# ADAPTIVE WINDOW SAMPLING AND FILTERING FOR CONTINUOUS MOBILE BEHAVIORAL AUTHENTICATION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MURATCAN KAPLAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
DATA INFORMATICS

AUGUST 2025

**ADAPTIVE WINDOW SAMPLING AND FILTERING FOR CONTINUOUS MOBILE BEHAVIORAL AUTHENTICATION**

submitted by **MURATCAN KAPLAN** in partial fulfillment of the requirements for the degree of **Master of Science in Data Informatics Department, Middle East Technical University** by,

Date:21.08.2025

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

**Name, Surname**: **Muratcan Kaplan**

**Signature** :

# ABSTRACT

## ADAPTIVE WINDOW SAMPLING AND FILTERING FOR CONTINUOUS MOBILE BEHAVIORAL AUTHENTICATION

Kaplan, Muratcan
M.S., Department of Data Informatics
Supervisor: Prof. Dr. Tuğba Taşkaya Temizel

August 2025, 81 pages

Continuous behavioral authentication verifies user identity in real-time through patterns in natural device interactions, such as touch gestures and hand movements. An important challenge in this domain is identifying segments of sensor data that consistently reflect individual behavior while excluding periods of noise or inactivity. This thesis presents an end-to-end framework that uses accelerometer data temporally aligned with touchscreen events on mobile devices. The approach begins with a balanced sampling strategy that constructs triplets of same-user and different-user interaction windows across different sessions, enabling effective Siamese network training. A subsequent filtering stage evaluates the variability of each candidate window, discarding those with low motion or anomaly fluctuations. By combining session-aware sampling with proposed filtering scheme, the method constructs a dataset of sensor windows that consistently reflect distinctive user behaviors. The resulting framework improves the quality of training data for continuous authentication models and contributes toward more robust and unobtrusive security mechanisms for everyday mobile use.

Keywords: Sensor-based Authentication, Behavioral Biometrics, Triplet Sampling Strategies, Entropy Filtering, Siamese Neural Networks

# ÖZ

## UYARLANABİLİR PENCERE ÖRNEKLEME VE FİLTRELEME İLE SÜREKLİ MOBİL DAVRANIŞSAL KİMLİK DOĞRULAMA

Kaplan, Muratcan
Yüksek Lisans, Veri Bilişimi Bölümü
Tez Yöneticisi: Prof. Dr. Tuğba Taşkaya Temizel

Ağustos 2025 , 81 sayfa

Sürekli davranışsal kimlik doğrulama, dokunmatik jestler ve el hareketleri gibi cihazla doğal etkileşimlerdeki örüntüleri kullanarak kullanıcı kimliğini gerçek zamanlı doğrular. Bu alandaki önemli bir zorluk, bireysel davranışı tutarlı bir şekilde yansıtan, gürültü veya hareketsizlik dönemlerini dışlayan sensör verisi bölümlerini tanımlamaktır. Bu tez, mobil cihazlardaki dokunmatik olaylarla zamansal olarak hizalanmış ivmeölçer verisini kullanan uçtan uca bir çerçeve sunar. Yöntem, öncelikle farklı oturumlarda aynı-kullanıcı ve farklı-kullanıcı etkileşim pencerelerinden oluşan üçlüler oluşturmak için dengeli bir örnekleme stratejisi ile başlar ve böylece Siamese ağların etkili eğitilmesini sağlar. Ardından, her aday pencerenin hareketliliğini değerlendirerek düşük hareketli veya anomali dalgalanmaları içeren pencereleri eleyen bir filtreleme aşaması uygulanır. Oturum farkındalıklı örneklemeyi önerilen filtreleme şemasıyla birleştirerek, ayırt edici kullanıcı davranışlarını tutarlı şekilde yansıtan sensör pencerelerinden oluşan yüksek kaliteli bir veri kümesi oluşturulur. Ortaya çıkan sistem, sürekli kimlik doğrulama modellerinin eğitim verisinin kalitesini artırır ve günlük mobil kullanımda daha güvenilir, rahatsız etmeyen güvenlik mekanizmalarına

katkıda bulunur.

Anahtar Kelimeler: Sensör Tabanlı Kimlik Doğrulama, Davranışsal Biyometri, Üçlü Örnekleme Stratejileri, Entropi Filtresi, Siyam Sinir Ağları

To those who keep the Inner Child alive — still dreaming, still learning, still laughing

## ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my thesis supervisor, Prof. Dr. Tuğba Taşkaya Temizel, for her invaluable guidance, support, and expertise throughout this journey. Her mentorship has played a vital role in shaping both this thesis and my development as a researcher.

I am grateful to the faculty members of the Informatics Institute at Middle East Technical University for providing a strong academic foundation, and to my friends who supported me throughout. I extend special thanks to Gürkan Gündüz for his encouragement and insightful contributions during this process.

I would like to extend my heartfelt thanks to my family. My mother Kezban, my father Murat, and my sister Mina have always been my unwavering sources of love, care, and strength. Their constant support has guided me through every step of this journey, and everything I do is driven by my desire to make them proud.

To those I have loved and lost — my grandfathers, grandmothers, and my dear mother-in-law — thank you for the warmth and wisdom you brought into my life. Though no longer with us, your memories continue to inspire me every day. As it is said, "The ones who love us never really leave us." I carry your presence with me always.

Finally, my most profound appreciation goes to my beloved wife, Özcan Dilara Kaplan. Words fall short in expressing what her love, patience, and unwavering belief in me mean. This thesis may represent a significant milestone in my academic journey, but my greatest pride in life is being with her. Her light has guided me through the darkest moments, and her companionship is my greatest accomplishment.

This study was supported by TÜBİTAK Bilim İnsanı Destek Programları Başkanlığı (BİDEB) under the 2210-Yurt İçi Lisansüstü Burs Programı, and I gratefully acknowledge their financial support.

## TABLE OF CONTENTS

ABSTRACT . . . . . iv

ÖZ . . . . . vi

ACKNOWLEDGMENTS . . . . . ix

TABLE OF CONTENTS . . . . . x

LIST OF TABLES . . . . . xiii

LIST OF FIGURES . . . . . xv

LIST OF ABBREVIATIONS . . . . . xviii

CHAPTERS

1 INTRODUCTION . . . . . 1

1.1 Research Questions . . . . . 3

1.2 Contributions of the Study . . . . . 4

1.3 Organization of the Thesis . . . . . 6

2 RELATED WORK . . . . . 7

2.1 Behavioral Biometric Authentication: Methodological Evolution and Task Diversity . . . . . 7

2.2 Datasets for Behavioral Biometric Authentication . . . . . 12

2.3 Keystroke Dynamics in Behavioral Biometrics . . . . . 12

2.4 Sensor Modalities in Behavioral Biometrics . . . . . 14

2.5 Preprocessing and Segmentation Challenges in Sensor-Based Behavioral Biometrics . . . 15

2.6 Decision and Fusion Strategies in Behavioral Biometrics . . . 16

2.7 Research Gap . . . 19

3 METHODOLOGY . . . 21

3.1 Dataset Description . . . 21

3.2 Preprocessing and Segmentation of Sensor Data for User Interaction Modeling . . . 23

3.3 Dual-Branch Feature Extraction Strategy . . . 32

3.4 Session-Balanced Random Window Sampling . . . 36

3.5 Session-Weighted Random Window Sampling . . . 38

3.6 Entropy-Based Window Filtering with Session-Weighted Sampling . . . 40

3.7 Index-Aligned Triplet Sampling . . . 43

3.8 Relative Position-Based Triplet Sampling . . . 45

3.9 Window-Ensembled Session Matching . . . 47

3.10 Baseline Model with Statistical Feature Aggregation . . . 49

4 EXPERIMENT RESULTS . . . 51

4.1 Experiment Setup . . . 51

4.2 Results . . . 52

4.2.1 Comparison of Baseline and Comparison of Session-Balanced Random Window Sampling (SBRWS) . . . 53

4.2.2 Comparison of Session-Balanced Random Window Sampling (SBRWS) and Session-Weighted Random Window Sampling (SWRWS) . . . 54

4.2.3 Comparison of Session-Weighted Sampling (SWRWS) and Entropy-Based Filtering with Session-Weighted Sampling (EBF-SWS) . . . 58

4.2.4 Comprehensive Comparison of Sampling Strategies Integrating Temporal Semantics . . . . . 59

5 DISCUSSIONS & LIMITATIONS . . . . . 61

5.1 Discussions on the Research Questions . . . . . 61

5.2 Limitations . . . . . 63

6 CONCLUSION AND FUTURE WORK . . . . . 65

REFERENCES . . . . . 69

APPENDICES . . . . . 77

APPENDIX A Sensor Length Distributions and Sampling Uniformity . . . . 79

APPENDIX B Baseline Model Hyperparameter Tuning . . . . . 81

# LIST OF TABLES

TABLES

Table 3.1 Behavioral tasks included in BehavePassDB. . . . . 22

Table 3.2 Summary of Siamese Model Architecture . . . . . 36

Table 3.3 Comparison of Triplet Sampling Strategies . . . . . 39

Table 3.4 Window counts at each preprocessing stage . . . . . 42

Table 3.5 List of extracted statistical features per sensor axis. Each feature is computed for $x$, $y$, and $z$ components, resulting in 75 total features per session. . . . . 50

Table 4.1 Best Hyperparameters for Siamese LSTM + MLP Model (Selected via WANDB Bayesian Optimization) . . . . . 52

Table 4.2 Performance Comparison: Baseline (XGBoost) vs. Siamese (Session-Balanced) . . . . . 54

Table 4.3 Comparison of Siamese Network Variants: Session-Balanced vs. Session-Weighted Sampling . . . . . 55

Table 4.4 Mean and standard deviation of Euclidean distances between intra-session (G1–G1) and inter-session (G1–G3) feature vectors for each user. . 57

Table 4.5 Mean and standard deviation of embedding distances between anchor-positive and anchor-negative pairs for selected users under two sampling strategies: Session-Balanced Random Window Sampling (SBRWS) and Session-Weighted Random Window Sampling (SWRWS). Anchor-positive pairs consist of two windows from the same user, while anchor-negative pairs include one window from the anchor user and one from a different user. . . . . 57

Table 4.6 Comparison of Session Weighted Sampling: With vs. Without Entropy Filtering . . . . . 59

Table 4.7 Comparison of Sampling Strategies Integrating Temporal Semantics 60

Table B.1 Bayesian optimization ranges and best configuration for XGBoost hyperparameters. . . . . 81

# LIST OF FIGURES

FIGURES

Figure 1.1 An illustration of the behavioral biometric authentication process. The visual depicts how sensor data from mobile devices are collected, processed, and analyzed to verify or reject user identity based on interaction patterns. . . . . 3

Figure 3.1 Distribution of estimated acquisition frequencies across user-session pairs in the `sensor_acc` data. The x-axis represents a sequential index over all user-session pairs, ordered first by session (g1 to g4), then by user ID (0 to 50). For example, index 0 corresponds to user 0's g1 session, index 50 to user 50's g1 session, and index 52 to user 0's g2 session. The y-axis shows the estimated sampling interval in seconds. This visualization highlights the variability in acquisition rates across sessions and users, justifying the need for adaptive downsampling. 26

Figure 3.2 Illustration of the timestamp correction process applied to accelerometer data. The x-axis represents the interpolated time index $i$ ranging from $0$ to $n-1$, where $n$ is the total number of uniformly spaced samples within a single user-session. For example, $i = 0$ corresponds to the first accelerometer reading in the session, while $i = 52$ refers to the 53rd reading. The y-axis shows the corresponding sensor values. The top plot shows the raw signal with non-uniform timestamps, leading to visible discontinuities. The bottom plot displays the corrected version after linear timestamp interpolation. . . . . 27

Figure 3.3 Distribution of time-series lengths for the keystroke-acceleration modality across all user-session pairs. The x-axis represents a sequential index over 204 user-session pairs, ordered first by session (from g1 to g4) and then by user ID (from 0 to 50). For example, index 0 corresponds to user 0 in session g1, index 50 to user 50 in g1, index 51 to user 0 in g2, and index 153 to user 0 in g4. The y-axis shows the number of recorded sensor samples per sequence. The long-tailed distribution and high variability necessitate fixed-length windowing to ensure input regularity and stable model convergence. . . . . 29

Figure 3.4 Illustration of filtering logic for User 1, Session g1. Only the sensor data temporally aligned with touch activity (red points) is retained. Three high-variance runs unrelated to touch are excluded. . . . . 31

Figure 3.5 Filtered accelerometer data for User 6, Session g1. Sensor segments aligned with user interaction are preserved, improving data quality for modeling. . . . . 31

Figure 3.6 Sensor-touch alignment for User 20, Session g3. Idle sensor noise is excluded; only periods overlapping with interaction are utilized for downstream tasks. . . . . 31

Figure 3.7 Dual-branch feature extraction pipeline for each 150-length window. Raw sensor data is converted into two modalities: time-series features ($150 \times 18$) and FFT features ($675 \times 1$), computed independently for $x$, $y$, and $z$ axes. . . . . 32

Figure 3.8 Siamese triplet architecture with dual-branch embedding. Each sample in the triplet (anchor, positive, negative) is passed through a shared model comprising a time-series LSTM encoder and a frequency-domain FFT-MLP encoder. The resulting embeddings are fused and used to compute the triplet loss. . . . . 34

Figure 3.9 Overview of Session-Balanced Random Window Sampling (SBRWS). For each user in G1 (training) or G2 (validation), $n_1$ anchor windows are sampled. Each anchor is paired with $n_2$ positive samples from the same user and $n_3$ negative samples from different users, both sampled across sessions. . . . . 37

Figure 3.10 Example of anchor and positive window similarity within the same session. Adjacent windows (highlighted) exhibit highly overlapping patterns, leading to poor discrimination in learned representations. 40

Figure 3.11 Low-entropy sensor segment for User 0—Session `g1`. Touch events are marked in red. . . . . 41

Figure 3.12 Low-entropy sensor segment for User 26—Session `g1`. Prolonged flat regions dominate the segment. . . . . 42

Figure 3.13 Low-entropy sensor segment for User 31—Session `g1`. Despite touch activity, informative variability is absent. . . . . 42

Figure 3.14 Illustration of index-aligned triplet sampling. Windows with matching temporal indices are selected across sessions and users to form semantically comparable anchor, positive, and negative samples. This alignment fosters meaningful intra-user and inter-user comparisons. 45

Figure 3.15 Relative Position-Based Sampling. Triplets are constructed by selecting anchor, positive, and negative windows with similar relative temporal positions ($\rho$) across sessions. This strategy allows flexible alignment even when session lengths vary. . . . . 47

Figure 3.16 Inference mechanism via window-aligned ensembling. For a given session pair, windows are aligned (e.g., by index), and pairwise embedding distances are computed. Final similarity is based on the mean distance across all aligned pairs. . . . . 48

Figure A.1 Sensor sequence lengths before frequency downsampling . . . . 79

Figure A.2 Sensor sequence lengths after frequency downsampling . . . . . 80

## LIST OF ABBREVIATIONS

| | |
|---|---|
| 3D | 3 Dimensional |
| AUC | Area Under the ROC Curve |
| EBFSWS | Entropy-Based Filtering with Session-Weighted Sampling |
| EER | Equal Error Rate |
| EMG | Electromyography |
| FFT | Fast Fourier Transform |
| GPS | Global Positioning System |
| LSTM | Long Short-Term Memory |
| MLP | Multilayer Perceptron |
| PPG | Photoplethysmography |
| ROC | Receiver Operating Characteristic |
| SSA | Singular Spectrum Analysis |
| SBRWS | Session-Balanced Random Window Sampling |
| SVM | Support Vector Machine |
| SWRWS | Session-Weighted Random Window Sampling |

# CHAPTER 1

# INTRODUCTION

Biometric authentication has gained significant attention in information security research for its potential to enable accurate and unobtrusive identity verification. Biometric technologies operate by recognizing individuals based on unique physiological or behavioral characteristics. While physiological biometrics include traits such as fingerprints, iris patterns, or facial features, behavioral biometrics focus on the patterns of human behavior, such as typing rhythm, gait, signature, and voice dynamics. These behavioral traits are acquired and analyzed to identify or verify users through their distinctive interaction patterns with devices, making them a natural fit for continuous and implicit authentication systems [1]. Figure 1.1 provides a visual overview of this process, highlighting the flow from sensor data acquisition on mobile devices to the analysis and decision-making stages used to authenticate user identity based on behavioral patterns.

