (displayed in Table 4.1), leading to a comparable or even slower deployment timeline. This efficiency, combined with CNN_fixLSTM's superior ability to autonomously learn hierarchical representations, further validates the practicality of the framework despite its computational overhead.

Chapter 4

Results

This chapter presents the results of the user authentication and identification experiments conducted using both shallow machine learning (ML) algorithms and deep learning (DL) models, across-activities and within-activity. Results are presented separately for authentication and identification tasks to highlight the distinct challenges and performance nuances associated with each of them. In the presented tables, the DL model **CNN_fixLSTM** will be abbreviated as **fixLSTM**. The performance of the LSTM-only model will not be included in the evaluation process as it is considered only as part of the final DL model. The ML models' feature set will be indicated as **allFT** (all features) or **topK** (top 20 features).

For the sake of comparison between the selected models in terms of authentication performance, only the averaged weighted EER and model sizes are presented per activity and model combination. This allows for a concise evaluation of model performance without delving into participant-specific details. However, any such participant-specific abnormalities or notable performance patterns are pointed out separately, and the full tables are also included in the Appendix A.

Lastly, model optimization results of the **CNN_fixLSTM** model are discussed.

Each subsection delves into specific aspects of the results, providing detailed metrics, comparative analyses, and insights into model performance across different configurations.

Table 4.1 presents the time required (in seconds) for two critical preprocessing steps essential for both Deep Learning (DL) and Machine Learning (ML) models. The Segmentation Time column details the duration needed to segment raw sensor data into 2-second windows with 50% overlap, a preprocessing step required for both model types. The Feature Extraction Time column shows the time taken to compute the full feature set for all 30 users, a necessary step specifically for ML models. The third column indicates the average time it takes to segment the data of a single participant recording and extract the full set of features from it, yielding an average of around 1.5 minutes overhead for across-activity segmentation + feature extraction time per person. This factor is important to consider when evaluating the efficiency of ML-based approaches, as it directly impacts computational cost and real-time feasibility.

The feature set derived in this study is based on established best practices in the field. However, for future implementations, domain experts could refine the feature selection process to identify a minimal yet highly discriminative subset of features. This optimization would significantly reduce extraction time while maintaining high recognition accuracy.

Activity	Segmentation Time (s)	Feature Extraction Time (s)	Total Avg. per Person
run_fast	0.386	218.289	23.622
run_slow	0.517	232.990	11.156
stairs_fast	0.509	214.497	11.813
stairs_slow	0.358	219.924	11.876
walk_fast	0.467	229.205	22.326
walk_slow	0.373	226.284	25.228
Total	2.610	1341.189	

Table 4.1: Segmentation and Feature Extraction Time per Activity

4.1 Authentication Results

This section details the performance of both ML and DL models in the authentication task, evaluated under *within-activity* and *across-activity* scenarios. The evaluation focuses on Equal Error Rate (EER) as the primary metric for authentication effectiveness and model size (MB) as a key factor for deployment feasibility on resource-constrained devices. The results highlight how different architectures perform across activities and users, and how feature selection impacts authentication accuracy.

4.1.1 Within-Activity Authentication.

This subsection analyzes the performance of various machine learning (ML) and deep learning (DL) models in user authentication on a within-activity basis.

Herein, within-activity authentication performance is evaluated using averaged results per activity per `model + feature` combination rather than presenting results on a per-user basis as these results would constitute a table of `180` rows. Reporting per-user results would thus significantly increase the complexity of result presentation, making it difficult to extract high-level trends. By focusing on per-activity averages, it is more intuitive to interpret model performance trends and compare different algorithms without overwhelming the analysis with excessive detail. Reporting per-user results may furthermore introduce unnecessary variability due to individual differences in biometric patterns, sensor noise, or session variability. Averaging results per activity ensures that conclusions drawn are representative of the authentication model's behavior rather than specific to individual participants. However, the said thorough results are available upon request.

Table 4.2 presents the averaged authentication results per activity for all ML and DL models. The values are aggregated across all participants performing each activity. The table includes the Equal Error Rate (EER) (%), which reflects the authentication accuracy, and the model size (MB). The top three models'results in terms of smallest size, and in terms of lowest EER, have been presented in bold.

One important observation is that, as topK features were derived per-activity basis, and not per-user basis, there is a significant increase in EER when comparing the performance of ML models from allFT to topK. **This suggests that the feature selection process may have removed important discriminatory features for individual authentication.** While this was done in this work to reduce computation costs, in a real-life scenario the features derived would be tailored to the one true user that the device belongs to.

Furthermore, to counter the increase in EER in the case where only one user would be considered the genuine one, more computationally heavy methods for feature selection like Recursive Feature Elimination (RFE) could be applied. As this work required the training of `180` models

Model + Set of Features + EER (in %)									
Activity	fixLSTM	MLP_AllFT	MLP_TopK	RF_AllFT	RF_TopK	SVM_AllFT	SVM_TopK	kNN_AllFT	kNN_TopK
Run_fast	0.09%	0.73%	5.25%	0.25%	2.68%	0.06%	2.91%	0.13%	2.96%
Run_slow	0.13%	1.09%	7.16%	0.13%	4.50%	0.05%	4.59%	0.13%	4.89%
Stairs_fast	2.45%	3.44%	15.92%	0.87%	12.68%	1.39%	14.89%	1.17%	14.39%
Stairs_slow	4.22%	4.57%	17.84%	1.58%	14.82%	2.12%	17.73%	2.27%	16.46%
Walk_fast	1.31%	1.51%	7.47%	0.24%	4.52%	0.27%	5.44%	0.38%	5.69%
Walk_slow	3.55%	3.32%	12.91%	0.94%	9.95%	1.30%	12.43%	1.23%	11.74%
Average	**1.96%**	2.44%	11.09%	**0.67%**	8.19%	**0.87%**	9.66%	0.89%	9.36%
Model + Set of Features + Model Size (in MB)									
Activity	fixLSTM	MLP_AllFT	MLP_TopK	RF_AllFT	RF_topK	SVM_AllFT	SVM_TopK	kNN_AllFT	kNN_TopK
Run_fast	0.05	0.21	0.03	0.49	0.85	0.17	0.03	6.69	0.78
Run_slow	0.05	0.21	0.03	0.54	1.03	0.19	0.03	6.7	0.79
Stairs_fast	0.05	0.22	0.03	0.67	1.84	0.27	0.05	6.55	0.74
Stairs_slow	0.05	0.22	0.03	0.78	2.01	0.3	0.06	6.57	0.74
Walk_fast	0.05	0.22	0.03	0.53	0.98	0.2	0.03	6.68	0.77
Walk_slow	0.05	0.22	0.03	0.69	1.72	0.26	0.05	6.82	0.78
Average	**0.05**	0.22	**0.03**	0.62	1.41	0.23	**0.04**	6.67	0.77

TABLE 4.2: Equal Error Rates (EER) (in %) and Model Sizes (in MB) for All ML and DL Models in User Authentication Within Activities (Averaged).

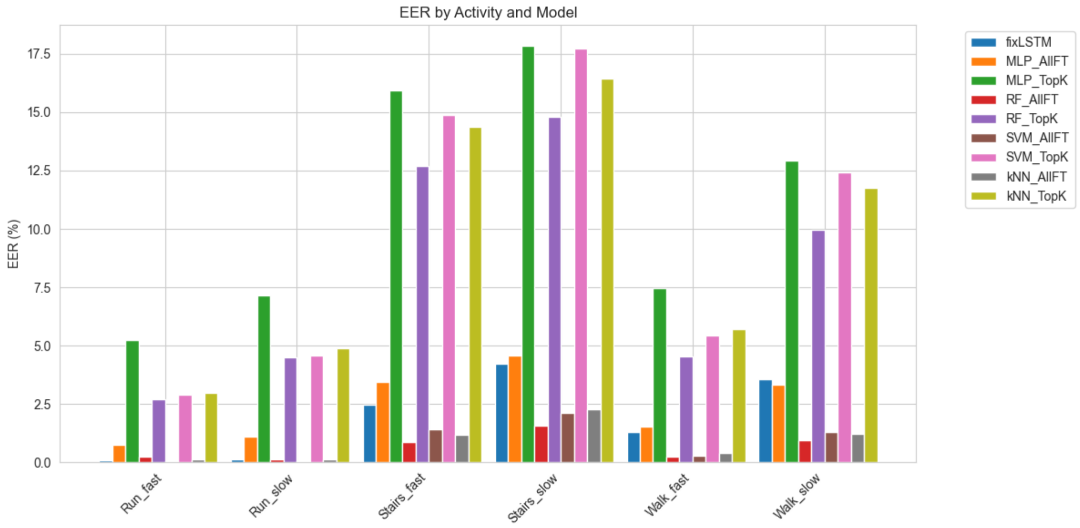


FIGURE 4.1: Detailed Grouped Bar Charts Across Activities and Feature Sets

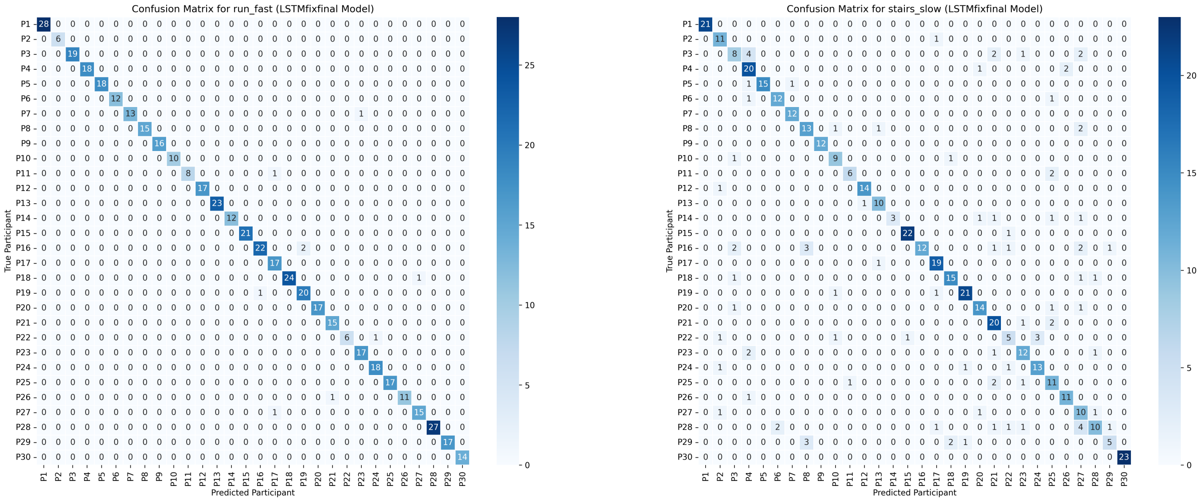
for each configuration, it was not within the resource availability of the device used to run such techniques at this scale.

RF and SVM experience the largest increase in size when comparng their topK to allFT sizes. To visualize the difference between feature sets, Figure 4.1 is provided as support.

As a general observation, it seems that **running-based activities (`Run_fast`, `Run_slow`) show the lowest EERs, indicating that these activities provide more consistent biometric patterns for authentication that allow for distinguishing participants clearly from one another**, followed by `walk_fast`. On the other hand, `stairs_slow` seems to be the activity hardest to authenticate participants on.

To visualize the worst (4.2a) and best (4.2b) activity for within-activity user authentication, the following confusion matrices generated by **CNN_fixLSTM** are provided in Figure 4.2. It seems that albeit the significant difference in performance, participant `P1` seems to have highly distinguishable gait patterns.