Behavioral biometrics have been widely explored for various security applications, ranging from online banking and mobile device protection to workplace authentication and access control. Among behavioral biometrics, modalities such as keystroke dynamics, touch gestures, gait analysis, and voice patterns have been employed for both verification and identification tasks [1, 2]. Unlike static passwords or tokens, behavioral biometrics do not require users to memorize specific information or carry physical tokens, providing a seamless and non-intrusive authentication experience [3]. Furthermore, behavioral traits are inherently difficult to replicate, which adds an additional layer of security against impostors.

A significant branch of this field is focused on behavioral biometrics for mobile and continuous authentication. In mobile contexts, individuals generate rich streams of

behavioral data through interactions such as screen touches, typing gestures, motion signals, and device handling patterns. Modern smartphones are equipped with multiple sensors including accelerometers, gyroscopes, magnetometers, and touchscreens, enabling the capture of multimodal behavioral data streams in real time [4]. This sensor data can be processed to develop user-specific behavioral profiles that support ongoing authentication throughout device usage rather than limiting verification to the point of entry. Such continuous authentication methods help to ensure that the authenticated user remains in possession of the device, reducing risks of unauthorized access after initial login [4].

Although behavioral biometrics offer considerable potential, unresolved issues concerning security and privacy continue to pose substantial challenges. Recent surveys highlight the risks of spoofing, replay attacks, and biometric data leakage, particularly when behavioral data is not adequately protected [5]. In addition, the trade-off between usability, security, and privacy is not trivial in biometric systems. Continuous authentication introduces additional complexity due to the dynamic nature of behavioral traits, which may vary over time or across contexts. Developing systems that maintain user privacy, operate reliably, and manage resources effectively continues to be a subject of ongoing investigation. [5].

Various behavioral biometric systems have been proposed that utilize different interaction scenarios. Some methods focus on fixed tasks such as typing a password, while others consider free-text typing, navigation gestures, or natural usage patterns [1, 2]. Continuous authentication systems may analyze multiple temporal windows of sensor data to dynamically model user behavior. However, capturing high-quality discriminative segments from long streams of behavioral data remains a challenge. Random sampling windows may include irrelevant or noisy information, leading to suboptimal model performance.

This thesis addresses the challenge by introducing an adaptive approach to window sampling and filtering within the context of continuous mobile behavioral authentication. The objective is to select the most informative and user-representative segments from sensor and touch data streams to improve both model training and inference, ultimately enhancing the accuracy and robustness of biometric authentication systems.

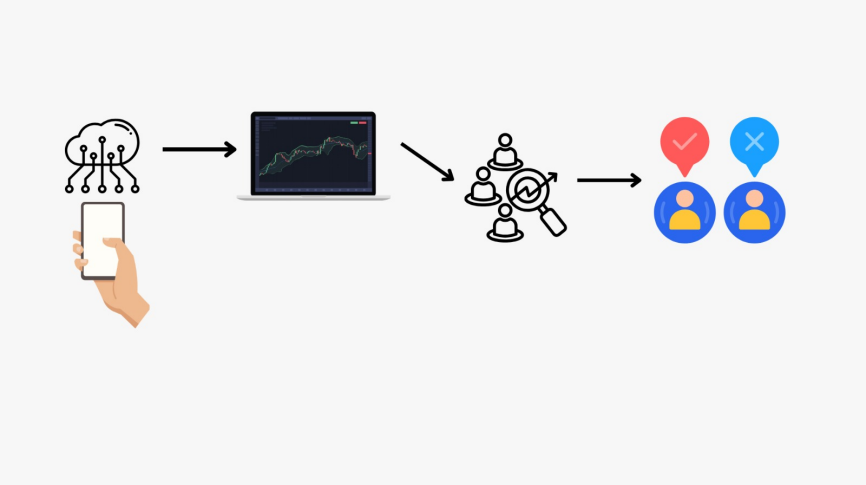


Figure 1.1: An illustration of the behavioral biometric authentication process. The visual depicts how sensor data from mobile devices are collected, processed, and analyzed to verify or reject user identity based on interaction patterns.

## 1.1 Research Questions

In mobile behavioral biometric systems, sensor modalities such as accelerometers continuously generate dense, time-series data that capture implicit patterns of user movement. However, directly utilizing all collected data without selection may result in suboptimal learning, increased computational cost, and decreased model generalization. In particular, for accelerometer-based continuous authentication, the quality and informativeness of sampled input windows critically influence model performance. Therefore, rather than assuming all sensor data to be equally valuable, careful window-level selection and filtering must be applied to ensure that only discriminative and representative segments are used during training and inference.

This thesis investigates adaptive strategies for data sampling and pair formation to address these challenges. The research is guided by the following questions:

**RQ1 How can behavioral biometric data be selectively sampled to maximize its discriminative value while minimizing redundancy?**

This question explores mechanisms for isolating relevant and representative

segments from continuous behavioral data streams. The aim is to develop data selection techniques that improve the efficiency and effectiveness of biometric authentication models by focusing on segments that carry the most identity-relevant information.

**RQ2 How can comparison pairs in biometric authentication models be constructed to meaningfully capture behavioral similarities and differences within the same user and between distinct user profiles?**

Authentication models based on similarity learning, such as Siamese networks, typically require constructing pairs of data segments that represent either the same individual or different individuals. This question investigates how to organize these pairs using structured and context-aware selection strategies instead of relying on random pairing. The objective is to improve the model's ability to learn genuine behavioral similarities and distinguish them from inter-user variations by using logically selected comparison pairs that capture consistent and realistic patterns of user behavior.

These research questions focus on improving adaptive data use in continuous behavioral biometric authentication, leading to more efficient and reliable systems.

## 1.2 Contributions of the Study

Recent works emphasize the importance of data efficiency and feature selection in behavioral biometrics [7]. Similarly, Stragapede et al. [8] highlight the need for careful management of multimodal behavioral data in continuous authentication, where the combination of touch and background sensor signals can introduce complexity and require strategies for robust feature extraction and efficient data selection. Their findings stress the importance of processing data segments that are contextually meaningful and relevant to the authentication task.

The contributions of this study are as follows:

- Proposing an adaptive filtering and sampling framework to select the most relevant segments of accelerometer-based behavioral biometric data collected from

mobile devices. This strategy reduces redundancy and prioritizes segments that are expected to be more informative for identity discrimination.

- Introducing a method for forming comparison pairs in behavioral biometric authentication. Rather than relying on random segment combinations, pairs are selected using temporal and contextual criteria to better represent realistic intra- and inter-user differences. This approach improves training and verification performance in similarity-based models such as Siamese networks.

- Presenting a comprehensive experimental pipeline that applies the proposed adaptive selection methods to the BehavePassDB dataset [6]. The dataset includes multiple behavioral modalities collected from smartphones across diverse interaction tasks and sessions, providing a challenging benchmark for evaluating continuous behavioral authentication systems under realistic conditions. In this context, the study demonstrates that constructing training and testing pairs with attention to session variability and task structure can substantially improve system performance, even when focused on a single modality.

In addition to these contributions, this study addresses a broader challenge in behavioral biometrics, which is optimizing the use of collected data. Conventional training pipelines often rely on large, unfiltered datasets, where many segments contribute little to authentication accuracy. By prioritizing relevant and contextually meaningful data, the proposed approach aims to improve both model efficiency and overall performance. Beyond the development of adaptive sampling and structured pair formation methods, this study also contributes to the broader goal of optimizing data utilization in biometric authentication systems. In traditional model training pipelines, large amounts of behavioral data are collected, but not all segments are equally informative or relevant for identity verification. Random sampling or unfiltered use of raw sensor data can introduce noise, redundancy, and inconsistencies, ultimately limiting model performance.

## 1.3 Organization of the Thesis

The organization of the thesis is as follows.

Chapter 2 reviews the related literature. It covers behavioral biometrics, biometric authentication, keystroke dynamics, and sensor-based methods. Preprocessing, filtering techniques, and decision strategies are discussed. The chapter ends by identifying the research gap. Chapter 3 describes the methodology. It explains the adaptive data sampling strategy and filtering techniques. It also details how informative segments are selected and how comparison pairs are formed. The dataset used is introduced. Chapter 4 presents the experimental results. It reports the performance of the proposed methods under different conditions. Comparisons with baseline approaches are provided. Chapter 5 summarizes the study. It discusses the main findings and contributions. The chapter concludes with limitations and suggestions for future work.

# CHAPTER 2

# RELATED WORK

This chapter provides a comprehensive overview of the existing literature on behavioral biometric authentication, with a particular emphasis on methods, data sources, and sensor modalities relevant to mobile environments. The review begins by tracing the methodological evolution of behavioral authentication systems and examining the diversity of tasks explored in prior research. It then discusses key public datasets used for evaluation, followed by focused analyses of keystroke dynamics and sensor-based modalities as prominent sources of behavioral data. The chapter also highlights important challenges in preprocessing and segmenting continuous interaction signals, reviews decision and fusion strategies for integrating multiple modalities, and concludes by identifying the research gap that motivates the methodological contributions of this thesis.

## 2.1 Behavioral Biometric Authentication: Methodological Evolution and Task Diversity

The development of behavioral biometric authentication systems initially focused on statistical modeling applied to controlled tasks such as PIN entry and virtual typing. Early research aimed to strengthen traditional authentication mechanisms by incorporating behavioral traits to improve security without compromising usability.

Feng et al. [9] proposed TAP (Typing Authentication and Protection), a system designed for mobile devices that captures virtual key typing dynamics during both login and post-login phases. The system collects features such as keypress duration, inter-keystroke timing, and touch pressure data. Their dataset was obtained through

a user study involving 40 participants, who entered controlled text inputs on Android devices using a virtual keyboard. For feature extraction, they included 40 frequently used key combinations, 41 key holding time features, and 41 pressure features—evaluated under three conditions: (i) time-only, (ii) time + pressure, and (iii) time + pressure with haptics feedback. Classification was performed using Random Forest (RF), Bayes Net (BN), and J48 decision tree classifiers. Results showed that using both time and pressure features substantially improved performance; for instance, Bayes Net achieved a False Reject Rate (FRR) of 10.2% and a False Accept Rate (FAR) of 16.9% in the login scenario. These findings underscore the importance of incorporating pressure-based features to enhance identity discrimination in keystroke dynamics.

Li et al. [10] introduced an active authentication method based on application usage profiling, where behavioral patterns were monitored over time using mobile usage logs collected from real users. The system extracted time-stamped features such as frequency and duration of application launches, transitions between apps, and user interaction timelines. A Bayesian framework with exponential smoothing was applied to capture temporal behavior consistency and reduce false alarms due to natural behavioral variability. Evaluations on a dataset collected over one week from 35 users showed that their model achieved continuous verification with an Equal Error Rate (EER) of 9.8%, demonstrating the feasibility of passive, usage-based profiling for mobile authentication.

At the system level, Song et al. [11] investigated user behavior through low-level event monitoring on desktop systems, including process creation, registry access, and file operations. Their data was collected over multiple weeks from 18 users using the RUU sensor installed on personal Windows machines. A total of 18 behavioral features were extracted in two-minute windows, including counts of unique processes, registry queries, and file touches. Fisher's criterion was used to rank and select the most discriminative features, and Gaussian Mixture Models (GMMs) were trained per user to represent behavioral fingerprints. Their system achieved strong user identification performance with AUC scores exceeding 0.85 for the majority of users, showing that carefully selected system-level features can effectively support continuous authentication under realistic conditions.

Touch-based authentication was also explored in numerous studies. Frank et al. [12] presented *Touchalytics*, a system analyzing swipe and tap dynamics for continuous authentication on smartphones. The dataset was collected from 41 users performing predefined tasks such as reading, image viewing, and short-form text writing on Android devices over multiple sessions. From each touch interaction, 30 features were extracted, including stroke direction, speed, acceleration, pressure, and inter-stroke timing. These features were computed for each swipe and aggregated over time windows to serve as input to classifiers such as k-Nearest Neighbors, Support Vector Machines, and Naive Bayes. Their approach achieved Equal Error Rates (EERs) between 4% and 5% in controlled scenarios. However, performance degraded when training and testing occurred across different sessions, highlighting the challenges of temporal behavioral variability and session-invariant feature representation in mobile touch-based biometrics. In addition, Gündüz et al. [13] proposed a feature distribution-based framework for touch-based authentication that utilizes Kullback-Leibler divergence and CNN-based Siamese networks to model and compare touch gesture distributions across sessions. Their approach emphasizes the discriminative power of distributional representations over handcrafted features.

Additionally, Feng et al. [9] extended virtual keyboard authentication by incorporating pressure sensors alongside timing features. Their results demonstrated EERs between 5% and 10%, depending on the classifier and feature combination, highlighting the potential of continuous virtual typing monitoring for mobile security.

Later, Nerini et al. [14] proposed an augmented PIN authentication method using motion sensor data combined with anomaly detection techniques. In their approach, smartphone movements were recorded during each keystroke using six motion-related sensors (accelerometer, gyroscope, linear acceleration, gravity, rotation vector, and orientation), capturing 17 distinct features per digit entry. These feature vectors were used to train unsupervised anomaly detection models, including PCA, Kernel PCA, One-Class SVM, and Local Outlier Factor. The anomaly scores from each digit were aggregated across the full PIN input to form a final decision. Experiments showed that their system achieved EERs as low as 4% for 6-digit PINs, with low latency suitable for real-time authentication.

These studies collectively demonstrate how behavioral biometrics have evolved from PIN and typing-based authentication to include richer data sources and statistical anomaly detection frameworks. The comparison of task types and methodologies highlights the trade-offs between security, usability, and computational efficiency.

As the field of behavioral biometric authentication matured, researchers began to recognize the limitations of statistical methods, particularly regarding scalability, session variability, and long-term usability. This motivated a gradual transition towards deep learning-based approaches, especially for unconstrained and continuous scenarios on mobile and wearable devices.

A major development during this transition was the systematic comparison of tasks and the assessment of different modalities. Li et al. [15] provided a comprehensive evaluation of PIN- and pattern-based behavioral biometric authentication. By developing a unified framework, they demonstrated that both PIN and pattern tasks could achieve comparable accuracy when combined with behavioral signals, using Dynamic Time Warping (DTW) and histogram-based verification methods. Their results showed histogram-based approaches to be more robust across sessions, addressing template aging issues.

The movement towards passive and continuous monitoring was further accelerated by multimodal data fusion. Studies such as Stragapede et al. [8] and Meng et al. [16] highlighted the integration of touchscreen data with background sensor signals (e.g., accelerometer and gyroscope), reporting Equal Error Rates (EER) between 4% and 9% over 3-second windows. These multimodal approaches improved robustness to session variability and increased system resilience against spoofing.

Deep metric learning has also become central in modern behavioral biometric systems. Several works have utilized Siamese and triplet networks to improve pairwise comparison and reduce reliance on large labeled datasets [17, 18, 19]. Wang et al. [17] introduced a Siamese network framework for touch and motion data, addressing small sample size problems, while Lis et al. [19] demonstrated effective keystroke dynamics-based verification using partial password inputs.

The integration of physiological and behavioral biometrics has further broadened the

authentication landscape. Choi [20] proposed an FPGA-optimized Siamese network for electromyogram (EMG)-based authentication in wearable devices, achieving 94% accuracy with minimal hardware resources. Bah et al. [21] explored combining PIN, fingerprint, and facial recognition into a unified two-level authentication pipeline for improved security in banking applications.

Several survey papers have analyzed these shifts. Sundararajan and Woodard [22], Minaee et al. [23], and Gruschina et al. [24] collectively emphasize that deep learning allows for automatic feature extraction, reducing the dependency on handcrafted features prominent in early works.

These advancements represent a clear methodological shift from task-constrained statistical approaches to data-driven adaptive learning systems. Modern frameworks use deep neural networks to handle diverse behavioral signals in both constrained and unconstrained environments, paving the way for scalable and secure continuous authentication systems.

Recent advancements in behavioral biometric authentication have introduced transformer-based models to address limitations in prior deep learning approaches. These models capture long-range temporal dependencies and better model multimodal sensor data, improving authentication accuracy in continuous scenarios. Nguyen et al. [25] proposed BehaveFormer, a dual-attention transformer that achieved state-of-the-art results on datasets such as HuMIdb and FETA, with EERs as low as 2.95% for keystroke and 3.67% for swipe dynamics. Hu et al. [26] developed AuthConFormer, combining convolutional layers with transformers to reduce computational complexity while achieving EERs near 1% on HMOG and IDNet datasets. Similarly, Zhao et al. [27] presented a spatio-temporal attention transformer with positional encoding for behavioral biometrics, reporting performance gains of over 25% compared to traditional LSTM and CNN-based methods. These approaches mark the latest evolution in the field, setting new benchmarks for robust and scalable continuous authentication.

## 2.2 Datasets for Behavioral Biometric Authentication

Several benchmark datasets have been developed to advance research in behavioral biometric authentication, varying in task types, data modalities, and collection scenarios. Early datasets such as Touchalytics [12] primarily focused on touch interactions, capturing controlled swipe gestures for continuous authentication using touchscreen data.Several recent studies have utilized datasets such as BehavePassDB [6], HuMIdb [8], and UMDAA-02 [28], which include multi-modal features by combining touch inputs with background sensor data such as accelerometer, gyroscope, and magnetometer, as well as contextual signals like GPS and Wi-Fi in some cases. These datasets incorporate a wider range of user activities, including free-text typing, scrolling, tapping, device picking, and phone holding, allowing for both task-specific and unconstrained continuous authentication research. Additionally, HMOG [29] introduced micro-movement analysis through inertial sensor data during tap and swipe actions, while BBAS [30] contributed data relevant to adversarial robustness, focusing on perturbation-resilient behavioral modeling. Collectively, these datasets enable comprehensive evaluation of authentication models, supporting comparisons across sensor modalities, task complexity, and real-world usage conditions.

## 2.3 Keystroke Dynamics in Behavioral Biometrics

Keystroke dynamics, a form of behavioral biometrics, has emerged as a compelling modality for user authentication on mobile devices. It captures the unique typing patterns of individuals—such as key press durations, inter-key latencies, and typing rhythms—offering a non-intrusive, cost-effective, and hardware-independent solution for continuous authentication [2, 31]. Unlike physiological biometrics (e.g., fingerprint or face recognition), which often require specialized sensors, keystroke dynamics can be seamlessly captured through the touchscreen or physical keyboards that users already interact with.

The decision to focus on keystroke dynamics in this thesis is grounded in both practical advantages and strong empirical performance. Numerous studies have demonstrated the modality's robustness across authentication contexts. For example, Mon-

rose and Rubin [31] highlighted the naturalness and uniqueness of typing patterns, likening them to handwritten signatures in their distinctiveness. More recent works have adapted these principles to mobile environments. Acien et al. [32] introduced TypeNet, a deep learning framework for mobile keystroke dynamics, which achieved competitive accuracy using only passive touchscreen interaction. Similarly, Buschek et al. [33] emphasized the usability of keystroke-based systems on mobile touchscreen devices, demonstrating their potential for deployment in real-world settings.

Comparative analyses consistently affirm the superiority of keystroke dynamics over alternative touch-based tasks such as swiping, tapping, or drawing gestures. In the benchmark study by Stragapede et al. [8], which used the HuMIdb dataset and RNN models trained with triplet loss, each touch modality including keystroke, swiping, and tapping, was processed through fixed-length windows with a duration of three seconds. Features extracted for each window included raw sensor values, their first- and second-order derivatives, and frequency-domain components via FFT. Keystroke tasks achieved the lowest EER among touch-based modalities, surpassing swiping and tapping both in fixed-text and free-text scenarios. In fact, while swiping gestures yielded an EER of approximately 19%, keystroke-based authentication achieved an EER as low as 9.2% for touchscreen keyboards, even in challenging conditions involving multiple devices and sessions. This advantage is also mirrored in the MobileB2C competition results, where keystroke dynamics consistently provided higher discriminative power in both random and skilled impostor settings compared to other tasks like tapping or gallery swiping [34].

Moreover, the stability and reproducibility of keystroke dynamics contribute to its effectiveness. Unlike gestures that may vary significantly depending on screen size, orientation, or context of use, typing behavior—especially in fixed-text tasks—remains relatively invariant. This makes keystroke dynamics more amenable to longitudinal studies and cross-session authentication. For instance, the BehavePassDB dataset [6] includes four sessions spaced at least 24 hours apart and demonstrates that keystroke-based features preserve a strong discriminative profile across time and device configurations, particularly when paired with sequence models like LSTMs. Additionally, its capacity to implicitly capture user identity, even under gesture variation and environmental changes, further reinforces its robustness and adaptability [35, 36]. These

factors collectively support its prioritization over other mobile touch-based behavioral traits in both experimental setups and practical applications.

## 2.4 Sensor Modalities in Behavioral Biometrics

In behavioral biometrics for mobile and wearable authentication, sensor-based modalities have gained prominence due to their unobtrusiveness and ability to continuously monitor user behavior without requiring explicit input. Commonly utilized sensors include accelerometers, gyroscopes, magnetometers, and orientation sensors. These sensors capture diverse physical movements, making them valuable for user profiling and identity verification. Among them, the accelerometer stands out for its high responsiveness to user-specific motion patterns and its ubiquitous integration in modern smartphones and wearables [4, 37]. Accelerometer data has been extensively employed in activity recognition tasks due to its ability to capture subtle motion patterns in wearable and mobile settings. Prior studies have demonstrated that body-worn accelerometers can effectively distinguish between a range of daily physical activities such as walking, standing, sitting, lying, and transitional movements. Gao et al. [38] conducted a comprehensive evaluation of activity recognition using accelerometer signals collected from various body positions, comparing single-sensor and multi-sensor configurations. Their results highlighted that, while multi-sensor setups can marginally enhance accuracy, a well-placed single accelerometer—particularly on the waist or wrist—is sufficient to capture informative motion cues for most classification tasks. These findings underscore the utility of accelerometer data as a primary sensing modality for behavior understanding, making it a natural candidate for biometric modeling based on mobile device interactions.

The decision to prioritize accelerometer data in this thesis is informed by both its demonstrated utility and comparative advantage over other modalities. Accelerometers have consistently shown superior performance in capturing nuanced behavioral traits, particularly in gait and gesture-based authentication. For instance, Mekruksavanich and Jitpattanakul [39] present a hybrid deep learning model using SE-DeepConvNet, where accelerometer-driven features achieve a 0% Equal Error Rate (EER) and 100% accuracy in several USC-HAD tasks, outperforming gyroscope and mag-

netometer counterparts. Similarly, Sitová et al. [29] in the HMOG dataset study illustrate the strong discriminative power of accelerometer data under both sitting and walking conditions.

Moreover, comparative studies affirm the accelerometer's effectiveness. For example, Goldstein et al. [40] and Killourhy and Maxion [41] show that sensor combinations involving accelerometers yield better biometric authentication results than those relying solely on gyroscope or magnetometer inputs. In [42], gyroscope signals were found to be more susceptible to noise and orientation variance, while accelerometer data preserved consistency across sessions and device placements. Finally, the accelerometer's energy efficiency and robustness in dynamic settings make it particularly suitable for real-time, continuous authentication [39, 43].

Given these advantages—ranging from discriminative capacity to environmental robustness—the use of accelerometer data in this thesis is both empirically justified and practically motivated. The next section will present a more focused comparison between accelerometer-based methods and those relying on alternative sensors, reinforcing its superiority within behavioral biometric systems.

## 2.5 Preprocessing and Segmentation Challenges in Sensor-Based Behavioral Biometrics

Preprocessing is a fundamental step in sensor-based machine learning systems, serving to improve signal quality, remove noise, and facilitate the extraction of meaningful features. In classical signal processing pipelines, various techniques such as wavelet transforms, low-pass and band-pass filters, and time-domain normalization have long been used to improve data quality and reduce redundancy [44, 45]. These operations help attenuate high-frequency noise, smooth signal fluctuations, and standardize the dynamic range of sensor recordings. In particular, wavelet-based methods have proven effective for denoising and trend removal in biosignal applications such as EEG or EMG [45]. However, while such techniques have matured in physiological contexts, their direct application to behavioral sensor signals presents distinct challenges due to the spontaneous and unstructured nature of human interactions.

In behavioral biometrics, raw data collected from mobile sensors often exhibit session-wise variability, sampling inconsistencies, and movement-induced artifacts. These factors necessitate preprocessing strategies that are not only robust but also adaptive to contextual changes. As highlighted in recent surveys [46, 47], standard preprocessing operations often fall short in deep learning pipelines when the data exhibits long durations, uninformative intervals, and inconsistent user engagement. These limitations become particularly pronounced when preprocessing precedes neural modeling, as inappropriate filtering or segmentation can suppress discriminative traits critical for user authentication.

A key difficulty in behavioral biometric preprocessing lies in selecting or isolating the most meaningful segments from lengthy sensor recordings. Unlike structured physiological signals, behavioral data often lack clear periodicity or task-aligned boundaries, which complicates windowing. Abdulrahman and Alhayani [48] highlight that segmentation quality significantly affects biometric system performance, particularly in dynamic modalities such as signature and EEG. They emphasize that naive fixed-size windowing may include irrelevant or noisy portions, degrading recognition accuracy. As a solution, the authors advocate for data-specific preprocessing pipelines, recommending adaptive segmentation techniques tailored to the modality's temporal dynamics. For EEG-based systems, they suggest entropy-aware filtering and wavelet-based signal decomposition to isolate cognitively salient regions, while for signature verification, they propose hybrid feature extraction over time-normalized strokes. These perspectives reinforce the view that segmentation is not a mere preprocessing step but a fundamental design choice in behavioral biometric pipelines. Motivated by these insights, this thesis investigates data-driven strategies for selecting informative windows in unconstrained sensor environments.

## 2.6 Decision and Fusion Strategies in Behavioral Biometrics

Decision-level strategies and ensemble approaches have been widely adopted in behavioral biometric systems to boost performance by combining complementary information sources. Ensemble methods span various fusion levels, including feature-level fusion, score-level fusion, and decision-level fusion [49]. In feature-level fusion,

raw or processed features from multiple sources are concatenated before classification; score-level fusion combines matching scores output by different models; and decision-level fusion aggregates final authentication decisions (e.g., majority voting) from multiple classifiers to make a robust prediction. Score-level fusion is particularly appealing in biometric authentication because it balances flexibility and performance without requiring strict feature alignment. Common strategies include weighted score summation, majority voting, and max-score rules. For instance, Nanni et al. [50] analyzed the effect of combining biometric matchers through simple rules, observing consistent improvements in both verification and identification tasks. Similarly, Jena et al. [51] explored deep neural ensemble fusion for smartphone-based biometrics, demonstrating that combining modalities like face and voice yielded significantly better results than single-modal systems.

Beyond modality-level fusion, recent works have also emphasized intra-modal ensemble strategies, where multiple features or outputs derived from the same sensor or task are fused. Cheng et al. [52] investigated multiple-sample fusion, showing that aggregating matching scores from repeated biometric samples strengthens robustness against noise and variability. Furthermore, the work by Alshazly et al. [53] showed that employing a committee of classifiers trained on different biometric characteristics or session views of the same input can improve accuracy under constrained data settings. As noted by Bermejo et al. [7], hybrid feature selection methods that combine filter and wrapper approaches are especially useful in balancing computational efficiency with performance gains. These methods reduce the dimensionality of data by focusing on features that most significantly contribute to the task, avoiding the costs of processing irrelevant or redundant inputs.

A complementary approach to classical ensemble learning is the exploitation of temporal or structural segmentation in behavioral signals—particularly in mobile and wearable systems. These systems often process long and noisy time series, where different segments of the signal (e.g., windows) may carry varying degrees of identity-discriminative information. Rather than making a decision on a single global representation, recent research has proposed post-inference strategies that aggregate local decisions from multiple windows. For example, in a PPG-based wearable authentication system, Zhao et al. [54] divided each session into multiple non-overlapping

short windows and performed identity verification on each individually, followed by session-level decision fusion through averaging of the verification scores. This approach improved authentication robustness under physiological variability by mitigating local signal noise and transient distortions. Likewise, Neiswanger and Xing [55] introduced a post-inference prior swapping mechanism in probabilistic models, allowing posterior distributions to be recomputed or adjusted after the initial inference step by reassessing the prior assumptions. While this method was not originally proposed for biometric authentication, the core idea of deferring final judgment and refining global predictions based on multiple local posteriors holds promise in biometric settings where noisy and heterogeneous windows are common. Moreover, Wahid et al. [56] proposed a multi-window majority voting strategy for EMG-based gesture recognition, where each window was independently classified using a trained model and final gesture classification was based on the most frequent class among window-level outputs. Their study demonstrated that such an approach increased recognition accuracy, especially in the presence of inconsistent or ambiguous segments. These methods highlight a growing recognition of the value in preserving and intelligently combining fine-grained temporal patterns rather than collapsing them into a singular embedding. Therefore, emphasizing the importance of window selection, structuring, and fusion constitutes a crucial frontier for improving decision reliability in behavioral biometrics—especially in the context of mobile and sensor-driven modalities. Concretely, similar window-level aggregation appears across modalities: in smartphone gait, IDNet fuses decisions from consecutive walking cycles in a multi-stage framework to reach low error with few cycles [57]; in free-text keystroke dynamics, TKCA performs majority voting over multiple short sequences to obtain the final decision under uncontrolled settings [58]; and within-device motion biometrics during form filling, Ray et al. fuse $k$ consecutive window scores (averaging or likelihood-ratio fusion) before a session-level verdict, reducing EER [59]. Closely related evidence from EMG gesture recognition shows that enlarging the window and the number of votes systematically improves accuracy [56], reinforcing the design choice to treat windowing and post-inference fusion as first-class components. In our system, this paradigm is operationalized by treating each window as an independent hypothesis and deferring the final accept/reject to a post-inference score/decision ensemble, with window length/stride and the number of aggregated windows selected

on validation; this mirrors best practice reported across modalities and directly targets robustness to transient artifacts and session variability.