To generalize, **movements that are high-paced and involve more head movement**



(A) CNN_fixLSTM run_fast results

(B) CNN_fixLSTM stairs_slow results

FIGURE 4.2: Comparison of Best and Worst Wihin-Activity User Authentication Confusion Matrices

seem to be the most distinctive. Furthermore, it is to derive that **CNN_fixLSTM**, **RF_allFT**, and **SVM_allFT** are the three models with highest performance averaged across the six activities. When comparing on Model Size, kNN_AllFT is the largest model, averaging 6.67 MB, indicating high memory requirements despite its competitive EER, while fixLSTM, MLP_topK, and SVM_topK are the smallest. Figure 4.3 represents the trade-off between EER and Model Size visually, and allows for examining where each model stands in terms of the trade-off between a lower error rate and smaller model size. It is clear that the clustered four models in the left-most lower corner constitute the best ratio.

The CNN_fixLSTM model achieves the best trade-off between Equal Error Rate (EER) and model size, maintaining a relatively low EER of 1.96% while being relatively compact compared to other models. Similarly, RF_allFT also demonstrates strong performance; however, it relies on a feature extraction pipeline that introduces additional computational overhead. As shown in Table 4.1, deriving the full set of 210 features per participant and activity notably increases preprocessing time—a constraint that CNN_fixLSTM bypasses by learning features directly from raw data. Given the priority of efficiency on resource-constrained devices, CNN_fixLSTM presents a more streamlined execution with fewer preprocessing steps, making it a more suitable candidate for deployment.

4.1.2 Across-Activity Authentication.

This subsection analyzes the performance of various machine learning (ML) and deep learning (DL) models in user authentication across different individuals.

A thorough per-participant reporting of results can be found in Table A.1 in the Appendix A.

Below, Table 4.4 provides a results overview in the form of a heatmap as a concise way to see where each `model + feature` combination excels or struggles for each participant. The shades of green indicate low EER (high authentication performance), while the red shades display poor authentication abilities.

It is evident that, as in Subsection 4.1.1, the worst performing authentication models are those that make use of the reduced set of features, for similar reasons as described in the previous section, namely suggesting that some critical features necessary for participant authentication

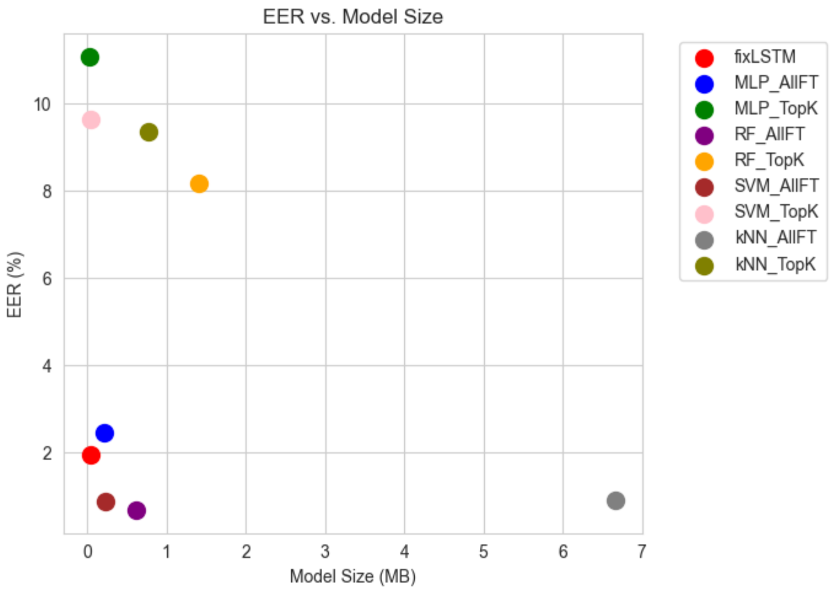


FIGURE 4.3: Trade-Off Plot: Average EER vs Average Model Size (Within-Activity Authentication)

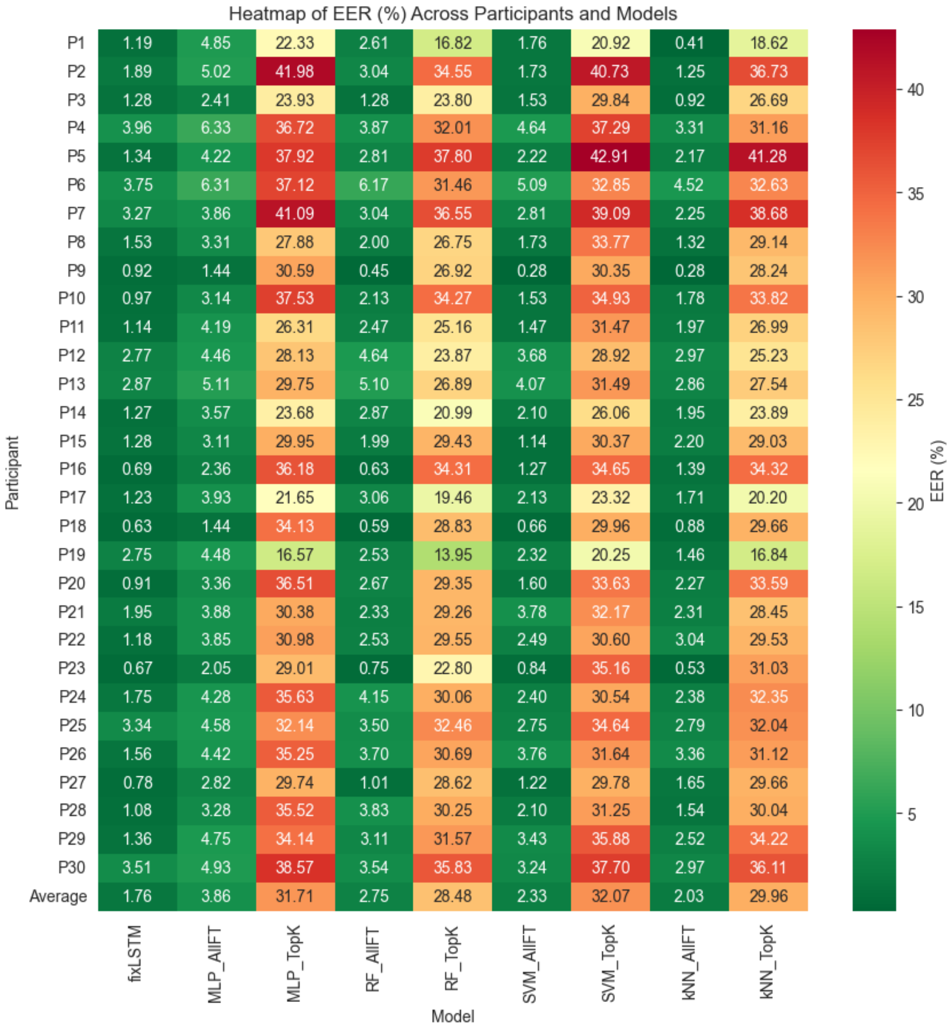
were removed, leading to worse classification performance. Again, more thorough person-specific feature selection is necessary in order to counteract the low performance.

Furthermore, **CNN_fixLSTM** seems to outperform all other considered models, with an average EER of **1.76%**. The low EER values indicate that it generalizes better across different users, meaning it is able to effectively separate individuals based on unique movement signatures it extracts from raw sensor data.

To look at **CNN_fixLSTM**'s performance in detail, a comparison of EER-based heatmaps is made between its worst and best authenticated participants, namely P18 (best) and P4 (worst), presented in Figure 4.5a, and Figure 4.5b, subsequently.

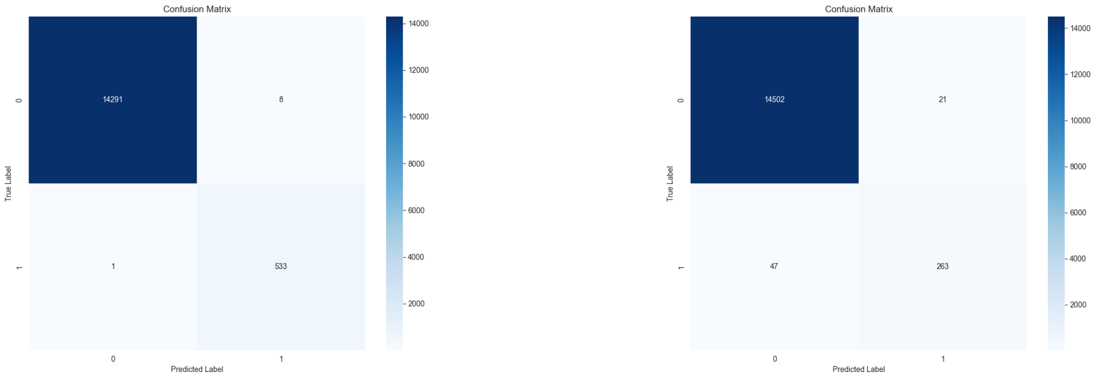
When talking about user authentication to grand device access, a critical part to security to focus on is the number of False Positives, indicating the case of granting access to an impostor, so minimizing FP should be the main focus in that context. On both parts of the figure, albeit the difference in EER performance, it can be seen that the number of FP are still the lowest the your confusion matrix (only 8 and 21 cases). This means that the system is highly secure in terms of preventing unauthorized access. The number of False Negatives (FN), or the number of cases where the legitimate user has been denied access, negatively influence usability levels of the system, yet seem to herein also remain relatively low. Security is prioritized over usability, which is often desirable for high-stakes authentication systems (e.g., banking, or secure devices).

Figure 4.6 represents the trade-off between EER and Model Size visually, and allows for examining where each model stands in terms of the trade-off between a lower error rate and smaller model size. It is clear that the clustered four models in the left-most lower corner constitute the best ratio, among which the best-performing model remains **CNN_fixLSTM**. With an average size of 0.05MB (from Appendix A.1), it is also the smallest model of the batch, making it immediately deployable on resource-constrained devices. kNN models seem to require significantly higher storage (≈ 38MB for **kNN_AllFT**), deeming them unsuitable for edge device deployment at least within this use case.