## 2.7 Research Gap

Existing literature in behavioral biometrics has explored a wide array of feature engineering strategies, sensor modalities, and classification techniques. Many recent works have focused on fusing features from multiple modalities [8, 60], employing ensemble learning at the feature or score levels to improve robustness.

Moreover, the integration of adaptive triplet selection algorithms, such as semi-hard or distance-weighted sampling, could further improve training by dynamically selecting more informative and challenging negative examples. These approaches could replace or complement the fixed sampling strategies introduced in this thesis. Additionally, since the dataset used in this work does not specify the contextual conditions under which users performed the biometric tasks (e.g., sitting, walking, or transitioning), future work could incorporate an activity recognition phase prior to or alongside the authentication pipeline. Recognizing the underlying physical activity may reveal patterns or confounding factors that influence sensor dynamics and authentication performance. Accelerometer-based activity recognition has proven effective in classifying such conditions across wearable settings [38], and its integration could provide a deeper understanding of behavioral variability during mobile interactions.

A key gap in the literature lies in the underexplored potential of window-level fusion for decision making. That is, while segment-wise or windowed data have been used in training or evaluation, the idea of combining multiple sub-segments at the inference stage—what may be termed post-inference window ensemble has not been systematically studied or optimized. Recent studies such as those by Neiswanger et al. [61] and Zhao et al. [54] show the benefit of posterior aggregation from model outputs, but they do not investigate strategies for selecting or combining temporal sub-units of a signal for a more informed final decision.

Despite extensive use of window-based processing in sensor-driven biometric systems, a methodological gap remains in identifying which segments are most infor-

mative and how to aggregate them meaningfully at the decision level. Most existing studies treat windows independently or rely on simple score averaging without prioritizing semantically rich segments. Zhao et al. [54], for example, applied majority voting across short PPG windows to improve authentication robustness, yet without an explicit mechanism for prioritizing the most discriminative regions. In contrast, this study proposes a task-aware and temporally guided ensembling framework that emphasizes selecting and aggregating informative sensor windows for robust decision-making.

# CHAPTER 3

# METHODOLOGY

This chapter presents the methodological framework developed for continuous mobile authentication using touch-aligned sensor data. The process begins with the temporal alignment and segmentation of multimodal sensor recordings into fixed-length windows, which serve as the basis for all subsequent analysis. These windows are then transformed into complementary feature representations and processed by a deep learning architecture designed to capture both temporal and frequency-domain patterns. Triplet construction is guided by alternative sampling strategies, some of which incorporate filtering and temporal constraints to improve discriminative capability. The evaluation procedure employs an ensemble-based matching approach to enhance robustness under behavioral variability. The chapter concludes with a description of the baseline model used for comparison.

## 3.1 Dataset Description

This study utilizes the BehavePassDB dataset [6, 8], a publicly available benchmark specifically designed for mobile behavioral biometrics research. The dataset was developed in the context of the MobileB2C competition [34], aiming to promote research on continuous authentication using behavioral traits collected from mobile devices under real-world Human-Computer Interaction (HCI) conditions.

The modalities recorded include touchscreen interactions, categorized into four tasks, and five stream of background sensors captured concurrently. Table 3.1 provides a summary of the behavioral tasks included in the dataset.

Table 3.1: Behavioral tasks included in BehavePassDB.

| Task | Description |
|---|---|
| **Texting** | Participants respond to several questions including name, age, and three short free-text prompts. In sessions 1 and 4, participants use their personal keyboards, while in sessions 2 and 3, a fixed keyboard records additional features like key release time and (x, y) coordinates. |
| **Text Reading** | Participants read a short passage by vertically swiping and then answer a comprehension question, followed by a drag-and-drop gesture. |
| **Gallery Swiping** | Participants horizontally swipe through four images and answer a question related to each image, again followed by a drag-and-drop gesture. |
| **Tapping** | Participants are instructed to tap on predefined screen locations as quickly as possible, followed by a drag-and-drop gesture. |

During each task, the following five background sensors were recorded concurrently:

- **Accelerometer** – captures linear acceleration in 3D space.
- **Gyroscope** – records angular velocity across the three axes.
- **Magnetometer** – captures Earth's magnetic field in 3D.
- **Linear Accelerometer** – measures acceleration without gravity.
- **Gravity Sensor** – isolates the gravity component from accelerometer data.

Each data stream in BehavePassDB adheres to a standardized format depending on its modality:

- **Keystroke:** [timestamp, ASCII code]

- **Touch (swipe, tap):** [timestamp, x / screen width, y / screen height, action_type] — where action_type $\in$ {0: down, 1: up, 2: move}
- **Accelerometer Sensor:** [timestamp, x, y, z], where $x$, $y$, and $z$ represent linear acceleration values along the three spatial axes, and *timestamp* indicates the sampling time of each measurement.

The dataset is divided into two subsets:

- **Development set:** Contains data from 51 users, used for both training and validation purposes. Each user contributes four sessions (denoted as g1, g2, g3, and g4), captured on different days to reflect temporal variation.
- **Evaluation set:** Contains a separate set of sessions that do not include additional users beyond the development set, but provide session pairs with labels indicating whether the sessions belong to the same individual, a random impostor, or a skilled impostor attempt.

In this thesis, the development set is used to construct the training and validation partitions. All available sessions from each user in this set are utilized without relying on impostor-type annotations. The evaluation set is reserved for final testing, where the classification task is reframed as a binary decision problem: determining whether two sessions originate from the same user, irrespective of the type of impostor label provided.

Each user in the development set performs four distinct sessions (g1 through g4), spaced at least 24 hours apart. This multi-session configuration enables modeling of intra-user behavioral variability across time and supports the development of longitudinal authentication strategies.

## 3.2 Preprocessing and Segmentation of Sensor Data for User Interaction Modeling

This study focuses on two specific modalities from the BehavePassDB dataset: keystroke timing information and accelerometer signals. The choice of these modalities is

grounded in findings from recent literature. Prior work has shown that temporal features of keystroke dynamics particularly inter-keystroke intervals, offer strong user discriminability and remain relatively stable across time and device changes, even in the absence of actual key content. Similarly, among motion-based sensors, accelerometer data has consistently demonstrated high reliability and discriminative power in continuous authentication settings, often outperforming alternatives such as gyroscopes, gravity sensors, and magnetometers [8].

In addition to these empirical advantages, the structure of the dataset itself further informed the modality choices. Notably, during the keystroke task, users were prompted to respond to questions using two different keyboard types. In sessions g1 and g4, the users interacted with their own custom keyboards, while sessions g2 and g3 utilized a fixed keyboard configured by the researchers. This deliberate design choice enables richer feature extraction from the fixed keyboard, such as key press-release intervals and spatial coordinates, while also capturing natural typing behavior through the inclusion of user-preferred keyboards [6]. This variation was exploited in defining training and validation sets, enabling more rigorous cross-device generalization analysis.

Accelerometer signals were subjected to a sequence of preprocessing operations aimed at minimizing noise, correcting temporal irregularities, and ensuring consistency across samples.

Timestamps were first interpolated to obtain a uniformly spaced time base. Given a raw sequence $\mathbf{X} = \{(t_i, x_i, y_i, z_i)\}_{i=1}^{N}$, where $N$ denotes the number of original sensor samples in the sequence, the non-uniform timestamps $t_i$ were rescaled to seconds and then re-interpolated via cubic interpolation to match the minimum sequence length across all sensors within a session.

Each axis ($x$, $y$, $z$) was then standardized using global mean $\mu$ and standard deviation $\sigma$ values derived from training data:

$$x_i^{\text{norm}} = \frac{x_i - \mu_x}{\sigma_x}, \quad y_i^{\text{norm}} = \frac{y_i - \mu_y}{\sigma_y}, \quad z_i^{\text{norm}} = \frac{z_i - \mu_z}{\sigma_z}$$

These parameters were stored and reused consistently for validation and test samples to avoid data leakage.

To suppress high-frequency noise and transient signal artifacts, median filtering with a kernel size of three was applied independently to each axis. The median filter, a well-established non-linear filtering method for signal denoising, replaces each data point with the median value of its neighbors and has been shown to preserve edge sharpness better than linear smoothing filters [62, 63]:

$$x_i^{\text{filt}} = \text{median}(x_{i-1}, x_i, x_{i+1})$$

and similarly for $y$ and $z$.

As acquisition frequencies varied across devices (typically within $\{50, 100, 200\}$ Hz), adaptive downsampling was performed. Following the frequency estimation approach in [8], the mean inter-sample interval $\bar{\Delta}t$ was used to compute a downsampling factor $D$:

$$D = \begin{cases} 1 & \text{if } \bar{\Delta}t < 0.015 \\ 2 & \text{if } 0.015 \leq \bar{\Delta}t < 0.03 \\ 4 & \text{otherwise} \end{cases}$$

The variability of acquisition frequencies across different user-session pairs is visualized in Figure 3.1. Each x-axis index represents a unique user-session pair, ordered first by session (from g1 to g4) and then by user ID (from 0 to 50). As an illustrative case, at index 50 (i.e., user 50 in session g1), the estimated sampling frequency is approximately 10 Hz. This variability across sessions and users highlights the necessity of adaptive downsampling to ensure consistent temporal resolution across the dataset.

All preprocessing operations were applied consistently across training, validation, and test samples. No synthetic augmentation techniques were employed. Instead, the focus remained on preserving the integrity of real user behavior while ensuring the comparability of sensor sequences.

Another crucial step in the preprocessing pipeline involves correcting erroneous timestamp values observed in the background sensor data, particularly in the accelerometer modality. While inspecting the dataset, a previously undocumented issue was discovered: within certain short temporal intervals, the timestamps remain constant across multiple consecutive samples, suggesting an anomaly in the recording mechanism. This pattern implies a possible fault in the sensing hardware or firmware, wherein

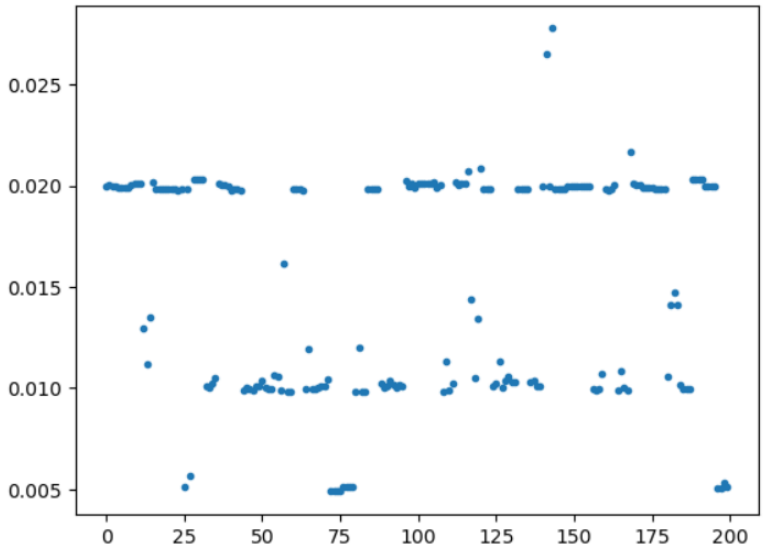



Figure 3.1: Distribution of estimated acquisition frequencies across user-session pairs in the `sensor_acc` data. The x-axis represents a sequential index over all user-session pairs, ordered first by session (g1 to g4), then by user ID (0 to 50). For example, index 0 corresponds to user 0's g1 session, index 50 to user 50's g1 session, and index 52 to user 0's g2 session. The y-axis shows the estimated sampling interval in seconds. This visualization highlights the variability in acquisition rates across sessions and users, justifying the need for adaptive downsampling.

high-frequency sensor readings are captured, but their timestamp metadata fails to update in real time.

To mitigate this, a custom correction routine was applied to reconstruct the timestamp sequence for such segments by linear interpolation. The assumption is that, although timestamps remain unchanged, the signal continues to evolve; therefore, evenly spaced synthetic timestamps were generated across the affected interval. The corrected sequence preserves the continuity of the temporal structure and allows reliable feature extraction during downstream processing. This correction process can be expressed as $\hat{t}_i = t_s + i \cdot \frac{t_e - t_s}{n-1},\ i = 0, \ldots, n-1$, where $t_s$ and $t_e$ are the starting and ending timestamps of a constant interval, and $n$ is the number of samples in that interval. The interpolation is applied blockwise over each detected constant segment.

The impact of this correction is visualized in Figure 3.2, where the top subplot illus-

trates a raw accelerometer signal segment with unchanging timestamps—notice the visible plateaus due to clustering of samples—while the bottom subplot displays the same signal after interpolation, yielding a more uniformly distributed sequence in the temporal domain.

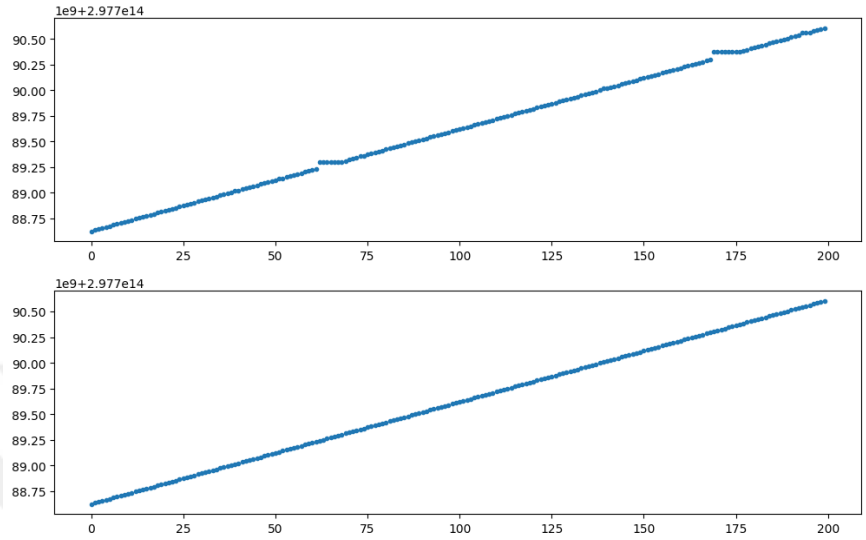



Figure 3.2: Illustration of the timestamp correction process applied to accelerometer data. The x-axis represents the interpolated time index $i$ ranging from $0$ to $n-1$, where $n$ is the total number of uniformly spaced samples within a single user-session. For example, $i = 0$ corresponds to the first accelerometer reading in the session, while $i = 52$ refers to the 53rd reading. The y-axis shows the corresponding sensor values. The top plot shows the raw signal with non-uniform timestamps, leading to visible discontinuities. The bottom plot displays the corrected version after linear timestamp interpolation.

While the original dataset documentation does not address this irregularity, prior work has acknowledged similar problems in large-scale time-series data collection and proposed general strategies for irregularity detection and cleansing. For instance, Zhang et al. [65] introduced a visual analytics framework for imaging timestamp inconsistencies and segment anomalies, while the study by Fang et al. [66] discusses timestamp realignment and temporal reconstruction in the context of mobile sensor platforms. Although these studies did not specifically apply the exact correction technique proposed here, they highlight the necessity of identifying and repairing struc-

tural irregularities prior to model training.

To address the challenges posed by variable-length sequences in behavioral biometrics, it is common practice to partition time-series data into fixed-length windows. This segmentation strategy has been widely adopted in various domains involving sensor-based analytics. For instance, Hidalgo et al. [67] proposed a flexible windowing mechanism to improve Fast Fourier Transform (FFT) measurements, while Walton et al. [68] examined the effects of sampling frequency and window size in the behavioral classification of livestock. These examples demonstrate the broader utility of windowing approaches for stabilizing temporal representation and improving downstream model performance.

In the context of this thesis, fixed-length windows of $M = 150$ samples were adopted for the accelerometer modality. This decision aligns with prior work on the BehavePassDB dataset, where the same window size was empirically validated for mobile continuous authentication [8]. Given the normalized acquisition rate of approximately 50 Hz, each window spans roughly 3 seconds of user activity, which has been shown to provide a balanced trade-off between temporal resolution and model learnability. To maintain consistency and avoid information leakage across time steps, non-overlapping windows were employed. For sequences where the total number of samples is not a multiple of the window size, the final window is zero-padded to preserve uniformity across all segments. Figure 3.3 illustrates the distribution of sequence lengths across user-session pairs. The observed variance—mean of approximately 7542 and standard deviation exceeding 3700 samples—further supports the need for segmentation, ensuring both computational tractability and consistency during training and evaluation phases.

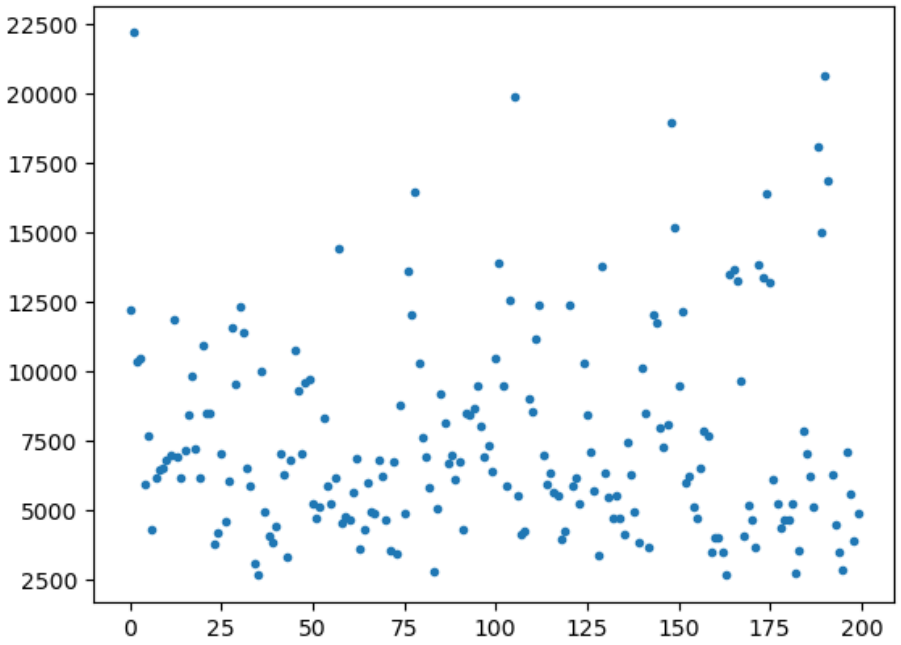



Figure 3.3: Distribution of time-series lengths for the keystroke-acceleration modality across all user-session pairs. The x-axis represents a sequential index over 204 user-session pairs, ordered first by session (from g1 to g4) and then by user ID (from 0 to 50). For example, index 0 corresponds to user 0 in session g1, index 50 to user 50 in g1, index 51 to user 0 in g2, and index 153 to user 0 in g4. The y-axis shows the number of recorded sensor samples per sequence. The long-tailed distribution and high variability necessitate fixed-length windowing to ensure input regularity and stable model convergence.

As detailed in Section 3.2, a frequency-aware downsampling procedure was applied to normalize the varying lengths of raw sensor sequences, accounting for minor acquisition rate differences across devices and sessions. This procedure, hereafter referred to as *frequency-aware downsampling*, ensures uniform temporal resolution and facilitates fixed-length window segmentation. To provide a clearer picture of the sensor sequence characteristics across sessions and users, complementary visualizations are presented in Figures A.1 and A.2, located in the appendices, which illustrate the length distributions before and after the frequency-aware downsampling procedure. These figures validate the necessity and effectiveness of the normalization step applied during preprocessing.

In practice, raw sensor recordings often span lengthy and noisy intervals that are not consistently aligned with active user input. To mitigate this, a filtering strategy was implemented to extract only those segments of background sensor data that temporally coincide with touch activity. This approach ensured that model inputs reflected meaningful interaction patterns rather than idle sensor noise or background drift.

Figures 3.4 through 3.6 illustrate representative examples of this filtering process. As seen in Figure 3.4, the raw accelerometer trace includes multiple brief bursts of activity unaccompanied by user touch, likely corresponding to spurious motion or device repositioning. These segments were excluded from analysis. In contrast, the highlighted region aligning with touch events (red dots) was retained, as it reflects intentional user interaction. Similar temporal filtering logic is demonstrated for other participants in Figures 3.5 and 3.6, where only the sensor regions surrounding active input are preserved. In Figures 3.4 through 3.6, the shaded regions correspond to local segments in the accelerometer signal where the short-term variance exceeded the 80th percentile, marking peaks of abrupt motion or potential noise. These peaks were detected across sliding windows and visualized with distinct colors for interpretability. Segments that exhibited such high variance but lacked any accompanying touch activity (e.g., purple and orange regions in Figure 3.4) were excluded from downstream analysis, as they are likely attributable to unintended movements such as device repositioning. In contrast, regions with elevated variance that temporally align with touch events (e.g., the green segment) were preserved, as they are more indicative of deliberate user interaction. This filtering strategy ensures that only contextually meaningful motion data is retained for authentication.

Such a quality-centric selection mechanism improves both interpretability and model training stability, as it removes outliers and non-interactive variance that could obscure behavioral patterns. Ultimately, this filtered alignment between sensor and touch data formed the foundation for constructing fixed-length time windows used in the embedding model.

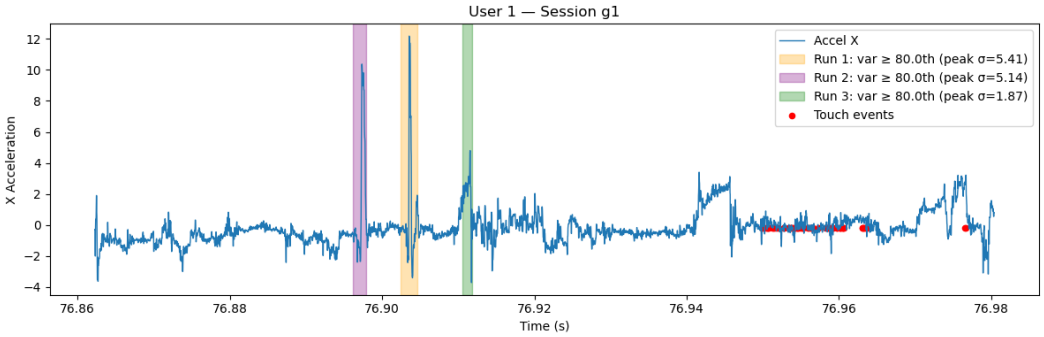



Figure 3.4: Illustration of filtering logic for User 1, Session g1. Only the sensor data temporally aligned with touch activity (red points) is retained. Three high-variance runs unrelated to touch are excluded.

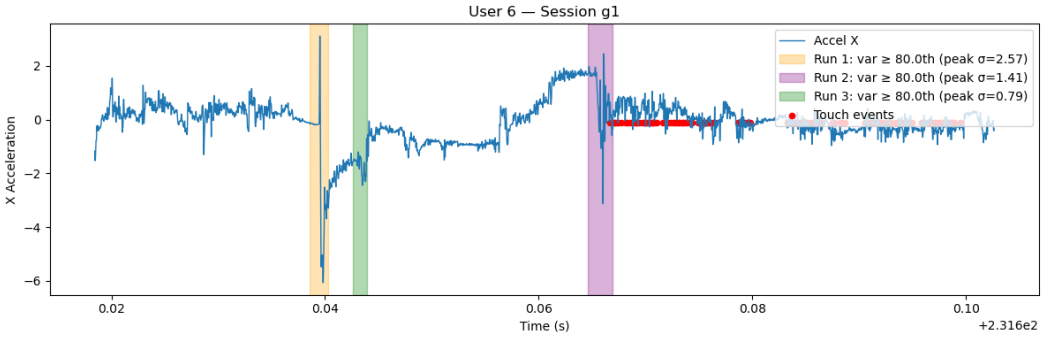



Figure 3.5: Filtered accelerometer data for User 6, Session g1. Sensor segments aligned with user interaction are preserved, improving data quality for modeling.

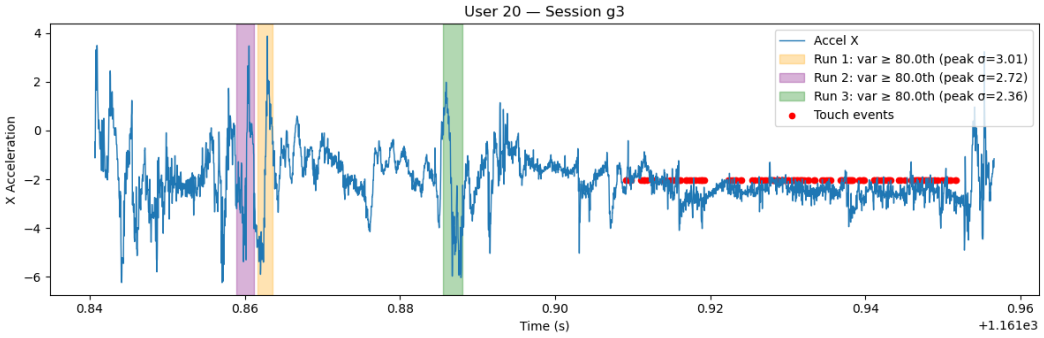



Figure 3.6: Sensor-touch alignment for User 20, Session g3. Idle sensor noise is excluded; only periods overlapping with interaction are utilized for downstream tasks.

In summary, the comprehensive preprocessing pipeline ensures that heterogeneous raw sensor recordings are standardized into temporally aligned, normalized, and fixed-length segments suitable for deep learning-based user modeling.

### 3.3 Dual-Branch Feature Extraction Strategy

Before constructing the pairwise training dataset, each windowed segment of the sensor data is transformed into two complementary feature representations to improve discriminatory capacity: a sequence-based time-series tensor and a frequency-based FFT vector. This dual-branch extraction is applied uniformly across the $x$, $y$, and $z$ axes of the accelerometer for all windows, as illustrated in Figure 3.7.

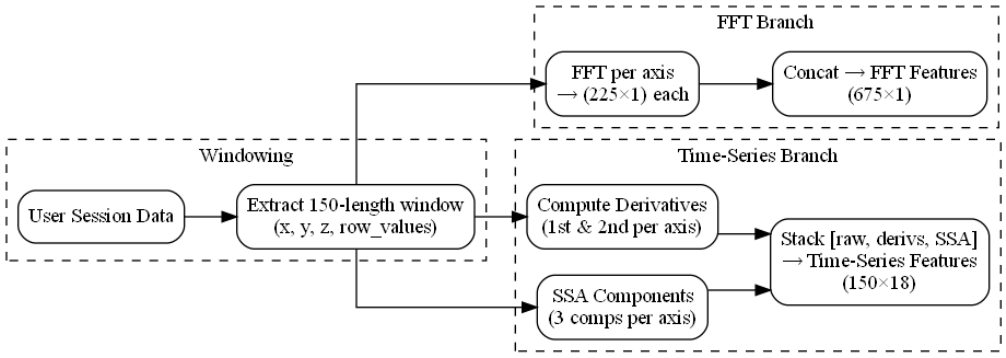



Figure 3.7: Dual-branch feature extraction pipeline for each 150-length window. Raw sensor data is converted into two modalities: time-series features ($150 \times 18$) and FFT features ($675 \times 1$), computed independently for $x$, $y$, and $z$ axes.

The frequency-domain representation is derived by applying a one-dimensional FFT independently to each axis of the sensor signal. This transformation decomposes the raw time-domain input into its constituent frequency components, yielding three primary outputs per axis: magnitude spectrum (power), phase spectrum, and frequency spectrum. Each of these components is truncated to a fixed length of 75, corresponding to the unique positive-frequency terms of the 150-point FFT. Since the FFT of a real-valued signal is symmetric, only the first half of the coefficients are retained to preserve all relevant spectral information while eliminating redundant components. The three resulting vectors per axis are then concatenated, producing

a 225-dimensional representation per axis. Finally, the complete FFT representation for the window is obtained by concatenating the axis-wise vectors for $x$, $y$, and $z$ axes, resulting in:

$$\mathbf{f}_{\text{FFT}} = \left[\mathbf{f}_x^{\text{FFT}},\ \mathbf{f}_y^{\text{FFT}},\ \mathbf{f}_z^{\text{FFT}}\right] \in \mathbb{R}^{675}$$

This formulation ensures that both high-frequency and low-frequency signal dynamics are consistently represented across the sensor axes and contributes complementary, time-independent information to the overall feature set.

In parallel, the time-series branch is designed to capture temporal dynamics and structural variations in the raw sensor input. Each window of 150 samples is transformed into a matrix of size $\mathbf{F}_{\text{TS}} \in \mathbb{R}^{150\times 18}$ by concatenating 18 distinct feature channels extracted from the $x$, $y$, and $z$ axes. These channels include:

- **Raw signal values:** The original accelerometer measurements for each axis $(x, y, z)$, yielding 3 feature channels.
- **First-order derivatives:** Velocity estimates, computed as the first temporal gradient of the signal. These derivatives highlight sudden directional changes, which often correspond to user-specific interactions.
- **Second-order derivatives:** Acceleration estimates, derived from the second temporal gradient, are indicative of the rate of change in motion intensity. These have shown to enrich temporal characterization, especially for abrupt or subtle gestures [8].
- **SSA components:** To further improve the expressiveness of the representation, Singular Spectrum Analysis (SSA) is applied individually to each axis. SSA is a data-adaptive, non-parametric technique for decomposing time series into interpretable components such as trend, periodicity, and noise [69]. Specifically, the top 3 principal components for each axis are extracted using the diagonal averaging step of SSA. These components isolate dominant temporal patterns at different scales, yielding a total of 9 channels (3 per axis). The implementation is guided by open-source applications such as the Kaggle notebook by D'Arcy [70].

As a result, the complete representation comprises:

- 3 raw channels $(x, y, z)$
- 6 derivative channels (first and second order for each axis)
- 9 SSA channels (3 components for each axis)

This yields a structured tensor of shape $\mathbf{F}_{\text{TS}} \in \mathbb{R}^{150\times 18}$, which encodes both low-level motion signals and higher-level temporal structures across time.

Building upon the extracted dual-branch features, the proposed architecture encodes each sensor window into a compact embedding using a Siamese neural network. Each data instance is organized into a triplet comprising an anchor, a positive sample from the same user, and a negative sample from a different user. For each of these samples, two distinct inputs are supplied to the network: (i) a time-series tensor $\mathbf{X}_{\text{TS}} \in \mathbb{R}^{150\times 18}$ capturing temporal dynamics, and (ii) a frequency-domain vector $\mathbf{X}_{\text{FFT}} \in \mathbb{R}^{675}$ representing global periodic patterns, as illustrated previously in Figure 3.7. These two modalities are processed by parallel branches—an LSTM encoder following the canonical formulation of Hochreiter and Schmidhuber [71] and an FFT-based MLP encoder—before being projected into a shared embedding space.