Participant	fixLSTM	MLP_AllFT	MLP_TopK	RF_AllFT	RF_TopK	SVM_AllFT	SVM_TopK	kNN_AllFT	kNN_TopK
P1	1.19	4.85	22.33	2.61	16.82	1.76	20.92	0.41	18.62
P2	1.89	5.02	41.98	3.04	34.55	1.73	40.73	1.25	36.73
P3	1.28	2.41	23.93	1.28	23.80	1.53	29.84	0.92	26.69
P4	3.96	6.33	36.72	3.87	32.01	4.64	37.29	3.31	31.16
P5	1.34	4.22	37.92	2.81	37.80	2.22	42.91	2.17	41.28
P6	3.75	6.31	37.12	6.17	31.46	5.09	32.85	4.52	32.63
P7	3.27	3.86	41.09	3.04	36.55	2.81	39.09	2.25	38.68
P8	1.53	3.31	27.88	2.00	26.75	1.73	33.77	1.32	29.14
P9	0.92	1.44	30.59	0.45	26.92	0.28	30.35	0.28	28.24
P10	0.97	3.14	37.53	2.13	34.27	1.53	34.93	1.78	33.82
P11	1.14	4.19	26.31	2.47	25.16	1.47	31.47	1.97	26.99
P12	2.77	4.46	28.13	4.64	23.87	3.68	28.92	2.97	25.23
P13	2.87	5.11	29.75	5.10	26.89	4.07	31.49	2.86	27.54
P14	1.27	3.57	23.68	2.87	20.99	2.10	26.06	1.95	23.89
P15	1.28	3.11	29.95	1.99	29.43	1.14	30.37	2.20	29.03
P16	0.69	2.36	36.18	0.63	34.31	1.27	34.65	1.39	34.32
P17	1.23	3.93	21.65	3.06	19.46	2.13	23.32	1.71	20.20
P18	0.63	1.44	34.13	0.59	28.83	0.66	29.96	0.88	29.66
P19	2.75	4.48	16.57	2.53	13.95	2.32	20.25	1.46	16.84
P20	0.91	3.36	36.51	2.67	29.35	1.60	33.63	2.27	33.59
P21	1.95	3.88	30.38	2.33	29.26	3.78	32.17	2.31	28.45
P22	1.18	3.85	30.98	2.53	29.55	2.49	30.60	3.04	29.53
P23	0.67	2.05	29.01	0.75	22.80	0.84	35.16	0.53	31.03
P24	1.75	4.28	35.63	4.15	30.06	2.40	30.54	2.38	32.35
P25	3.34	4.58	32.14	3.50	32.46	2.75	34.64	2.79	32.04
P26	1.56	4.42	35.25	3.70	30.69	3.76	31.64	3.36	31.12
P27	0.78	2.82	29.74	1.01	28.62	1.22	29.78	1.65	29.66
P28	1.08	3.28	35.52	3.83	30.25	2.10	31.25	1.54	30.04
P29	1.36	4.75	34.14	3.11	31.57	3.43	35.88	2.52	34.22
P30	3.51	4.93	38.57	3.54	35.83	3.24	37.70	2.97	36.11
Average	1.76	3.86	31.71	2.75	28.48	2.33	32.07	2.03	29.96

FIGURE 4.4: Heatmap of EER (%) Across Participants and Model Configurations (Across-Activity Authentication)



(A) CNN_fixLSTM P18 (best) results

(B) CNN_fixLSTM P4 (worst) results

FIGURE 4.5: Comparison of Best and Worst Across-Activity User Authentication Confusion Matrices

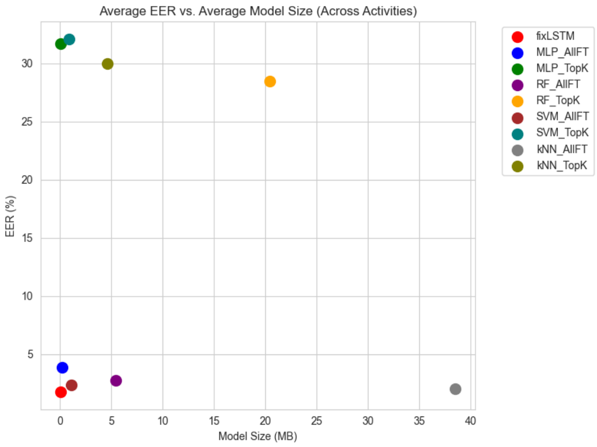


Figure 4.6: Trade-Off Plot: Average EER vs. Average Model Size (Across-Activities Authentication)

4.2 Identification Results

This subsection outlines the performance of ML and DL models in the identification task, assessed under *within-activity* and *across-activity* scenarios. The evaluation focuses on weighted F1 score (WFS) as the primary metric for identification effectiveness and model size (MB) and computation time (s) as a key factor for deployment feasibility on resource-constrained devices. The results highlight how different architectures perform on each individual activity, and across activities (activity-neutral); which activities seem to be the most discriminative when it comes to uniquely identifying participants; and how feature selection impacts identification performance.

4.2.1 Within-Activity Identification

This subsection analyzes the performance of various machine learning (ML) and deep learning (DL) models in user identification on a within-activity basis.

Figure 4.7 paired with Table 4.3 provide a visualization that allows for comparing the different combinations of classifiers and feature sets, and how they perform for each activity. The best overall classifiers in terms of F1 seem to be SVM (0.973 for all features, 0.959 for top20) and RandomForest (0.960 for all, 0.947 for top20). `run_fast, run_slow, and walk_fast`, like previously noted in 4.1.1, show the highest F1 scores for all classifiers, while the most challenging movement remains `stairs_slow`, suggesting higher inter-user similarity. **CNN_fixLSTM** performs the worst overall. `P1` remains a highly distinguishable participant among all activity-based classifications.

Activity	**RandomForest**		**SVM**		**kNN**		**MLP**		**fixLSTM**
	all	**top20**	**all**	**top20**	**all**	**top20**	**all**	**top20**	
run_fast	0.992	0.978	0.998	0.990	0.972	0.979	0.952	0.909	0.956
run_slow	0.994	0.980	0.988	0.994	0.988	0.986	0.955	0.955	0.954
stairs_fast	0.939	0.919	0.963	0.937	0.915	0.940	0.896	0.870	0.837
stairs_slow	0.896	0.873	0.929	0.887	0.814	0.869	0.775	0.810	0.635
walk_fast	0.992	0.980	0.994	0.986	0.973	0.982	0.955	0.910	0.844
walk_slow	0.950	0.949	0.966	0.960	0.910	0.945	0.860	0.773	0.733
Average	0.960	0.947	0.973	0.959	0.929	0.950	0.899	0.871	0.826

Table 4.3: Weighted F1 (all vs. top20) for Within-Activity Identification.

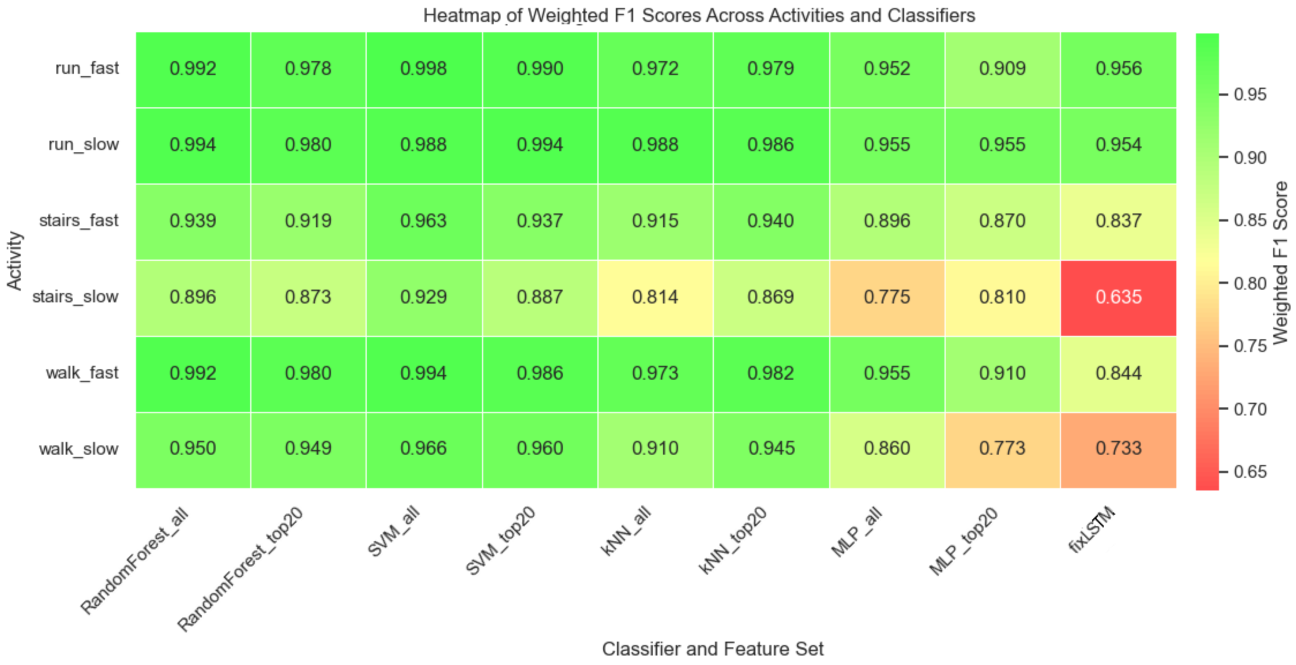


FIGURE 4.7: Heatmap of Weighted F1 Scores Across Activities and Classifiers

Activity	RandomForest		SVM		kNN		MLP		CNNfix
	all	top20	all	top20	all	top20	all	top20	
run_fast	11.557	11.499	3.061	0.546	3.462	0.405	0.248	0.033	0.312
run_slow	12.836	13.004	3.156	0.623	3.423	0.405	0.246	0.068	0.312
stairs_fast	15.053	13.997	3.444	0.651	3.511	0.395	0.257	0.067	0.312
stairs_slow	19.469	22.325	3.778	0.827	3.659	0.410	0.257	0.068	0.312
walk_fast	7.893	12.053	3.213	0.580	3.417	0.395	0.251	0.066	0.312
walk_slow	17.412	16.525	3.593	0.723	3.749	0.429	0.253	0.066	0.312
Average	14.037	14.567	3.374	0.658	3.537	0.407	**0.252**	**0.061**	**0.312**

TABLE 4.4: Model sizes (MB) (all vs. top20) for Within-Activity Identification.

The effect on performance when switching from training the ML algorithms with all features to using the reduced set of features is negligible (with kNN slightly improving (+2.15%) and MLP experiencing the highest drop (-2.77%), while the reduction in model size is substantial (Table 4.4), namely approximately 60%. These findings indicate that, for activity-based user identification, the reduced feature set is sufficient to distinguish between users with high confidence.

The CNN_fixLSTM model performs worse than the ML classifiers for within-activity participant identification, achieving the lowest Weighted F1 Score (0.826) on average (Table 4.3). DL models require larger datasets to generalize well. Since within-activity identification restricts training to only one type of movement, the available training samples per user are relatively limited, making it harder for CNN_fixLSTM to learn user-specific gait patterns effectively. The ML models examined rely on handcrafted features that explicitly capture discriminative motion characteristics, making them evidently more robust to small dataset sizes. The CNN_fixLSTM model is also the largest in size at 0.312 MB, making it a worse choice for device deployment.

When considering Training Time (Table 4.5), kNN is the fastest to train, while CNN_fixLSTM is extremely slow. Reducing features significantly speeds up training for SVM and RandomForest. However, it should be noted that, as similarly shown in Table 4.1, the amount of time (in seconds) required to extract the full set of features is not accounted for herein. This amount of time should be considered when evaluating ML model efficiency. It averages to an additional nearly 4 minutes overhead per activity; the total training time per ML model then becomes longer than that of the DL model. The automatic feature extraction part of the DL model as an advantage justifies the cost of training time.

Activity	**RandomForest**		**SVM**		**kNN**		**MLP**		**CNNfix**
	all	**top20**	**all**	**top20**	**all**	**top20**	**all**	**top20**	
run_fast	9.042	4.947	5.150	6.162	0.678	0.410	8.658	13.199	51.669
run_slow	20.385	4.616	18.251	2.104	0.619	0.211	9.117	6.665	189.650
stairs_fast	20.006	4.146	10.063	2.012	0.456	0.135	11.953	4.678	183.339
stairs_slow	8.491	5.551	13.979	3.684	0.486	0.210	10.030	5.696	181.795
walk_fast	10.763	10.011	8.221	5.958	0.528	0.382	11.105	14.189	196.300
walk_slow	10.265	13.398	17.226	8.271	1.304	0.412	24.274	11.847	189.797
Average	13.825	7.778	12.815	4.698	**0.679**	**0.293**	12.523	9.712	165.758

TABLE 4.5: Model Training Time (s) for Within-Activity Identification.