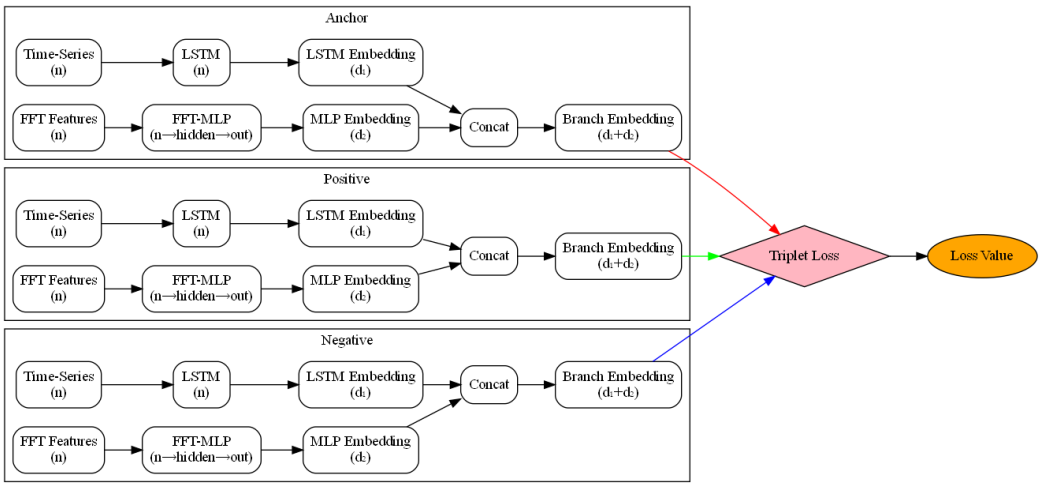



Figure 3.8: Siamese triplet architecture with dual-branch embedding. Each sample in the triplet (anchor, positive, negative) is passed through a shared model comprising a time-series LSTM encoder and a frequency-domain FFT-MLP encoder. The resulting embeddings are fused and used to compute the triplet loss.

The time-series branch applies a three-layer LSTM network, each with a hidden size of 64. The final hidden states from the forward and backward directions of the last layer are concatenated to produce a 128-dimensional representation:

$$\mathbf{e}_{\text{LSTM}} \in \mathbb{R}^{128}$$

Simultaneously, the FFT branch maps the frequency vector through a feedforward network composed of four fully connected layers (FC-1024 $\rightarrow$ FC-512 $\rightarrow$ FC-256 $\rightarrow$ FC-128), each followed by ReLU activation and dropout regularization. This yields a second embedding of identical dimension:

$$\mathbf{e}_{\text{FFT}} \in \mathbb{R}^{128}$$

The two embeddings are concatenated to form a combined representation:

$$\mathbf{e}_{\text{concat}} = [\mathbf{e}_{\text{LSTM}}, \mathbf{e}_{\text{FFT}}] \in \mathbb{R}^{256}$$

A final projection MLP (FC-256 $\rightarrow$ FC-128 $\rightarrow$ FC-64) compresses this into the final latent embedding used for comparison:

$$\mathbf{e}_{\text{final}} \in \mathbb{R}^{64}$$

The training objective is formulated using the standard triplet loss [72], which encourages embeddings from the same user to be closer than those from different users:

$$\mathcal{L}_{\text{triplet}} = \max\left(\|\mathbf{e}_a - \mathbf{e}_p\|_2^2 - \|\mathbf{e}_a - \mathbf{e}_n\|_2^2 + \alpha,\ 0\right)$$

where $\mathbf{e}_a$, $\mathbf{e}_p$, and $\mathbf{e}_n$ denote the embeddings for the anchor, positive, and negative samples respectively, and $\alpha$ is the margin (set to 0.2). Table 3.2 summarizes the architectural components and dimensional mappings.

This design, grounded in early applications of Siamese learning for verification [73], has been validated in domains such as face recognition [72] and keystroke-based authentication [74]. The current model employs both frequency and temporal features to capture user-specific interaction styles in a compact vector form.

The anchor and positive samples are selected from the same user and task, while the negative is chosen from another user. The way these triplets are generated, randomly or through hard negative mining, directly affects the discriminative capacity

Table 3.2: Summary of Siamese Model Architecture

| Input Type | Layer | Configuration | Output Dim |
|---|---|---|---|
| Time-Series ($150 \times 18$) | LSTM | 3 layers, hidden layer size = 64 | 128 |
| FFT (675) | MLP | FC-1024 $\rightarrow$ FC-512 $\rightarrow$ FC-256 $\rightarrow$ FC-128 | 128 |
| Concat | – | Concatenate LSTM + FFT embeddings | 256 |
| Fusion MLP | MLP | FC-256 $\rightarrow$ FC-128 $\rightarrow$ FC-64 | 64 |
| Output | – | Final embedding vector | 64 |

of the model. The process described enables a compact and discriminative representation of user-specific sensor patterns, optimized through triplet loss on anchor, positive, and negative embeddings. The formulation of triplet combinations and sampling strategies, essential to model training, is addressed in detail in the subsequent methodological components of this thesis.

### 3.4 Session-Balanced Random Window Sampling

To construct training and validation triplets for the Siamese network, a structured random sampling strategy was devised, referred to as *Session-Balanced Random Window Sampling (SBRWS)*. This strategy aims to balance intra-session and inter-session comparisons, while incorporating variation due to keyboard changes across sessions.

The triplet generation procedure follows three main phases: anchor selection, positive/negative pool construction, and triplet composition. The process is illustrated in Figure 3.9, and is formally defined below.

Let $\mathcal{U}$ denote the set of users in the training partition. For a given user $u \in \mathcal{U}$, let $\mathcal{S}_{G1}^{(u)}$ and $\mathcal{S}_{G3}^{(u)}$ represent the sets of windows extracted from that user's sessions G1 and G3, respectively. Likewise, let $\mathcal{S}_{G1}^{(v)}$ and $\mathcal{S}_{G3}^{(v)}$ denote windows from other users $v \neq u$.

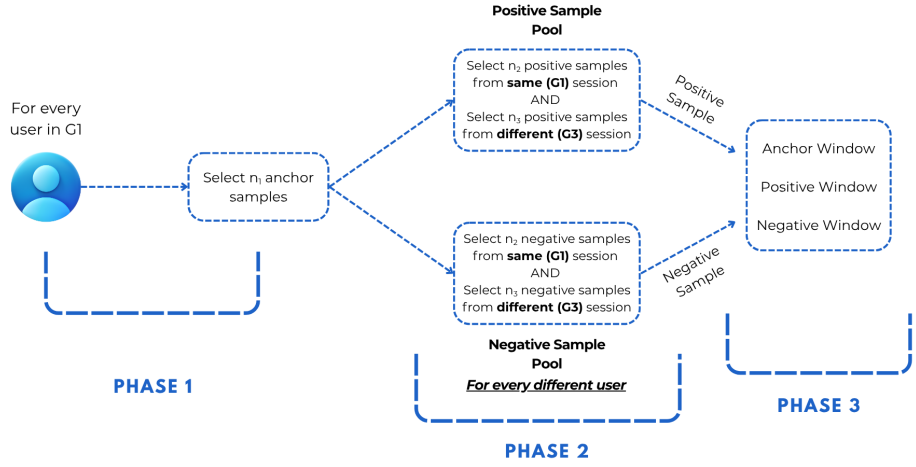



Figure 3.9: Overview of Session-Balanced Random Window Sampling (SBRWS). For each user in G1 (training) or G2 (validation), $n_1$ anchor windows are sampled. Each anchor is paired with $n_2$ positive samples from the same user and $n_3$ negative samples from different users, both sampled across sessions.

**Phase 1: Anchor Selection.** For each user $u$, a total of $n_1 = 5$ anchor windows are randomly sampled from session G1:

$$\mathcal{A}^{(u)} = \left\{ \mathbf{a}_i^{(u)} \mid \mathbf{a}_i^{(u)} \in \mathcal{S}_{G1}^{(u)},\ i = 1, \ldots, n_1 \right\}.$$

**Phase 2: Positive and Negative Sample Pools.** For each anchor window $\mathbf{a}_i^{(u)}$, a set of positive and negative windows is constructed:

- *Positive Pool (same user)*: $n_2 = 7$ positive windows are selected from $\mathcal{S}_{G1}^{(u)} \setminus \{\mathbf{a}_i^{(u)}\}$ and $\mathcal{S}_{G3}^{(u)}$ yielding total of 14 windows:

$$\mathcal{P}_{\text{same}}^{(u,i)} = \left\{ \mathbf{p}_j^{(u)} \in \mathcal{S}_{G1}^{(u)} \cup \mathcal{S}_{G3}^{(u)} \,\middle|\, j = 1, \ldots, 2n_2 \right\}$$

- *Negative Pool (different users)*: For each user $v \neq u$, $n_3 = 7$ windows are drawn from $\mathcal{S}_{G1}^{(v)}$ and $\mathcal{S}_{G3}^{(v)}$, yielding total of 14 windows:

$$\mathcal{N}^{(u,i)} = \bigcup_{v \neq u} \left\{ \mathbf{n}_j^{(v)} \in \mathcal{S}_{G1}^{(v)} \cup \mathcal{S}_{G3}^{(v)} \,\middle|\, j = 1, \ldots, 2n_3 \right\}.$$

**Phase 3: Triplet Composition.** Triplets are generated by pairing each anchor with all positive and negative windows in the respective pools:

$$\mathcal{T}^{(u,i)} = \left\{ (\mathbf{a}_i^{(u)}, \mathbf{p}_j^{(u)}, \mathbf{n}_k^{(v)}) \,\middle|\, \mathbf{p}_j^{(u)} \in \mathcal{P}_{\text{same}}^{(u,i)},\ \mathbf{n}_k^{(v)} \in \mathcal{N}^{(u,i)} \right\}.$$

Each anchor yields $14 \times 14 = 196$ triplets, resulting in $5 \times 196 = 980$ triplets per user, ensuring both high inter-session diversity and user-discriminative contrast. A similar approach is used during validation, with anchors selected from session G2 and positive/negative pools drawn from sessions G2 and G4.

**Sampling Summary**

- $n_1$ — Anchor windows per user: $5$
- $n_2 = n_3$ — Positive/negative windows per anchor (from same and different sessions): $7$
- Ratio of total positives to total negatives per anchor: $14/(50 \times 14) \approx 1:50$

The rationale behind SBRWS is to expose the model to both intra-user variation across sessions and inter-user dissimilarity. Notably, the negative sample pool is intentionally diversified by incorporating all other users, promoting stronger separation in the embedding space. However, this introduces a class imbalance, as the number of potential negative samples ($\sim 50 \times 2n_3$) significantly exceeds that of positives ($2n_2$), making the triplet composition inherently asymmetric. This imbalance, while increasing the diversity of impostor samples, may introduce challenges in convergence or overly penalize positive matches if not addressed during training. While this strategy is used throughout the main experiments, alternative sampling techniques are explored and evaluated in subsequent sections.

### 3.5 Session-Weighted Random Window Sampling

While the SBRWS strategy ensures an equal number of samples across user sessions, its uniform distribution may result in overly similar anchor-positive pairs, particularly

when windows are sampled from temporally adjacent regions. As illustrated in Figure 3.10, consecutive windows within a single session (e.g., G1) often exhibit minimal variation in signal dynamics. This leads to a positive pair pool that is overly homogeneous, reducing the model's ability to distinguish fine-grained temporal differences and potentially inducing overfitting.

To mitigate this effect, a Session-Weighted Random Window Sampling (SWRWS) strategy is introduced. In this formulation, anchor windows are again selected from session G1, but a greater emphasis is placed on incorporating session-diverse samples in the positive and negative pools. Specifically, for each anchor $a_i^{(u)}$, the positive sample pool is constructed using $n_2 = 5$ windows from session G1 (same session) and $n_3 = 10$ from session G3 (different session), sampled from the same user. This session-skewed configuration promotes temporal variability and enforces cross-session generalization, addressing the over-reliance on same-session coherence seen in the balanced case. The negative pool continues to consist of samples drawn from all other users across the same two sessions.

The resulting strategy not only balances user diversity but also adjusts for session-specific variance. The values of $n_1$, $n_2$, and $n_3$ in both approaches were selected such that the total number of generated triplets remains comparable between strategies, ensuring a fair basis for performance comparison. The total number of generated triplets is comparable to the balanced variant, with $56{,}348$ unique triplets formed under the SWRWS scheme. A formal summary of triplet construction parameters for both strategies is provided in Table 3.3, highlighting the shift in positive sample emphasis.

Table 3.3: Comparison of Triplet Sampling Strategies

| **Sampling Strategy** | $n_1$ **(Anchors)** | $n_2$ **(Same-Session)** | $n_3$ **(Diff-Session)** | **Total Triplets** |
|---|---|---|---|---|
| Balanced Sampling | 5 | 7 | 7 | 57,375 |
| Weighted Sampling | 5 | 5 | 10 | 56,348 |

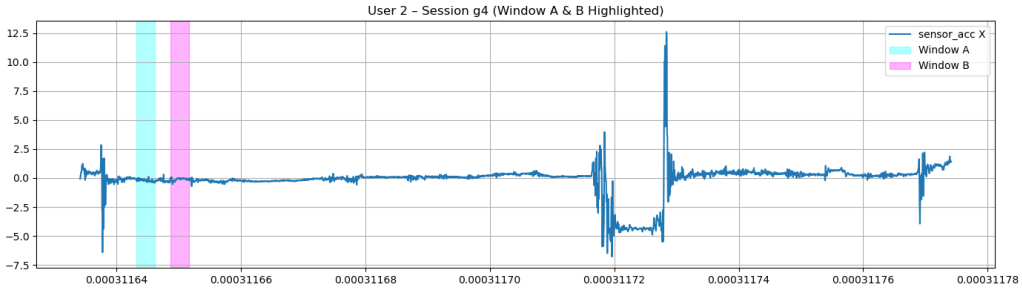



Figure 3.10: Example of anchor and positive window similarity within the same session. Adjacent windows (highlighted) exhibit highly overlapping patterns, leading to poor discrimination in learned representations.

## 3.6 Entropy-Based Window Filtering with Session-Weighted Sampling

Despite extensive alignment and preprocessing steps, windows with minimal signal variation—often corresponding to prolonged user inactivity—can persist in the dataset. Such low-variability windows lack discriminative information and, if retained, may degrade model performance by introducing noise into the training process. To address this, an entropy-based filtering strategy was applied to eliminate windows that exhibit insufficient amplitude diversity. The model trained using this filtered subset, in combination with session-weighted sampling, is hereafter referred to as *Entropy-Based Filtering with Session-Weighted Sampling (EBFSWS)*.

For each fixed-length window $w$, the discrete Shannon entropy $H(w)$ is computed as:

$$H(w) = -\sum_{i=1}^{B} p_i \log_2 p_i,$$

where the amplitude range is divided into $B = 30$ equally spaced bins, and $p_i$ denotes the empirical probability of a sample falling into bin $i$. A threshold $H_{\text{th}}$ is then defined as the 20th percentile of entropy values across all windows in the training set. Any window satisfying $H(w) < H_{\text{th}}$ is discarded. This criterion effectively removes both flat segments and low-amplitude fluctuations, even in cases where such windows overlap with legitimate touch events. The choice of entropy as a filtering criterion is supported by prior work on entropy-based filtering for time-series noise suppression

and feature preservation in sensor streams [76].

Figures 3.11 through 3.13 illustrate examples of low-entropy windows across multiple users, where the selected segments (highlighted in gray) span long durations with minimal sensor dynamics despite overlapping with touch activity (red dots). These low-variance regions typically occur during extended pauses between keystrokes, where the user's hand remains mostly stationary. Although these segments coincide with touch events, they are often uninformative for authentication due to the absence of motion patterns or due to potential sensor noise. It was speculated that the highlighted windows may reflect transient device movements or subtle tremors occurring between keystroke actions such as finger repositioning or unintended phone movement, yet they fail to exhibit sufficient dynamic content. In particular, the conditions under which keystroke events were recorded have not been disclosed in the dataset documentation, limiting the ability to attribute sensor patterns to specific user actions with full certainty. These observations collectively validate the necessity of the entropy-based filtering step to discard low-information windows that may otherwise hinder model training. The blue line in each figure denotes the raw acceleration signal on the X-axis, while the shaded gray regions mark the specific windows flagged as low-entropy. As seen in these examples, several gray regions are aligned with ongoing touch activity but exhibit very limited motion, underscoring the need for a content-aware selection strategy.

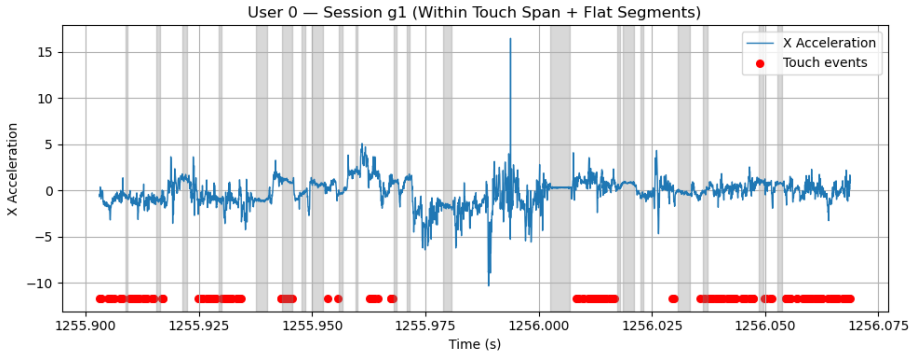



Figure 3.11: Low-entropy sensor segment for User 0—Session `g1`. Touch events are marked in red.

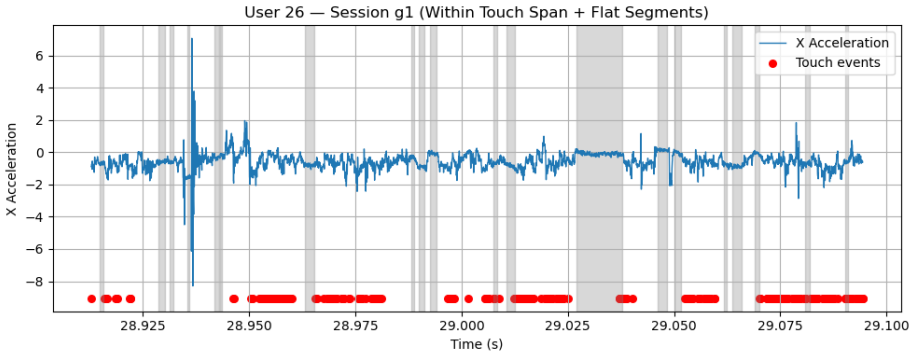



Figure 3.12: Low-entropy sensor segment for User 26—Session `g1`. Prolonged flat regions dominate the segment.

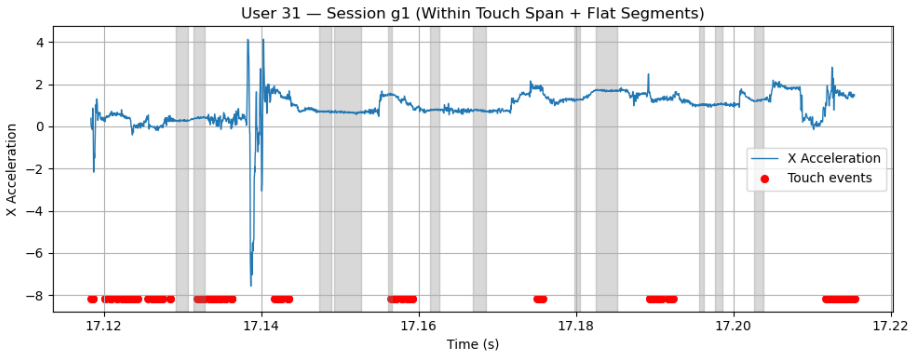



Figure 3.13: Low-entropy sensor segment for User 31—Session `g1`. Despite touch activity, informative variability is absent.

This filtering process substantially reduced the dataset size while preserving its most informative segments. Table 3.4 summarizes the impact of the entropy-based filter across preprocessing stages.

Table 3.4: Window counts at each preprocessing stage

| **Stage** | **# Windows** | **Reduction vs. Raw** | **% Remaining** |
|---|---|---|---|
| Raw sensor data | 10,254 | – | 100% |
| Touch-aligned to events | 3,076 | 70% lower | 30% |
| Entropy-filtered (final set) | 2,256 | 78% lower | 22% |

On the final filtered dataset, the session-weighted sampling strategy—originally intro-

duced in Section 3.5—was employed to construct triplet pairs. By restricting anchor-positive samples to high-entropy windows, the quality of training examples was further improved, mitigating the risk of overfitting to trivial patterns.

## 3.7 Index-Aligned Triplet Sampling

Although session-weighted sampling successfully introduces session-level diversity into the triplet construction process, it does not fully consider the temporal structure of user interactions. In particular, randomly sampled window pairs may correspond to semantically distant moments within a session, e.g., pairing the beginning of one session with the end of another. This issue is conceptually similar to the challenge of hard triplet mining, where semantically similar negatives must be selected to form more informative training samples [78]. To address this limitation, an Index-Aligned Triplet Sampling strategy was developed, windows of which are paired based on their temporal alignment across user sessions.

Let $w_{i,j}^{(u,s)}$ denote the $j$-th window from user $u$ in session $s$, where $i$ indexes the triplet (anchor, positive, or negative). Under the index-aligned scheme, for each anchor window $w_{a,j}^{(u,s_1)}$, the corresponding positive and negative windows are selected such that:

$$w_{p,j}^{(u,s_2)} \quad \text{and} \quad w_{n,j}^{(v,s_2)} \quad \text{with} \quad u \neq v \quad \text{and} \quad s_1 \neq s_2$$

That is, all three windows share the same temporal index $j$ but originate from different users and sessions. This enforces:

$$\text{Index constraint:} \quad \text{Index}(w_a) = \text{Index}(w_p) = \text{Index}(w_n)$$

Let $W_i^{(u,s)}$ denote the $i^{\text{th}}$ window of user $u$ in session $s$, where $i$ indicates the window index (starting from the beginning of the session), $u$ is the user ID, and $s$ refers to the session label (e.g., g1 to g4). For example, $W_1^{(1,\text{g1})}$ represents the first window of User 1 in session g1, and $W_7^{(1,\text{g3})}$ represents the seventh window of the same user in session g3.

To illustrate this idea, consider a toy scenario involving User 1 across sessions g1 and g3. Let $W_1^{(1,\text{g1})}$ denote the first window of User 1's session g1, and let $W_1^{(1,\text{g3})}$

denote the first window of the same user's session g3. In index-aligned sampling, these two windows are considered a valid positive pair since they occupy the same relative position. In contrast, $W_1^{(1,\text{g1})}$ would not be matched with $W_7^{(1,\text{g3})}$, which might correspond to an entirely different interaction phase such as prolonged idle holding, due to potential semantic misalignment.

In contrast, random window sampling selects anchor, positive, and negative windows without regard to their temporal position. Using the same example, $W_1^{(1,\text{g1})}$ could be arbitrarily paired with $W_7^{(1,\text{g3})}$ or even a flat segment from $W_{10}^{(1,\text{g2})}$. While this introduces diversity, it increases the risk of mismatching semantically unrelated segments, especially in sessions containing repetitive or idle behavior. Such mismatches may bias the model toward learning superficial or trivial distinctions.

Index-aligned sampling prevents such mismatches by enforcing structural consistency. In our toy example, $W_1^{(1,\text{g1})}$ is strictly compared with $W_1^{(1,\text{g3})}$, or with windows $W_1^{(u,s)}$ from other users or sessions that occupy the same relative index. This constraint helps stabilize the comparison space, particularly in scenarios with high intra-session behavioral variation.

This constraint imposes an implicit alignment of interaction phases across samples, promoting more semantically coherent and temporally synchronized comparisons. The motivation behind this sampling design draws inspiration from the positional encoding principles introduced in Transformer architectures [77], where temporal location within a sequence contributes to the expressiveness of learned representations. By analogously preserving temporal locality in triplet formation, the sampling process encourages the model to distinguish nuanced biometric patterns that are aligned in their relative interaction phase.

Although it is not strictly a hard triplet selection approach, the index-based strategy can be viewed as an extension to session-weighted sampling, motivated by similar principles. As noted earlier, session-weighted sampling emphasizes more confounding anchor-positive relationships by incorporating intra-session similarities. This idea naturally leads to enforcing temporal coherence across triplet elements, thereby controlling for trivial or overly dissimilar comparisons. The result is a more discriminative training signal and greater robustness against session-level or behavioral noise.

An illustration of the index-based window alignment logic is provided in Figure 3.14, where window positions are temporally matched across users and sessions.

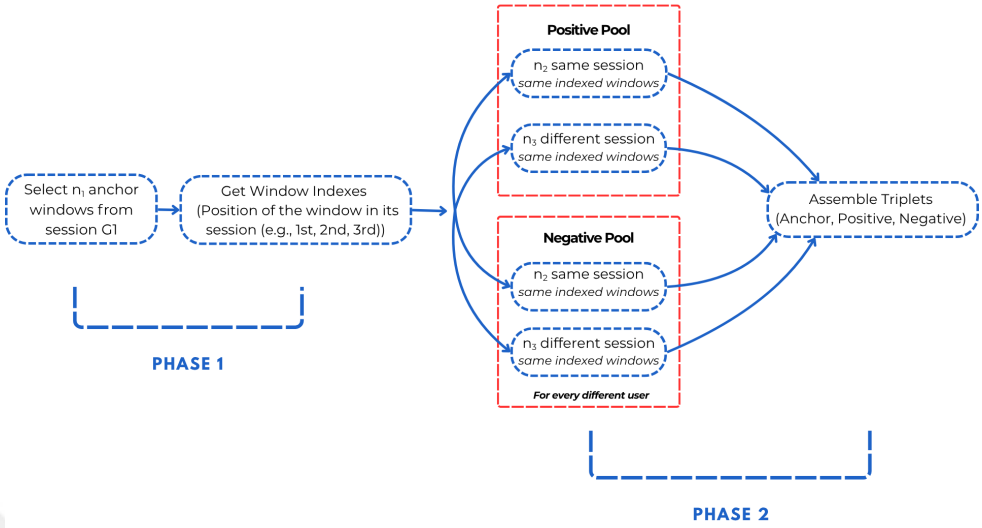



Figure 3.14: Illustration of index-aligned triplet sampling. Windows with matching temporal indices are selected across sessions and users to form semantically comparable anchor, positive, and negative samples. This alignment fosters meaningful intra-user and inter-user comparisons.

## 3.8 Relative Position-Based Triplet Sampling

As the final variant in the sampling framework, a more flexible strategy is proposed to overcome the limitations of strict index-based alignment. While exact positional matching ensures temporal synchrony across windows, it may be overly restrictive in real-world biometric data where session lengths vary considerably. To address this, a *Relative Position-Based Triplet Sampling* strategy is introduced, wherein triplet elements are selected based on their normalized temporal location within a session.

Let $w_j^{(u,s)}$ represent the $j$-th window of user $u$ in session $s$, and $L^{(u,s)}$ denote the total number of windows in that session. The relative position $\rho$ of window $w_j^{(u,s)}$ is defined as:

$$\rho_j^{(u,s)} = \frac{j}{L^{(u,s)}}$$

For each anchor window $w_a = w_j^{(u,\texttt{G1})}$, the corresponding positive and negative can-

didates are filtered based on their relative distance to the anchor:

$$\left| \rho_k^{(\cdot)} - \rho_j^{(u,\mathrm{G1})} \right| \leq \epsilon$$

where $\rho_k^{(\cdot)}$ is the relative position of a candidate window from another session (same or different user), and $\epsilon$ is a predefined tolerance margin. This flexible threshold permits semantically coherent but not strictly identical temporal pairing, enabling comparisons between interaction phases that are functionally similar but not identically positioned in absolute terms.

To clarify the relative position constraint, consider a toy example involving User 2 across sessions g1 and g3. Suppose that session g1 contains $L^{(2,\mathrm{g1})} = 10$ windows, and session g3 contains $L^{(2,\mathrm{g3})} = 20$ windows. Let the anchor window be $w_3^{(2,\mathrm{g1})}$, which has a relative position of $\rho_3^{(2,\mathrm{g1})} = \frac{3}{10} = 0.3$. A candidate window $w_6^{(2,\mathrm{g3})}$ from session g3 has a relative position of $\rho_6^{(2,\mathrm{g3})} = \frac{6}{20} = 0.3$, and thus qualifies as a positive match if $\epsilon \geq 0.0$. Similarly, a window such as $w_8^{(2,\mathrm{g3})}$ with $\rho = 0.4$ would also be accepted if the tolerance is relaxed to $\epsilon \geq 0.1$. In contrast, distant candidates like $w_{18}^{(2,\mathrm{g3})}$, with $\rho = 0.9$, would be rejected. This relative thresholding enables comparisons across sessions of differing lengths while maintaining temporal correspondence in behavioral dynamics.

The sampling process is illustrated in Figure 3.15. In Phase 1, anchor windows are selected and their relative positions are computed. In Phase 2, candidate windows from same-user (positive) and different-user (negative) sessions are filtered by enforcing the relative position constraint. Finally, in Phase 3, valid triplets are assembled based on this temporal correspondence.

Compared to absolute index matching, this formulation improves generalizability in sessions of varying length and captures higher-level behavioral alignment. For instance, two users may exhibit similar interaction dynamics early or late in a session, even if their sessions contain different numbers of windows. The proposed method adapts to this variability while still enforcing temporal coherence, thereby improving the discriminative quality of triplet construction and preventing trivial pairings due to session structure mismatch.

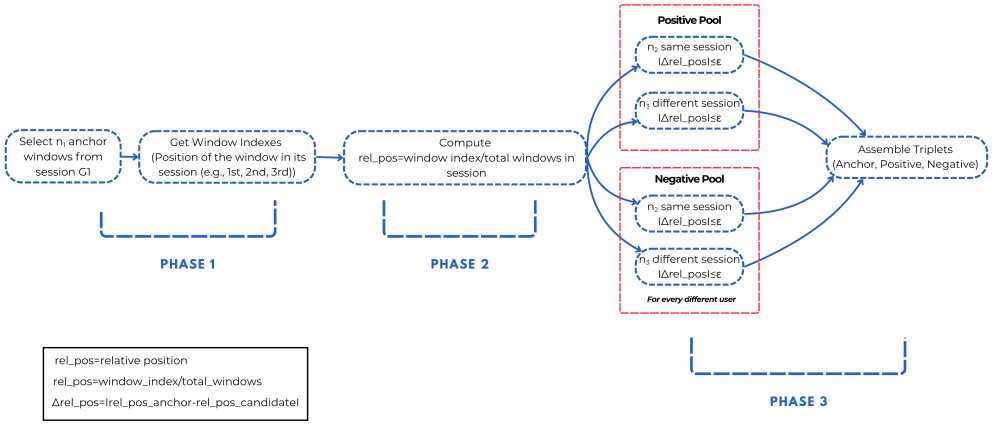



Figure 3.15: Relative Position-Based Sampling. Triplets are constructed by selecting anchor, positive, and negative windows with similar relative temporal positions ($\rho$) across sessions. This strategy allows flexible alignment even when session lengths vary.