4.2.2 Across-Activity Identification

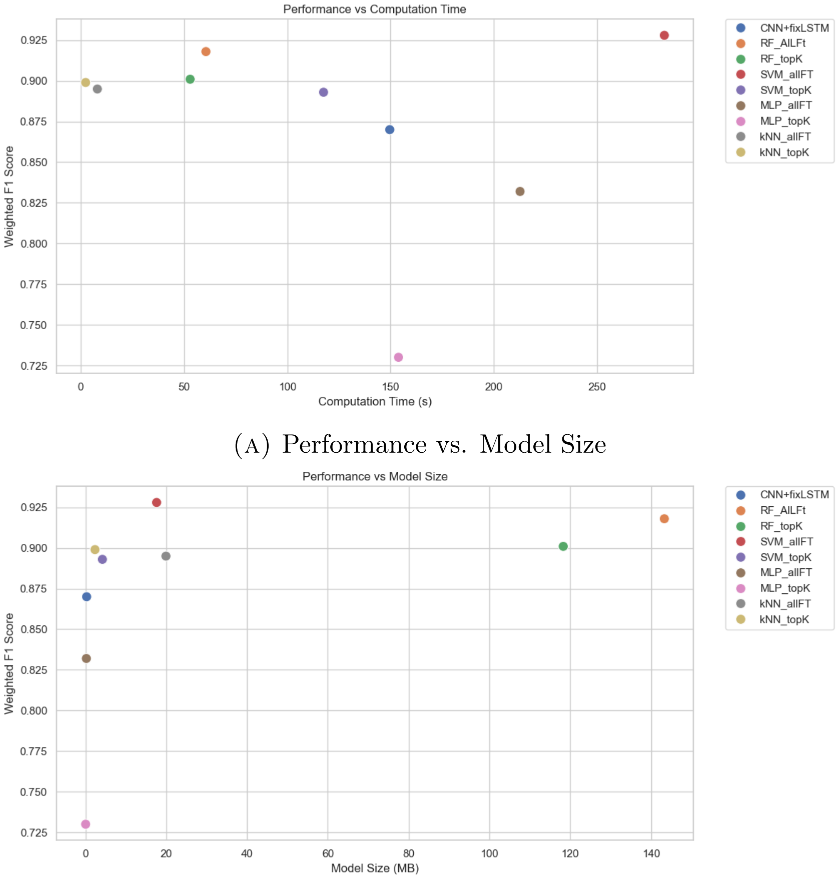
This subsection analyzes the performance of various machine learning (ML) and deep learning (DL) models in user identification across different individuals.

In terms of identification performance, as displayed in Table 4.6, SVM_allFT achieves the highest Weighted F1 score of 0.928, indicating superior classification performance among all models. RF_allFT and RF_topK also demonstrate strong performance with F1 scores of 0.918 and 0.901, respectively. **CNN_fixLSTM** performs on the lower side, with a score of 0.87. MLP is the worst performing model. A trend similar to that in the Within-Activity Identification results (Subsection 4.2.1) is also observed here, namely a relatively negligible drop in performance when working with reduced features compared to a halved model size.

Model	WeightedF1	ComputationTime_s	ModelSize_MB
CNN_fixLSTM	0.870	149.631	0.312
RF_AlLFt	0.918	60.624	143.268
RF_topK	0.901	52.975	118.257
SVM_allFT	0.928	282.619	17.627
SVM_topK	0.893	117.573	4.209
MLP_allFT	0.832	212.748	0.244
MLP_topK	0.731	153.874	0.067
kNN_allFT	0.895	8.044	19.941
kNN_topK	0.899	2.371	2.380

TABLE 4.6: Weighted F1, Computation TIme (s), and Model Size (MB) for Across-Activity Identification.

When considering size, **CNN_fixLSTM** and MLP are the smallest models, with sizes below 1 MB, making them suitable for direct on-device deployment. They do, however, exhibit higher computation times ranging from `149.631s` to `212.748s`. To better visualize the mentioned trade-off between computation time, size, and performance, Figure 4.8 below has been provided, with Figure 4.8a representing the Performance vs. Model Size trade-off, and Figure 4.8b displaying the Performance vs. Computation Time ratio. From the figures, it seems evident that **kNN_topK and SVM_topK provide the best balance between performance and model size,** wherein kNN is a better choice for devices where both storage and processing power are limited, while SVM is a more viable option for moderately resource-constrained devices.



(A) Performance vs. Model Size

(B) Performance vs. Computation Time

Figure 4.8: Trade-Off Plot: Performance vs. Computation Time vs. Model Size (Across-Activities Identification)

4.3 Deep Learning Model Optimization Results

This section compares the optimization processes and outcomes for the DL model, namely the participant identification **CNN_fixLSTM**, highlighting the efficiency, convergence behavior, and performance improvements achieved through the chosen optimization techniques.

4.3.1 Post-Training Quantization (PTQ) Results.

Table 4.7 provides a summary of the effects of Post-Training Quantization (PTQ) on within-activities classification, and across-activities classification (last row). The table compares initial (non-quantized) model results with quantized metrics, including validation accuracy, weighted F1 score, training time, file size, inference time, and model size.

For within-activity identification, most activities experience slight reductions in validation accuracy and weighted F1 score post-quantization. For instance, `run_fast` sees a decrease in WFS from 0.950 to 0.948, indicating negligible impact. Activities like `stairs_fast` and `stairs_slow` exhibit more substantial declines in WFS (from 0.849 to 0.834 and 0.678 to 0.650, respectively). As opposed to the minor performance effect, quantization consistently halves the model file size from 0.318 MB to 0.159 MB across all activities.

For across-activity identification, validation accuracy experiences a minimal drop from 0.883 to 0.881, while weighted F1 score remains stable at 0.881 post-quantization. However, the model size is almost halved (from 0.318 MB to 0.159 MB), and inference time significantly decreases too. **This affirms PTQ as a highly effective method for DL model optimization.** For future references, as PTQ can only be applied to the LSTM component of the model, other optimization techniques can be applied to the CNN part, and their collective effect on size, inference time, and performance can be evaluated.

Activity	**ValAcc**	**Quant_ValAcc**	**WF1**	**Quant_WF1**	**Size(MB)**	**Quant_Size(MB)**	**Inference_Time(s)**
run_fast	0.948	0.950	0.948	0.950	0.318	0.159	0.960
run_slow	0.951	0.949	0.951	0.949	0.318	0.159	1.040
stairs_fast	0.853	0.847	0.849	0.834	0.318	0.159	0.660
stairs_slow	0.690	0.676	0.678	0.650	0.318	0.159	0.850
walk_fast	0.849	0.845	0.852	0.848	0.318	0.159	0.850
walk_slow	0.750	0.755	0.746	0.753	0.318	0.159	0.950
Across-Activities	0.882	0.881	0.883	0.881	0.312	0.163	4.050

Table 4.7: Effects of Post-Training Quantization on Within-Activity and Across-Activities Identification of the CNN_fixLSTM Model

4.3.2 Quantization-Aware Training (QAT) Results.

Results of the performed QAT on the CNN part of **CNN_fixLSTM**, and its implications on performance are discussed below, supported by Table 4.8.

It can be seen that QAT reduces the model size per activity from 0.318 MB to 0.131 MB (and from 0.328 MB to 0.122 MB in the across-activities scenario). This approximately 60% reduction is consistent with int8 quantization replacing float32 parameters in the CNN portion. In some activities, e.g., `run_fast`, validation accuracy and Weighted F1 are slightly higher with QAT. In others, like `stairs_slow`, they slightly decrease (from 0.690 to 0.624 in ValAcc; from 0.678 to 0.609 in Weighted F1). Overall, **QAT leads to modest accuracy changes** (within ±5–10%). The largest drops appear in the more challenging activities (`stairs_fast`, `stairs_slow`), which already had lower baseline performance to begin with.

For across-activities evaluation, the validation accuracy and Weighted F1 both drop slightly (from 0.883 to 0.869 and 0.881 to 0.868, respectively). The model size plummets from 4.050 MB to 0.131 MB, which greatly benefits edge-device deployment. Inference time, however, soars to 19449.58 s, reflecting the significant overhead of QAT on a large multi-activity dataset. **This deems QAT less suitable than PTQ for this specific use case**.

Activity	Initial ValAcc	QAT ValAcc	Initial WF1	QAT WF1	QAT Inference Time (s)	Initial Size (MB)	QAT Size (MB)
run_fast	0.948	0.954	0.948	0.955	3871.45	0.318	0.131
run_slow	0.951	0.937	0.951	0.938	3986.64	0.318	0.131
stairs_fast	0.853	0.773	0.849	0.750	3769.81	0.318	0.131
stairs_slow	0.690	0.624	0.678	0.609	3863.28	0.318	0.131
walk_fast	0.849	0.851	0.852	0.848	3781.74	0.318	0.131
walk_slow	0.750	0.710	0.746	0.700	1144.91	0.318	0.131
Across-Activities	0.882	0.869	0.883	0.868	19449.58	0.312	0.122

TABLE 4.8: Effects of Quantization Aware Training (QAT) on Within-Activity and Across-Activities Identification of the CNN_fixLSTM Model

4.4 Summary of Findings

This subsection synthesizes the key insights derived from the authentication and identification experiments, comparing the strengths and limitations of ML and DL approaches across different scenarios. It highlights trends in model performance, discusses the trade-offs between accuracy and computational efficiency.

The experiments presented in this work comprehensively evaluate both shallow machine learning (ML) and deep learning (DL) models in two primary tasks: user identification and user authentication. The dataset is segmented into "within-activity" and "across-activity" scenarios, allowing for a nuanced view of performance under different types of motion.

With regards to **authentication**, the Equal Error Rate (EER) is the principal metric. Within-activity analysis confirms that the CNN_fixLSTM model, as well as Random Forest and SVM with the full feature set (allFT), achieve consistently low EER values. In particular, running-based activities (fast or slow), and fast-paced walking, enable these models to drop below one percent EER, evidencing a clear advantage for movements that exhibit prominent and repetitive acceleration patterns. By contrast, activities such as `stairs_slow` produce higher EER, indicating that mild or less dynamic gait patterns introduce more overlap among participants and are thus more challenging for the classifiers to distinguish. When all features (allFT) are extracted, model sizes, especially for ensembles like Random Forest or for non-parametric methods like kNN, can become large; however, these same models demonstrate high accuracy. The **CNN_fixLSTM** architecture, although more computationally intensive to train, remains highly compact because it automatically learns features directly from the raw sensor data. Across-activity authentication confirms these observations by showcasing low EER values for **CNN_fixLSTM** in a range of dynamic movements, while again highlighting difficulties in activities with subtle gait patterns. Overall, the **CNN_fixLSTM** approach is shown to strike the most favorable balance between low EER, small model footprint, and relatively straightforward deployment.

In the case of **identification**, the Weighted F1 Score (WFS) is the primary measure of classification performance. On a within–activity basis, SVM and Random Forest with the full feature set achieve some of the highest Weighted F1 Scores, occasionally nearing perfect classification (F1 > 0.95) for activities like `run_fast` and `walk_fast`. However, these configurations also demand greater storage and, in the case of RF, risk becoming prohibitively large. The **CNN_fixLSTM** model continues to be competitive, though it does not always surpass the best-performing shallow methods in smaller per-activity training sets. When moving to across-activity identification, the deep model's capability to integrate spatiotemporal information across varied movements becomes more apparent while SVM_allFT can still produce the highest numerical F1, the **CNN_fixLSTM** model offers advantages in a more uniform performance across user populations. The cost, however, is greater training time relative to some of the shallow algorithms - an overhead that becomes more balanced once the substantial time spent on feature extraction for ML pipelines is factored in.

In terms of **resource optimization**, the model optimization applied to the ML architectures, namely feature reduction from allFT to a top20 feature set (topK), consistently lowers model sizes and speeds up training. This simplification sometimes causes only marginal drops in performance, as seen in certain configurations of SVM and kNN for identification. Yet in authentication tasks, the absence of user-specific feature selection reveals performance degradations when applying topK, presumably due to the omission of discriminative attributes that help distinguish one individual's gait from another.
For DL, Post-Training Quantization (PTQ) further demonstrates that the CNN_fixLSTM model size can be halved while preserving a high level of identification accuracy, and Quantization-Aware Training (QAT) pushes the size reduction even further but, in many configurations, demands considerable training overhead.

To summarize, **running-based motions are inherently more separable than slower or more subtle gait patterns (notably stair climbing).** Shallow ML methods exhibit the strengths of clear interpretability, fast training on topK, and high accuracy when utilizing comprehensive features—whereas the deep **CNN_fixLSTM** model advances the field by significantly cutting down on storage space, removing the need for manual feature extraction and selection, and by performing robustly when data from multiple activities must be combined (across-activity recognition).

4.4.1 Discussion

These results consistently present logical patterns tied to both the data and the model architectures. The marked improvement of high-motion activities—like running—over more subtle or variable movements—like slow stair climbing—makes sense because distinctive signals in acceleration and rotation are more pronounced in brisk tasks. **When motion changes are vigorous and periodic, small differences among individuals become more evident, thus driving down the EER in authentication and boosting the Weighted F1 Score in identification**.

Likewise, the strong performance of Random Forest and SVM on the full feature reflects the fact that large, well-selected feature collections can powerfully represent human gait and that robust, mature ML algorithms can capitalize on such representations. Yet said models tend to also grow in size or in training cost, which explains why they can become less appealing in resource-constrained scenarios like on-device inference. **The deep learning alternative, CNN_fixLSTM, shows a capacity for feature learning from raw signals, circumventing elaborate pre-processing.**

The observation that performance sometimes dips for CNN_fixLSTM in single-activity scenarios, particularly on limited participant-specific data, affirms that deep networks require bigger training datasets, and confining training to a smaller dataset can negatively impact performance. Nonetheless, when the model is given richer data, as in across-activity identification, it narrows the gap in performance with, or surpasses, competing methods.

From the results, **PTQ suffices for many tasks with minimal accuracy loss, while QAT offers deeper yet more computationally expensive compression**, which aligns with standard trade-offs in neural network deployment on microcontrollers [4]. Overall, the observed results align with expectations and known properties of ML and DL models, while highlighting gait-specific findings, namely that high-energy movements simplify discrimination, and that hand-picked features can be effective if well-curated to the movement in question.

Chapter 5

Conclusion

This work presented a new approach to gait-based recognition systems on earables by introducing a lightweight and effective learning approach. It made several notable contributions in the domain of gait-based recognition using ear-worn inertial sensors, contributing to bridging gaps in prior work on biometric authentication and identification in low-power, resource-constrained settings.

A novel dataset was created specifically for gait-based recognition via an IMU-equipped earable, addressing the scarcity of open-source data collected from in-ear accelerometers and gyroscopes. The dataset captures diverse locomotion types—running, walking, and stair climbing—at both slow and fast intensities, collected from 30 participants, and enables evaluation of user recognition tasks and further research efforts. The dataset is intended to be submitted to the public.

Unlike existing works that predominantly study smartphones, smartwatches, or specialized body-worn sensors, this research concentrates on in-ear IMUs with the presumption that the majority of commercial devices come already equipped with said sensors, so a way for securing device access based on the said sensors was needed. The chosen device (OpenEarable) and experimental design reflect typical earbud usage, confirming that it can accurately capture uniquely identifiable biometric gait patterns despite sensor placement challenges and potential signal noise near the head.

By applying both shallow ML (Random Forest, SVM, kNN, MLP) and deep learning models (LSTMOnly, CNN_fixLSTM), this work offers direct performance comparisons across tasks (identification vs. authentication) and scenarios (within-activity vs. across-activity). This dual perspective reveals how classifiers perform under different constraints—informing future design choices.

A key objective of this work was to ensure that solutions could fit on low-power microcontrollers. Several complementary optimization strategies were proposed and tested, namely Feature Set Minimization (ML models) through top-K feature selection, reducing model size and processing overhead; and Quantization (DL models) via Post-Training Quantization (PTQ) and Quantization-Aware Training (QAT). Applying said methods, and evaluating the results, theoretically confirmed the feasibility of deploying neural networks on the OpenEarable device as this reduced the model size to fit within the device's memory capacity. Furthermore, the detailed experiments—encompassing model accuracy (EER, Weighted F1), training/inference times, and memory usage—offer a comprehensive view of real-world trade-offs. Observations about which gait-based activities are most distinguishable and how each classifier scales in size/time highlight best practices for future earable-based biometric deployments.

5.1 Limitations and Future Implications

Although this work demonstrates the practicality and advantages of gait-based recognition systems using IMU-equipped earables, several limitations have been discovered in the process. By recognizing these constraints, future research can better conduct subsequent investigations and real-world implementations.

1. Real-World Generalization and Activity Breadth The dataset collected is relatively constrained in activity scope (running, walking, stair climbing) and is captured under partially controlled conditions. While these activities provide clear gait patterns, everyday movements can be more varied and sporadic. In practice, earables may be worn in highly dynamic environments and across many more actions including the need for an ability to capture action transitions that the current algorithm does not account for. For future work, a larger, more diverse datasets in naturalistic settings (e.g., outdoor spaces) should be collected to capture full-spectrum user motion along with terrain-specific and close-dependent changes. For this purpose, the data of a single participant should be captured on several occasions over a longer period of time. Expanding the range of activities to everyday and unstructured movements, tracking how gait-based recognition adapts to short transitions, context shifts, and multi-person interactions would enhance the targeted recognition capabilities of the device.

2. Limited Participant Pool and Demographic Representation The current dataset involves a limited number of participants that were healthy at the time of conducting the experiement. Gait greatly varies with health conditions. Consequently, including elderly populations, pediatric cohorts, and individuals with mobility impairments, would ensure more robust generalization capabilities, and could increase confidence in clinical or consumer applications.

3. Earbud Placement and Orientation Variance Earable sensors can shift or rotate within the ear, producing noisy or inconsistent signals. While our experiments used relatively stable placements, in practice earbuds may loosen, be reinserted, or be swapped between ears. In this regard, the algorithms could be designed to be orientation-invariant, or could incorporate a wider set of the sensors the earbud is equipped with, e.g., magnetometer, to enable continuous orientation tracking through magnetometer-based corrections. This work did not conduct any prior calibration, adaptive recalibration could be applied next.

4. Timestamp Misalignments and Sensor Fusion Challenges This work assumed synchronous accelerometer and gyroscope data. Android-based or custom earable solutions may face random delays, dropped packets, and cross-channel resynchronization as established with the current OpenEarable device as well, specifically random switch to recording at lower frequency, and missing data points. If extending approaches to multiple sensors (e.g., magnetometer, microphone), a more careful data fusion and labeling strategy will be necessary.

5. Model Size and Computation Constraints Although hardware-aware optimization (e.g., quantization, feature reduction) proved effective, certain deep learning architectures still require considerable training time or memory before quantization. Additionally, the selected `topK` features for ML were not tailored to individual users, occasionally removing user-specific gait cues. As a future implication, a wider variety of optimization techniques, along with inherently lightweight frameworks (EdgeML) should be explored.

Despite the above constraints, the findings of this work confirm the broader potential of earable-based gait-based recognition in the scenarios of, among others, device access security, as gait-based biometrics remain a soft metric yet, fused with other hard biometrics, can enhance device security. This study establishes a solid baseline for earable-based gait-based recognition, demonstrating both its feasibility and need for careful handling of real-world complexities. By extending data collection, refining model robustness, and tackling synchronization issues, future research can deliver on-device gait biometrics collected through the ear for daily consumer use.

Although the field of earables is relatively new and remains underexplored, the outcomes of this work show promising results for broader adoption of earable-based recognition as a subtle and non-intrusive measure for securing device access, at least as a complementary metric to existing methods.

Appendix A

Additional Data

This appendix contains additional data referred in this work.

User	fixLSTM	MLP_AllFT	MLP_TopK	RF_AllFT	RF_TopK	SVM_AllFT	SVM_TopK	kNN_AllFT	kNN_TopK	fixLSTM_MB	MLP_AllFTMB	MLP_TopKMB	RF_AllFTMB	RF_TopKMB	SVM_AllFTMB	SVM_TopKMB	kNN_AllFTMB	kNN_TopKMB
P1	1,19%	4,85%	22,33%	2,61%	16,82%	1,76%	20,92%	0,41%	18,62%	0,05	0,209	0,035	2,971	10,261	0,580	0,392	39,153	4,672
P2	1,89%	5,02%	41,98%	3,04%	34,55%	1,73%	40,73%	1,25%	36,73%	0,05	0,209	0,031	3,413	19,339	0,767	0,792	39,003	4,654
P3	1,28%	2,41%	23,93%	1,28%	23,80%	1,53%	29,84%	0,92%	26,69%	0,05	0,209	0,034	4,030	19,812	0,943	0,792	38,221	4,561
P4	3,96%	6,33%	36,72%	3,87%	32,01%	4,64%	37,29%	3,31%	31,16%	0,05	0,210	0,031	5,523	23,668	1,054	1,211	38,493	4,594
P5	1,34%	4,22%	37,92%	2,81%	37,80%	2,22%	42,91%	2,17%	41,28%	0,05	0,210	0,035	8,859	23,620	1,809	1,090	38,273	4,567
P6	3,75%	6,31%	37,12%	6,17%	31,46%	5,09%	32,85%	4,52%	32,63%	0,05	0,210	0,030	6,140	23,989	1,457	1,205	38,358	4,578
P7	3,27%	3,86%	41,09%	3,04%	36,55%	2,81%	39,09%	2,25%	38,68%	0,05	0,209	0,030	4,819	18,607	0,909	0,725	38,890	4,641
P8	1,53%	3,31%	27,88%	2,00%	26,75%	1,73%	33,77%	1,32%	29,14%	0,05	0,209	0,035	4,066	21,678	0,683	1,079	38,216	4,561
P9	0,92%	1,44%	30,59%	0,45%	26,92%	0,28%	30,35%	0,28%	28,24%	0,05	0,210	0,040	5,296	18,682	1,103	0,646	39,003	4,654
P10	0,97%	3,14%	37,53%	2,13%	34,27%	1,53%	34,93%	1,78%	33,82%	0,05	0,209	0,031	5,581	23,945	1,197	1,121	38,211	4,560
P11	1,14%	4,19%	26,31%	2,47%	25,16%	1,47%	31,47%	1,97%	26,99%	0,05	0,209	0,037	4,474	17,204	0,925	0,648	38,939	4,647
P12	2,77%	4,46%	28,13%	4,64%	23,87%	3,68%	28,92%	2,97%	25,23%	0,05	0,211	0,034	7,195	22,238	1,459	0,994	38,262	4,566
P13	2,87%	5,11%	29,75%	5,10%	26,89%	4,07%	31,49%	2,86%	27,54%	0,05	0,210	0,031	8,008	23,366	1,625	1,110	38,224	4,562
P14	1,27%	3,57%	23,68%	2,87%	20,99%	2,10%	26,06%	1,95%	23,89%	0,05	0,209	0,031	5,806	18,895	1,131	0,681	38,503	4,595
P15	1,28%	3,11%	29,95%	1,99%	29,43%	1,14%	30,37%	2,20%	29,03%	0,05	0,209	0,035	4,636	18,540	0,792	0,700	38,888	4,641
P16	0,69%	2,36%	36,18%	0,63%	34,31%	1,27%	34,65%	1,39%	34,32%	0,05	0,209	0,031	4,662	21,945	1,208	1,129	38,460	4,590
P17	1,23%	3,93%	21,65%	3,06%	19,46%	2,13%	23,32%	1,71%	20,20%	0,05	0,209	0,036	6,396	17,506	1,177	0,602	38,367	4,579
P18	0,63%	1,44%	34,13%	0,59%	28,83%	0,66%	29,96%	0,88%	29,66%	0,05	0,208	0,032	3,412	21,510	0,684	0,967	38,401	4,583
P19	2,75%	4,48%	16,57%	2,53%	13,95%	2,32%	20,25%	1,46%	16,84%	0,05	0,209	0,033	3,077	9,293	0,615	0,352	39,188	4,677
P20	0,91%	3,36%	36,51%	2,67%	29,35%	1,60%	33,63%	2,27%	33,59%	0,05	0,210	0,031	5,068	22,053	1,029	0,951	38,525	4,597
P21	1,95%	3,88%	30,38%	2,33%	29,26%	3,78%	32,17%	2,31%	28,45%	0,05	0,209	0,030	5,604	21,174	1,154	0,954	38,597	4,606
P22	1,18%	3,85%	30,98%	2,53%	29,55%	2,49%	30,60%	3,04%	29,53%	0,05	0,209	0,039	4,765	19,016	1,025	0,758	38,665	4,614
P23	0,67%	2,05%	29,01%	0,75%	22,80%	0,84%	35,16%	0,53%	31,03%	0,05	0,208	0,031	3,277	21,847	0,763	0,968	38,187	4,557
P24	1,75%	4,28%	35,63%	4,15%	30,06%	2,40%	30,54%	2,38%	32,35%	0,05	0,210	0,036	7,089	22,752	1,536	1,043	38,267	4,567
P25	3,34%	4,58%	32,14%	3,50%	32,46%	2,75%	34,64%	2,79%	32,04%	0,05	0,209	0,031	6,564	20,357	1,483	0,948	38,587	4,605
P26	1,56%	4,42%	35,25%	3,70%	30,69%	3,76%	31,64%	3,36%	31,12%	0,05	0,210	0,035	7,005	23,167	1,459	0,992	38,297	4,570
P27	0,78%	2,82%	29,74%	1,01%	28,62%	1,22%	29,78%	1,65%	29,66%	0,05	0,210	0,031	4,637	21,280	1,147	0,914	38,173	4,555
P28	1,08%	3,28%	35,52%	3,83%	30,25%	2,10%	31,25%	1,54%	30,04%	0,05	0,210	0,035	7,149	21,917	1,361	0,954	38,192	4,558
P29	1,36%	4,75%	34,14%	3,11%	31,57%	3,43%	35,88%	2,52%	34,22%	0,05	0,210	0,030	6,511	22,729	1,482	1,232	38,267	4,567
P30	3,51%	4,93%	38,57%	3,54%	35,83%	3,24%	37,70%	2,97%	36,11%	0,05	0,210	0,038	7,342	23,222	1,707	1,127	38,420	4,585
Average	1,76%	3,86%	31,71%	2,75%	28,48%	2,33%	32,07%	2,03%	29,96%	0,05	0,209	0,033	5,446	20,454	1,142	0,903	38,508	4,595

TABLE A.1: EER (%) and Model Sizes (MB) for All Models in Across-Activity User Authentication.

Bibliography

[1] Earphones and headphones market growth, size, share, trends, and forecast 2032, 2023. URL: https://www.zionmarketresearch.com/report/earphones-and-headphones#:~:text=According%20to%20the%20report%20published,12.70%25%20during%20the%20forecast%20period.

[2] Earset dataset: A multi-modal dataset for in-ear ppg and imu sensing, 2023. Available at Zenodo. doi:10.5281/zenodo.8142332.

[3] Temitayo Caroline Adeniran, Rasheed G Jimoh, Emmanuel U Abah, Nasir Faruk, Emmanuel Alozie, and Agbotiname Lucky Imoize. Vulnerability assessment studies of existing knowledge-based authentication systems: A systematic review. *Sule Lamido University Journal of Science & Technology*, 8(1):34–61, 2024.

[4] Subhrangshu Adhikary, Subhadeep Biswas, Arindam Ghosh, and Subrata Nandi. Tiny-biogait—embedded intelligence and homologous time approximation warping for gait biometric authentication from imu signals. *Smart Health*, 34:100515, 2024.

[5] Sumeyye Agac and Ozlem Durmaz Incel. User authentication and identification on smart glasses with motion sensors. *SN Computer Science*, 4(6):761, 2023.

[6] N. Al-Naffakh, N. Clarke, F. Li, and P. Haskell-Dowland. Unobtrusive gait recognition using smartwatches. In *Proceedings of the 2017 International Conference on Biometrics and Security (BIOSIG)*, 2017. doi:10.23919/BIOSIG.2017.8053523.

[7] Neamah Al-Naffakh, Nathan Clarke, and Fudong Li. Continuous user authentication using smartwatch motion sensor data. In *Trust Management XII: 12th IFIP WG 11.11 International Conference, IFIPTM 2018, Toronto, ON, Canada, July 10–13, 2018, Proceedings 12*, pages 15–28. Springer, 2018.

[8] Hadeel Alobaidi, Nathan Clarke, Feng Li, and Abdulrahman Alruban. Real-world smartphone-based gait recognition. *Computers Security*, 2022. DOI unavailable.

[9] Yasith Amarasinghe, Darshana Sandaruwan, Thilina Madusanka, Indika Perera, and Lakmal Meegahapola. Multimodal earable sensing for human energy expenditure estimation. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–4. IEEE, 2023.

[10] Toshiyuki Ando, Yuki Kubo, Buntarou Shizuki, and Shin Takahashi. Canalsense: Face-related movement recognition system based on sensing air pressure in ear canals. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*, pages 679–689, New York, NY, USA, 2017. Association for Computing Machinery. doi:10.1145/3126594.3126649.

[11] Lloyd Vincent R Asuncion, Joan Xyrel P De Mesa, Patrick Kyle H Juan, Nathaniel T Sayson, and Angelo R Dela Cruz. Thigh motion-based gait analysis for human identification using inertial measurement units (imus). In *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pages 1–6. IEEE, 2018.

[12] S. N. Basharzad and M. Fazeli. Knowledge based dynamic password. In *2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pages 367–372. IEEE, 2017. doi:10.1109/KBEI.2017.8325004.

[13] Lauren C Benson, Christian A Clermont, Sean T Osis, Dylan Kobsar, and Reed Ferber. Classifying running speed conditions using a single wearable sensor: Optimal segmentation and feature extraction methods. *Journal of biomechanics*, 71:94–99, 2018.

[14] Mehrab Bin Morshed, Harish Kashyap Haresamudram, Dheeraj Bandaru, Gregory D Abowd, and Thomas Plötz. A personalized approach for developing a snacking detection system using earbuds in a semi-naturalistic setting. In *Proceedings of the 2022 ACM international symposium on wearable computers*, pages 11–16, 2022.

[15] Nam Bui, Nhat Pham, Jessica Jacqueline Barnitz, Zhanan Zou, Phuc Nguyen, Hoang Truong, Taeho Kim, Nicholas Farrow, Anh Nguyen, Jianliang Xiao, et al. ebp: an ear-worn device for frequent and comfortable blood pressure monitoring. *Communications of the ACM*, 64(8):118–125, 2021.

[16] Clara Piris Burgos, Lea Gärtner, Miguel A González Ballester, Jérôme Noailly, Fabian Stöcker, Martin Schönfelder, Tim Adams, and Simone Tassani. In-ear accelerometer-based sensor for gait classification. *IEEE Sensors Journal*, 20(21):12895–12902, 2020.

[17] Qian Cao, Fei Xu, and Huiyong Li. User authentication by gait data from smartphone sensors using hybrid deep learning network. *Mathematics*, 10(13):2283, 2022.

[18] Yetong Cao, Chao Cai, Fan Li, Zhe Chen, and Jun Luo. Enabling passive user authentication via heart sounds on in-ear microphones. *IEEE Transactions on Dependable and Secure Computing*, 2024.

[19] Jin-Young Choi, Seonghee Jeon, Hana Kim, Jaeyoung Ha, Gyeong-suk Jeon, Jeong Lee, Sung-il Cho, et al. Health-related indicators measured using earable devices: Systematic review. *JMIR mHealth and uHealth*, 10(11):e36696, 2022.

[20] Seokmin Choi, Junghwan Yim, Yincheng Jin, Yang Gao, Jiyang Li, and Zhanpeng Jin. Earppg: Securing your identity with your ears. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 835–849, 2023.

[21] Romit Roy Choudhury. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, pages 147–153, 2021.

[22] Hu Chunsheng, Wang De, Zhao Huidong, and Li Guoli. Human gait feature data analysis and person identification based on imu. In *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, pages 437–442. IEEE, 2020.

[23] Christian A Clermont, Lauren C Benson, Sean T Osis, Dylan Kobsar, and Reed Ferber. Running patterns for male and female competitive and recreational runners based on accelerometer data. *Journal of sports sciences*, 37(2):204–211, 2019.

[24] Arianna De Vecchi, Alice Scandelli, Federica Bossi, Benedetta Caterina Casadei, Marco Boschi, and Federica Villa. On-the-edge gait analysis using a smart earable inertial measurement unit. In *2024 IEEE Sensors Applications Symposium (SAS)*, pages 1–6, 2024. doi:10.1109/SAS60918.2024.10636526.

[25] Arianna De Vecchi, Alice Scandelli, Federica Bossi, Benedetta Caterina Casadei, Marco Boschi, and Federica Villa. On-the-edge gait analysis using a smart earable inertial measurement unit. In *2024 IEEE Sensors Applications Symposium (SAS)*, pages 1–6. IEEE, 2024.

[26] Omid Dehzangi, Mojtaba Taherisadr, and Raghvendar ChangalVala. Imu-based gait recognition using convolutional neural networks and multi-sensor fusion. *Sensors*, 17(12):2735, 2017.

[27] Paula Delgado-Santos, Ruben Tolosana, Richard Guest, Ruben Vera-Rodriguez, Farzin Deravi, and Aythami Morales. Gaitprivacyon: Privacy-preserving mobile gait biometrics using unsupervised learning. *Pattern Recognition Letters*, 161:30–37, 2022.

[28] Mohammad Omar Derawi, Claudia Nickel, Patrick Bours, and Christoph Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *Proceedings of the 2010 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'10)*, 2010.

[29] Richard O Duda, Peter E Hart, et al. *Pattern classification*. John Wiley & Sons, 2006.

[30] Andrea Ferlini, Dong Ma, Robert Harle, and Cecilia Mascolo. Eargate: gait-based user identification with in-ear microphones. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 337–349, 2021.

[31] Andrea Ferlini, Alessandro Montanari, Andreas Grammenos, Robert Harle, and Cecilia Mascolo. Enabling in-ear magnetic sensing: Automatic and user transparent magnetometer calibration. In *2021 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–8. IEEE, 2021. doi:10.1109/PerCom50583.2021.9439041.

[32] Kenneth R Foster, Robert Koprowski, and Joseph D Skufca. Machine learning, medical diagnosis, and biomedical engineering research-commentary. *Biomedical engineering online*, 13:1–9, 2014.

[33] Davrondzhon Gafurov. Performance and security analysis of gait-based user authentication. 2008.

[34] Davrondzhon Gafurov, Kirsi Helkala, and Torkjel Søndrol. Biometric gait authentication using accelerometer sensor. *J. comput.*, 1(7):51–59, 2006.

[35] Yang Gao, Wei Wang, Vir V. Phoha, Wei Sun, and Zhanpeng Jin. Earecho: Using ear canal echo for wearable authentication. 3(3), September 2019. doi:10.1145/3351239.

[36] Shkurta Gashi, Chulhong Min, Alessandro Montanari, Silvia Santini, and Fahim Kawsar. A multidevice and multimodal dataset for human energy expenditure estimation using wearable devices. *Scientific Data*, 9(1):537, 2022.

[37] Grand View Research. Earphone and headphone market size, share trends analysis report by product, by technology, by price range, by application, by region, and segment forecasts, 2022 - 2030, 2022. URL: https://www.grandviewresearch.com/industry-analysis/earphone-and-headphone-market.

[38] Ying Gu, Evy Cleeren, Jonathan Dan, Kasper Claes, Wim Van Paesschen, Sabine Van Huel, and Borbála Hunyadi. Comparison between scalp eeg and behind-the-ear eeg for development of a wearable seizure detection system for patients with focal epilepsy. *Sensors*, 18(1):29, 2018. doi:10.3390/s18010029.

[39] Feiyu Han, Panlong Yang, Shaojie Yan, Haohua Du, and Yuanhao Feng. Breathsign: Transparent and continuous in-ear authentication using bone-conducted breathing biometrics. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2023.

[40] Lin He, Chen Ma, Chang Tu, and Yanyong Zhang. Gait2vec: Continuous authentication of smartphone users based on gait behavior. In *2022 IEEE 25th International Conference on Computational Science and Engineering (CSE)*. IEEE, 2022. DOI unavailable.

[41] Tahera Hossain, Md Shafiqul Islam, Md Atiqur Rahman Ahad, and Sozo Inoue. Human activity recognition using earable device. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 81–84, 2019.

[42] Chengli Hou. A study on imu-based human activity recognition using deep learning and traditional machine learning. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, pages 225–234. IEEE, 2020.

[43] Changshuo Hu, Thivya Kandappu, Yang Liu, Cecilia Mascolo, and Dong Ma. Breathpro: Monitoring breathing mode during running with earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2):1–25, 2024.

[44] Changshuo Hu, Xiao Ma, Xinger Huang, Yiran Shen, and Dong Ma. Lr-auth: Towards practical implementation of implicit user authentication on earbuds. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(4):1–27, 2024.

[45] Changshuo Hu, Xiao Ma, Dong Ma, and Ting Dang. Lightweight and non-invasive user authentication on earables. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications*, pages 36–41, 2023.

[46] Haohua Huang, Pan Zhou, Ye Li, and Fangmin Sun. A lightweight attention-based cnn model for efficient gait recognition with wearable imu sensors. *Sensors*, 21(8):2866, 2021.

[47] Shawn Hymel, Colby Banbury, Daniel Situnayake, Alex Elium, Carl Ward, Mat Kelcey, Mathijs Baaijens, Mateusz Majchrzycki, Jenny Plunkett, David Tischler, Alessandro Grande, Louis Moreau, Dmitry Maslov, Artie Beavis, Jan Jongboom, and Vijay Janapa Reddi. Edge impulse: An mlops platform for tiny machine learning, 2023. URL: https://arxiv.org/abs/2212.03332, arXiv:2212.03332.

[48] Erifili Ichtiaroglou. Gait verification using deep learning models, accelerometers and gyroscope data, 2020.

[49] Author(s) (if available). Dataset title (replace with actual title), Year. Accessed: Month, Year. doi:10.5281/zenodo.6420886.

[50] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2704–2713, 2018.

[51] Anil K. Jain, Arun Ross, and Sharath Pankanti. Biometrics: A tool for information security. *IEEE Transactions on Information Forensics and Security*, 1(2):125–143, 2006. doi:10.1109/TIFS.2006.873653.

[52] Majid Janidarmian, Atena Roshan Fekr, Katarzyna Radecka, and Zeljko Zilic. A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors*, 17(3):529, 2017.

[53] D. Jeong, J. Jeong, Y. Chae, and H. Choi. Identification of attention state for menu-selection using in-ear eeg recording. In *2017 5th International Winter Conference on Brain-Computer Interface (BCI)*, pages 112–114. IEEE, 2017. doi:10.1109/IWW-BCI.2017.7858176.

[54] Nan Jiang, Terence Sim, and Jun Han. Earwalk: towards walking posture identification using earables. In *Proceedings of the 23rd Annual International Workshop on Mobile Computing Systems and Applications*, pages 35–40, 2022.

[55] Felix Juefei-Xu, Chandrasekhar Bhagavatula, Aaron Jaech, Unni Prasad, and Marios Savvides. Gait-id on the move: Pace independent human identification using cell phone accelerometer dynamics. In *Proceedings of the 2012 IEEE 5th International Conference on Biometrics: Theory, Applications, and Systems (BTAS'12)*, Los Alamitos, CA, 2012. IEEE.

[56] Adil Khan, Omar Galarraga, Sonia Garcia-Salicetti, and Vincent Vigneron. Deep learning for quantified gait analysis: a systematic literature review. *IEEE Access*, 2024.

[57] Jure Kovač, Vitomir Štruc, and Peter Peer. Frame–based classification for cross-speed gait recognition. *Multimedia Tools and Applications*, 78(5):5621–5643, 2019.

[58] Rajesh Kumar, Can Isik, and Vir V Phoha. Treadmill assisted gait spoofing (tags) an emerging threat to wearable sensor-based gait authentication. *Digital Threats: Research and Practice*, 2(3):1–17, 2021.

[59] Rajesh Kumar, Vir V Phoha, and Anshumali Jain. Treadmill attack on gait-based authentication systems. In *2015 IEEE 7th international conference on biometrics theory, applications and systems (BTAS)*, pages 1–7. IEEE, 2015.

[60] Maxime Laporte, Priyansh Baglat, Sami Gashi, Martin Gjoreski, Silvia Santini, and Marc Langheinrich. Detecting verbal and non-verbal gestures using earables. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, pages 283–286. ACM, 2021. doi:10.1145/3460418.3479288.

[61] Robyn Larracy, Angkoon Phinyomark, and Erik Scheme. Gait representation: from vision-based to floor sensor-based gait recognition. In *2023 IEEE Sensors Applications Symposium (SAS)*, pages 1–6, 2023. doi:10.1109/SAS58821.2023.10254014.

[62] Galina Lavrentyeva, Sergey Novoselov, Egor Malykh, Alexander Kozlov, Oleg Kudashev, and Vadim Shchemelinin. Audio replay attack detection with deep learning frameworks. In *Proceedings of INTERSPEECH*, pages 82–86, 2017.

[63] Euihyoek Lee, Chulhong Min, Jeaseung Lee, Jin Yu, and Seungwoo Kang. Automatic detection of reactions to music via earable sensing. *arXiv preprint arXiv:2304.03295*, 2023.

[64] Wei-Han Lee and Ruby B Lee. Multi-sensor authentication to improve smartphone security. In *2015 International conference on information systems security and privacy (ICISSP)*, pages 1–11. IEEE, 2015.

[65] Wei-Han Lee and Ruby B Lee. Implicit smartphone user authentication with sensors and contextual machine learning. In *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 297–308. IEEE, 2017.

[66] Jiao Li, Yang Liu, Zhenjiang Li, and Jin Zhang. Earpass: Continuous user authentication with in-ear ppg. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, pages 327–332, 2023.

[67] Ke Li, Ruidong Zhang, Bo Liang, François Guimbretière, and Cheng Zhang. Eario: A low-power acoustic sensing earable for continuously tracking detailed facial movements. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–24, 2022.

[68] Jianwei Liu, Wenfan Song, Leming Shen, Jinsong Han, and Kui Ren. Secure user verification and continuous authentication via earphone imu. *IEEE Transactions on Mobile Computing*, 22(11):6755–6769, 2022.

[69] Darrell Loh, Tien J Lee, Shaghayegh Zihajehzadeh, Reynald Hoskinson, and Edward J Park. Fitness activity classification by using multiclass support vector machines on head-worn sensors. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 502–505. IEEE, 2015.

[70] Jiwen Lu and Erhu Zhang. Gait recognition for human identification based on ica and fuzzy svm through multiple views fusion. *Pattern Recognition Letters*, 28(16):2401–2411, 2007.

[71] Ioan Lucan Orășan, Ciprian Seiculescu, and Cătălin Daniel Căleanu. A brief review of deep neural network implementations for arm cortex-m processor. *Electronics*, 11(16):2545, 2022.

[72] Maria De Marsico and Alessio Mecca. A survey on gait recognition via wearable sensors. *ACM Computing Surveys (CSUR)*, 52(4):1–39, 2019.

[73] Chenren Min, Akhil Mathur, and Fahim Kawsar. Exploring audio and kinetic sensing on earable devices. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications*, pages 33–38. ACM, 2018. doi:10.1145/3211960.3211966.

[74] Bendik B Mjaaland, Patrick Bours, and Danilo Gligoroski. Walk the walk: Attacking gait biometrics by imitation. In *Information Security: 13th International Conference, ISC 2010, Boca Raton, FL, USA, October 25-28, 2010, Revised Selected Papers 13*, pages 361–380. Springer, 2011.

[75] Alessandro Montanari, Andrea Ferlini, Ananta Narayanan Balaji, Cecilia Mascolo, and Fahim Kawsar. Earset: A multi-modal dataset for studying the impact of head and facial movements on in-ear ppg signals. *Scientific Data*, 10(1):850, 2023.

[76] Muhammad Muaaz and René Mayrhofer. Smartphone-based gait recognition: From authentication to imitation. *IEEE Transactions on Mobile Computing*, 16(11):3209–3221, 2017.

[77] Muhammad Muaaz and Claudia Nickel. Influence of different walking speeds and surfaces on accelerometer-based biometric gait recognition. In *Proceedings of the 2012 35th International Conference on Telecommunications and Signal Processing (TSP'12)*, 2012.

[78] Sheikh Nooruddin. On the design of efficient deep learning methods for human activity recognition in resource constrained devices. Master's thesis, University of Waterloo, 2023.

[79] Nobuyuki Oishi, Benedetta Heimler, Lloyd Pellatt, Meir Plotnik, and Daniel Roggen. Detecting freezing of gait with earables trained from vr motion capture data. In *Proceedings of the 2021 ACM International Symposium on Wearable Computers*, pages 33–37, 2021.

[80] OpenEarable Contributors. Openearable: Open source dashboard and firmware. https://github.com/OpenEarable/open-earable/blob/main/README.md#Arduino-IDE.

[81] Ioannis Papavasileiou, Zhi Qiao, Chenyu Zhang, Wenlong Zhang, Jinbo Bi, and Song Han. Gaitcode: Gait-based continuous authentication using multimodal learning and wearable sensors. *Smart Health*, 19:100162, 2021. URL: https://www.sciencedirect.com/science/article/pii/S2352648320300544, doi:10.1016/j.smhl.2020.100162.

[82] Anubha Parashar, Apoorva Parashar, Andrea F Abate, Rajveer Singh Shekhawat, and Imad Rida. Real-time gait biometrics for surveillance applications: A review. *Image and Vision Computing*, page 104784, 2023.

[83] P Jonathon Phillips, Alvin Martin, Charles L Wilson, and Mark Przybocki. An introduction evaluating biometric systems. *Computer*, 33(2):56–63, 2000.

[84] Xueteng Qian, Xiuzhen Guo, Yongjie Yang, Xiaoran Fan, and Longfei Shangguan. Headfi ii: Toward more resilient earable computing platform. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 827–828, 2022.

[85] David Quintero, Daniel J Lambert, Dario J Villarreal, and Robert D Gregg. Real-time continuous gait phase and speed estimation from a single sensor. In *2017 IEEE Conference on Control Technology and Applications (CCTA)*, pages 847–852. IEEE, 2017.

[86] Yanzhi Ren, Yingying Chen, Mooi Choo Chuah, and Jie Yang. User verification leveraging gait recognition for smartphone-enabled mobile healthcare systems. *IEEE Transactions on Mobile Computing*, 14(9):1961–1974, 2014.

[87] Tobias Röddiger, Christopher Clarke, Paula Breitling, Tim Schneegans, Haibin Zhao, Hans Gellersen, and Michael Beigl. Sensing with earables: A systematic literature review and taxonomy of phenomena. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 6(3):1–57, 2022.

[88] Tobias Röddiger, Tobias King, Dylan Ray Roodt, Christopher Clarke, and Michael Beigl. Openearable: Open hardware earable sensing platform. In *Proceedings of the 1st International Workshop on Earable Computing*, EarComp'22, pages 29–34, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3544793.3563415.

[89] Abdul Saboor, Triin Kask, Alar Kuusik, Muhammad Mahtab Alam, Yannick Le Moullec, Imran Khan Niazi, Ahmed Zoha, and Rizwan Ahmad. Latest research trends in gait analysis using wearable sensors and machine learning: A systematic review. *Ieee Access*, 8:167830–167864, 2020.

[90] Swapnil Sayan Saha, Sandeep Singh Sandha, Siyou Pei, Vivek Jain, Ziqi Wang, Yuchen Li, Ankur Sarker, and Mani Srivastava. Auritus: An open-source optimization toolkit for

training and development of human movement models and filters using earables. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 6(2):1–34, 2022.

[91] Francesca Salis, Stefano Bertuletti, Tecla Bonci, Marco Caruso, Kirsty Scott, Lisa Alcock, Ellen Buckley, Eran Gazit, Clint Hansen, Lars Schwickert, et al. A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. *Frontiers in Bioengineering and Biotechnology*, 11:1143248, 2023.

[92] Rubén San-Segundo, Ricardo Cordoba, Javier Ferreiros, and Luis Fernando D'Haro-Enríquez. Frequency features and gmm-ubm approach for gait-based person identification using smartphone inertial signals. *Pattern Recognition Letters*, 73:60–67, 2016.

[93] Giovanni Marco Scalera, Andrea Cereatti, and Ugo Della Croce. Gait regularity assessed by wearable sensors: Comparison between accelerometer and gyroscope data for different sensor locations and walking speeds in healthy subjects. *Journal of Biomechanics*, 113:110115, 2020. doi:10.1016/j.jbiomech.2020.110115.

[94] Werner Schiehlen and Daniel García-Vallejo. Walking dynamics from mechanism models to parameter optimization. *Procedia IUTAM*, 2:199–211, 2011.

[95] Suranga Seneviratne, Yining Hu, Tham Nguyen, Guohao Lan, Sara Khalifa, Kanchana Thilakarathna, Mahbub Hassan, and Aruna Seneviratne. A survey of wearable devices and challenges. *IEEE Communications Surveys & Tutorials*, 19(4):2573–2620, 2017.

[96] Shu Shen, Shao-Shan Sun, Wen-Juan Li, Ru-Chuan Wang, Peng Sun, Sen Wang, and Xin-Yu Geng. A classifier based on multiple feature extraction blocks for gait authentication using smartphone sensors. *Computers and Electrical Engineering*, 108:108663, 2023.

[97] Jasvinder Pal Singh, Sanjeev Jain, Sakshi Arora, and Uday Pratap Singh. Vision-based gait recognition: A survey. *Ieee Access*, 6:70497–70527, 2018.

[98] Isaac Skog, Peter Handel, John-Olof Nilsson, and Jouni Rantakokko. Zero-velocity detection—an algorithm evaluation. *IEEE transactions on biomedical engineering*, 57(11):2657–2666, 2010.

[99] Sebastijan Sprager and Matjaz B Juric. Inertial sensor-based gait recognition: A review. *Sensors*, 15(9):22089–22127, 2015.

[100] Tanmay Srivastava, Shijia Pan, Phuc Nguyen, and Shubham Jain. Jawthenticate: Microphone-free speech-based authentication using jaw motion and facial vibrations. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, pages 209–222, 2023.

[101] Ioannis Stylios, Spyros Kokolakis, Olga Thanou, and Sotirios Chatzis. Behavioral biometrics continuous user authentication on mobile devices: A survey. *Information Fusion*, 66:76–99, 2021. URL: https://www.sciencedirect.com/science/article/pii/S1566253520303493, doi:10.1016/j.inffus.2020.08.021.

[102] Shuai Tao, Xiaowei Zhang, Huaying Cai, Zeping Lv, Caiyou Hu, and Haiqun Xie. Gait based biometric personal authentication by using mems inertial sensors. *Journal of Ambient Intelligence and Humanized Computing*, 9:1705–1712, 2018.

[103] Auritus Team. Auritus dataset, 2022. Accessed: 2024-11-23. URL: https://github.com/nesl/auritus/tree/main/Dataset.

[104] Technavio. Earphone and headphone market by product, technology, and geography - industry analysis and forecast 2021-2025. https://www.technavio.com/report/earphone-and-headphone-market-industry-analysis. Accessed: 2024-11-07.

[105] TECO Research Group. Openearable: Wearable device platform. https://open-earable.teco.edu/. Accessed: 2024-10-30.

[106] Hoang Minh Thang, Vo Quang Viet, Nguyen Dinh Thuc, and Deokjai Choi. Gait identification using accelerometer on mobile phone. In *2012 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pages 344–348. IEEE, 2012.

[107] Luke K Topham, Wasiq Khan, Dhiya Al-Jumeily, Atif Waraich, and Abir J Hussain. Gait identification using limb joint movement and deep machine learning. *IEEE Access*, 10:100113–100127, 2022.

[108] Dante Trabassi, Mariano Serrao, Tiwana Varrecchia, Alberto Ranavolo, Gianluca Coppola, Roberto De Icco, Cristina Tassorelli, and Stefano Filippo Castiglia. Machine learning approach to support the detection of parkinson's disease in imu-based gait analysis. *Sensors*, 22(10):3700, 2022.

[109] Lam Tran, Thang Hoang, Thuc Nguyen, Hyunil Kim, and Deokjai Choi. Multi-model long short-term memory network for gait recognition using window-based data segment. *IEEE Access*, 9:23826–23839, 2021.

[110] Dhruv Verma, Sejal Bhalla, Dhruv Sahnan, Jainendra Shukla, and Aman Parnami. Expressear: Sensing fine-grained facial expressions with earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–28, 2021.

[111] Sudip Vhaduri and Christian Poellabauer. Multi-modal biometric-based implicit authentication of wearable device users. *IEEE Transactions on Information Forensics and Security*, 14(12):3116–3125, 2019.

[112] Changsheng Wan, Li Wang, and Vir V Phoha. A survey on gait recognition. *ACM Computing Surveys (CSUR)*, 51(5):1–35, 2018.

[113] Huandong Wang, Qiaohong Yu, Yu Liu, Depeng Jin, and Yong Li. Spatio-temporal urban knowledge graph enabled mobility prediction. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 5(4):1–24, 2021.

[114] Yong Wang, Tianyu Yang, Chunxiao Wang, Feng Li, Pengfei Hu, and Yiran Shen. Budsauth: Toward gesture-wise continuous user authentication through earbuds vibration sensing. *IEEE Internet of Things Journal*, 11(12):22007–22020, 2024. doi:10.1109/JIOT.2024.3380811.

[115] Yong Wang, Tianyu Yang, Chunxiao Wang, Feng Li, Pengfei Hu, and Yiran Shen. Budsauth: Towards gesture-wise continuous user authentication through earbuds vibration sensing. *IEEE Internet of Things Journal*, 2024.

[116] Zi Wang. *Towards Ubiquitous User Authentication and Sensing Based on the Ear Canal and Toothprint Biometrics Using Ear Wearables*. The Florida State University, 2022.

[117] Zi Wang, Yilin Wang, and Jie Yang. Earslide: a secure ear wearables biometric authentication based on acoustic fingerprint. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(1):1–29, 2024.

[118] J. Wu and J. Wang. Pca-based svm for automatic recognition of gait patterns. *Journal of Applied Biomechanics*, 24(1):83–87, 2008.

[119] Jianning Wu, Yuanbo Liu, and Xiaoyan Wu. Early identification of gait asymmetry using a dual-channel hybrid deep learning model based on a wearable sensor. *Symmetry*, 15(4):897, 2023.

[120] Yadong Xie, Fan Li, Yue Wu, Huijie Chen, Zhiyuan Zhao, and Yu Wang. Teethpass: Dental occlusion-based user authentication via in-ear acoustic sensing. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*, pages 1789–1798, 2022. doi:10.1109/INFOCOM48880.2022.9796951.

[121] Wenzhi Xu, Yan Shen, Chao Luo, Jiawen Li, Wei Li, and Albert Y. Zomaya. Gait-watch: A gait-based context-aware authentication system for smart watch via sparse coding. *Ad Hoc Networks*, 2020.

[122] Zhenye Xu and Yao Guo. Gccrr: A short sequence gait cycle segmentation method based on ear-worn imu. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 650–654, 2024.

[123] Li Yang, Xi Li, Zhuoru Ma, Lu Li, Neal Xiong, and Jianfeng Ma. Irga: An intelligent implicit real-time gait authentication system in heterogeneous complex scenarios. *ACM Transactions on Internet Technology*, 23(2):1–29, 2023.

[124] Kuo-Hui Yeh, Chunhua Su, Wayne Chiu, and Lu Zhou. I walk, therefore i am: Continuous user authentication with plantar biometrics. *IEEE Communications Magazine*, 56(2):150–157, 2018.

[125] Xin Zeng, Xiaomei Zhang, Shuqun Yang, Zhicai Shi, and Chihung Chi. Gait-based implicit authentication using edge computing and deep learning for mobile devices. *Sensors*, 21(13), 2021. URL: https://www.mdpi.com/1424-8220/21/13/4592, doi:10.3390/s21134592.

[126] Linghan Zhang, Sheng Tan, Zi Wang, Yili Ren, Zhi Wang, and Jie Yang. Viblive: A continuous liveness detection for secure voice user interface in iot environment. In *Proceedings of the 36th Annual Computer Security Applications Conference*, pages 884–896, 2020.

[127] Yuqi Zhang, Yongzhen Huang, Liang Wang, and Shiqi Yu. A comprehensive study on gait biometrics using a joint cnn-based method. *Pattern Recognition*, 93:228–236, 2019.

[128] Qin Zou, Yanling Wang, Qian Wang, Yi Zhao, and Qingquan Li. Deep learning-based gait recognition using smartphones in the wild. *IEEE Transactions on Information Forensics and Security*, 15:3197–3212, 2020.

[129] Yongpan Zou, Jianhao Weng, Haibo Lei, Danyang Wang, Victor CM Leung, and Kaishun Wu. Earprint: Earphone-based implicit user authentication with behavioural and physiological acoustics. *IEEE Internet of Things Journal*, 2024.