## 3.9 Window-Ensembled Session Matching

While the training and validation strategies—including session-wise sampling and triplet construction—have been discussed in earlier sections, this chapter focuses on how predictions are made during inference. The evaluation set is designed to test the model's ability to determine whether two given sessions—composed of raw multi-modal sensor streams—originate from the same user. Each test sample is a session pair $(S_a, S_b)$, where the goal is to assign a binary decision indicating user similarity.

Instead of computing this decision from a single feature pair, the inference pipeline employs an ensemble-like approach over fixed-length windows extracted from each session. As illustrated in Figure 3.16, both sessions are segmented into $n$ windows, which are then paired based on a pre-defined logic. Specifically, the figure corresponds to the index-based alignment configuration discussed in Section 3.7, where window $i$ from session $S_a$ is paired with window $i$ from session $S_b$. In other configurations—such as relative position-based pairing—the window pairs are selected

using their normalized temporal indices, as defined by their fractional position within a session. Each configuration reuses the same core inference mechanism but differs in how windows are aligned.

For each aligned pair $(w_i^{(a)}, w_i^{(b)})$, their corresponding embeddings $\mathbf{e}_i^{(a)}$ and $\mathbf{e}_i^{(b)}$ are passed through the trained Siamese model to compute a distance score $d_i$ in the latent space:

$$d_i = \|\mathbf{e}_i^{(a)} - \mathbf{e}_i^{(b)}\|_2$$

The final score for the entire session pair is then computed by aggregating these distances, typically by taking the mean:

$$D_{\text{final}} = \frac{1}{n} \sum_{i=1}^{n} d_i$$

Since lower distances imply higher similarity, this raw score is transformed into a normalized confidence score using a threshold $\tau$, determined empirically on the validation set. The decision function is then formulated as:

$$\hat{y} = \begin{cases} 1 & \text{if } D_{\text{final}} < \tau \\ 0 & \text{otherwise} \end{cases}$$

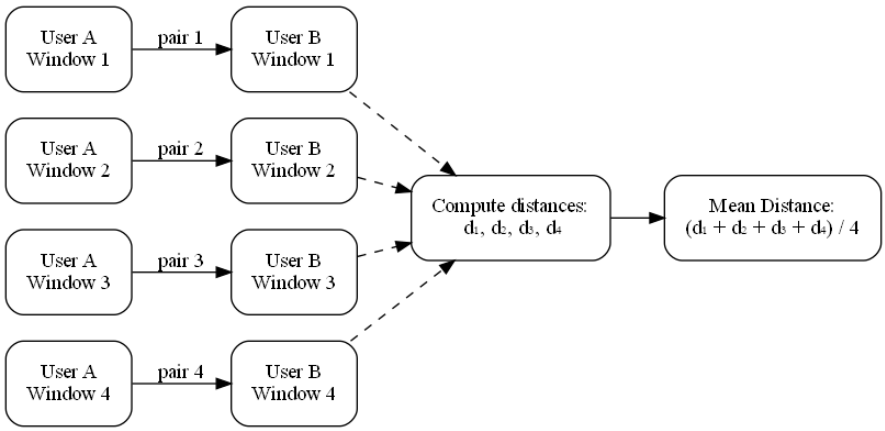



Figure 3.16: Inference mechanism via window-aligned ensembling. For a given session pair, windows are aligned (e.g., by index), and pairwise embedding distances are computed. Final similarity is based on the mean distance across all aligned pairs.

This approach, termed *window-ensembled post-inference*, enables robust verification by capturing the temporal diversity within sessions and aggregating evidence across

the full signal rather than relying on a singular snapshot. The ensemble mechanism ensures that noisy or non-informative windows do not disproportionately affect the final decision. Furthermore, different sampling logics (e.g., index-based, relative position-based) allow for flexible pairing strategies that reflect varied temporal characteristics of user behavior. This strategy is analogous to majority voting or confidence fusion in traditional ensemble learning, but operates in the embedding distance domain, offering resilience against intra-session variability.

### 3.10 Baseline Model with Statistical Feature Aggregation

A baseline model for comparison was constructed using statistical features extracted from the sensor accelerometer modality. After applying the preprocessing procedures described in the previous section, raw tri-axial signals ($x$, $y$, $z$) were transformed into descriptive feature vectors. For each axis, a set of 15 statistical descriptors was computed, including mean, median, standard deviation, minimum, maximum, range, interquartile range (IQR), quartiles (Q1, Q3), percentiles (10th, 25th, 75th, 90th), median absolute deviation (MAD), zero-crossing rate, and signal energy. These values capture both the central tendency and variation in the signals, as well as temporal dynamics. The full list of features used is presented in Table 3.5. Concatenating the feature sets from all three axes resulted in a 45-dimensional representation per session. During training, a pairwise comparison setup was used: for each training sample, two user sessions were randomly selected, and their 45-dimensional feature vectors were concatenated, yielding a 90-dimensional input to the model. The goal of the model was to determine whether the two input feature sets belong to the same individual. XGBoost, a gradient boosting decision tree algorithm known for its scalability and performance [79], was employed as the classification model. This strategy mirrors approaches taken in earlier studies, such as Touchalytics [12], which used handcrafted statistical features for biometric analysis on touch modalities. While several works have demonstrated the efficacy of statistical summarization techniques for sensor signals, such as keystroke-based user identification using inter-key latency features [80] or motion-derived descriptors in smartphone authentication [29], this study extends such strategies to a consistent accelerometer-based formulation. Training was

performed using sessions g1, g2, and g3, while validation was conducted on session g4 to simulate deployment-time variability across time and session contexts. To optimize the baseline model's performance, hyperparameter tuning was conducted using a Bayesian search strategy. The optimal parameter configuration obtained from this process is detailed in as shown in Appendices Table B.1.

Table 3.5: List of extracted statistical features per sensor axis. Each feature is computed for $x$, $y$, and $z$ components, resulting in 75 total features per session.

| **Feature Name** | **Description** |
|---|---|
| Mean | Average of the signal values |
| Median | Central value of the signal |
| Standard Deviation | Dispersion of signal values |
| Minimum | Smallest value in the signal |
| Maximum | Largest value in the signal |
| Range | Difference between max and min |
| Interquartile Range (IQR) | Q3 minus Q1 |
| 1st Quartile (Q1) | 25th percentile |
| 3rd Quartile (Q3) | 75th percentile |
| 10th Percentile | Value below which 10% fall |
| 25th Percentile | Value below which 25% fall |
| 75th Percentile | Value below which 75% fall |
| 90th Percentile | Value below which 90% fall |
| Median Absolute Deviation (MAD) | Median of absolute deviations from the median |
| Zero-Crossing Rate | Frequency of signal crossing zero |
| Signal Energy | Sum of squared values |

# CHAPTER 4

# EXPERIMENT RESULTS

This section presents the experimental evaluation of the proposed sampling strategies and model configurations introduced in Chapter 3. A series of controlled comparisons are conducted to assess the impact of session-aware sampling, entropy-based window filtering, and temporal alignment techniques on authentication performance. The experiments are designed to highlight how each component contributes to model generalizability, robustness, and sensitivity to intra- and inter-session variability. Metrics such as EER, ROC AUC and user-specific confusion matrices are reported to provide a complete understanding of the strengths and limitations of each strategy.

## 4.1 Experiment Setup

Prior to presenting the evaluation metrics and comparative analyses across triplet sampling strategies, a comprehensive hyperparameter optimization procedure was conducted for the dual-branch Siamese architecture combining LSTM and MLP modules. For each sampling strategy including Session-Balanced, Session-Weighted, Index-Based, and Relative Position-Based triplet construction, a separate hyperparameter tuning phase was executed to ensure fair evaluation. The search was performed via Bayesian optimization using the Weights & Biases (WANDB) platform, focusing on maximizing validation F1-SCORE.

The optimization was applied independently to each model branch: the LSTM pathway processing time-series features, and the MLP pathway handling FFT vectors. A search space was defined for key architectural and training parameters, such as LSTM depth, hidden dimensions, bidirectionality, MLP layer sizes, and training-specific pa-

rameters like learning rate and batch size. The best-performing configurations identified through this process were retained for final training and evaluation. For each strategy same configuration gives the best F1-Score.

Table 4.1: Best Hyperparameters for Siamese LSTM + MLP Model (Selected via WANDB Bayesian Optimization)

| **Component** | **Selected Configuration** |
|---|---|
| LSTM Layers | 3 |
| LSTM Hidden Size | 64 |
| MLP Layers (FFT Branch) | [1024, 512, 256, 128] |
| Fusion MLP Layers | [256, 128, 64] |
| Dropout (applied to all hidden layers) | 0.3 |
| Learning Rate | $1 \times 10^{-3}$ |
| Batch Size | 64 |
| Margin ($\alpha$) for Triplet Loss | 0.2 |
| Optimizer | Adam |
| Training Epochs | 50 |

## 4.2 Results

The following subsections provide a comparative analysis of various model and sampling configurations. Section 4.2.1 begins with an evaluation of the baseline Siamese model versus the session-balanced variant to assess the effect of balanced anchor-positive pair selection. Section 4.2.2 extends this analysis by comparing session-balanced and session-weighted sampling strategies, highlighting how the weighting mechanism affects representation learning. Section 4.2.3 focuses on the contribution of entropy-based filtering in eliminating uninformative segments, while Section 4.2.4 integrates all proposed enhancements—including relative position-based sampling—to provide a comprehensive evaluation of temporal semantics in triplet construction.

### 4.2.1 Comparison of Baseline and Comparison of Session-Balanced Random Window Sampling (SBRWS)

This subsection presents a comparative evaluation between the traditional XGBoost-based baseline model and the deep Siamese network trained using the Session-Balanced Random Window Sampling (SBRWS) strategy. The Baseline model, which relies on hand-crafted statistical features inspired by prior literature, exhibits notable overfitting behavior. While its training AUC reaches 0.853, performance deteriorates significantly on the validation and test sets, yielding AUC scores of 0.647 and 0.598, respectively. Similarly, the F1-score and Recall metrics follow this downward trajectory, indicating that the model fails to generalize effectively. A detailed performance comparison between the baseline and Session-Balanced Siamese models is provided in Table 4.2.

In contrast, the Siamese model trained under the SBRWS configuration achieves substantially higher recall scores across all sets, with nearly perfect recall (0.997) on the training data. This suggests that the model tends to classify most samples as belonging to the same user, thus capturing nearly all true positives. However, this behavior results in lower precision and AUC on the test set, signaling overfitting driven by a small and homogeneous positive sample pool. As previously discussed in Section 3.5, the high degree of similarity among within-session positive windows biases the model toward positive predictions. Consequently, the Siamese model's recall-oriented behavior leads to higher Equal Error Rates (EER) on unseen data, though still outperforming the baseline.

Overall, while the Siamese model demonstrates improved validation metrics, the discrepancy between training and testing results highlights the limitations of the Session-Balanced strategy and motivates further refinements addressed in subsequent sections. The pronounced overfitting observed in the recall-dominant behavior suggests that the model's understanding of user-specific patterns is limited when exposed only to intra-session window similarities. This restricts its ability to generalize effectively to more varied or temporally distant user behaviors. In practice, such performance inconsistency poses challenges for real-world biometric systems, which must operate reliably across diverse interaction contexts and usage patterns. These findings emphasize the

necessity of incorporating richer session diversity and more robust sampling logic, which are systematically explored in the following sampling strategies.

Table 4.2: Performance Comparison: Baseline (XGBoost) vs. Siamese (Session-Balanced)

| Metric | Baseline (XGBoost) | | | Siamese (Session-Balanced) | | |
|---|---|---|---|---|---|---|
| | Train | Validation | Test | Train | Validation | Test |
| AUC | 0.853 | 0.647 | 0.598 | 0.778 | 0.726 | 0.643 |
| F1-score | 0.773 | 0.589 | 0.562 | 0.667 | 0.694 | 0.552 |
| Precision | 0.779 | 0.592 | 0.571 | 0.502 | 0.687 | 0.491 |
| Recall | 0.768 | 0.578 | 0.552 | 0.997 | 0.703 | 0.612 |
| EER | 0.401 | 0.397 | 0.429 | 0.305 | 0.309 | 0.335 |

### 4.2.2 Comparison of Session-Balanced Random Window Sampling (SBRWS) and Session-Weighted Random Window Sampling (SWRWS)

This subsection evaluates the effectiveness of the Session-Weighted Random Window Sampling (SWRWS) strategy in comparison to the Session-Balanced Random Window Sampling (SBRWS) approach described earlier. As shown in Table 4.3, the SWRWS configuration outperforms SBRWS across nearly all performance metrics, especially on the test set. This improvement can be attributed to the core logic of SWRWS, where more positive samples are drawn from different sessions (i.e., G3) rather than the same session (i.e., G1). This sampling strategy exposes the model to higher intra-user variability and increases the semantic diversity of positive pairs during training.

In particular, the recall score of the SBRWS model on the training set remains nearly perfect (0.997), indicating that the model tends to predict most samples as positive due to the excessive similarity of intra-session windows. This results in lower test precision and F1-score, reflecting an overfitting tendency towards recall. On the other hand, SWRWS significantly reduces this overfitting effect by incorporating more challenging anchor-positive pairs that better represent cross-session user behavior. As a result, the recall becomes more reasonable (0.860 on training, 0.595 on test),

and the model achieves higher test AUC (0.688 vs. 0.643) and F1-score (0.580 vs. 0.552), indicating improved generalization and class balance.

This comparison highlights the importance of carefully balancing session diversity during triplet sampling. By prioritizing session-wise dissimilarity, SWRWS ensures that the model learns more discriminative representations that generalize better to unseen data.

Table 4.3: Comparison of Siamese Network Variants: Session-Balanced vs. Session-Weighted Sampling

| **Metric** | **Siamese (SBRWS)** | | | **Siamese (SWRWS)** | | |
|---|---|---|---|---|---|---|
| | **Train** | **Validation** | **Test** | **Train** | **Validation** | **Test** |
| AUC | 0.778 | 0.726 | 0.643 | 0.886 | 0.739 | 0.688 |
| F1-score | 0.667 | 0.694 | 0.552 | 0.784 | 0.704 | 0.580 |
| Precision | 0.502 | 0.687 | 0.491 | 0.687 | 0.632 | 0.565 |
| Recall | 0.997 | 0.703 | 0.612 | 0.860 | 0.775 | 0.595 |
| EER | 0.348 | 0.362 | 0.374 | 0.318 | 0.324 | 0.349 |

To better understand the behavioral implications of sampling strategies, a session-level analysis of feature similarity was conducted. The primary motivation behind this analysis was to evaluate how consistent a user's behavior is within a session compared to across different sessions. If windows from the same session are highly similar in terms of extracted features, while windows from different sessions are more diverse—even for the same user—this would support that **cross-session pairing is more informative** for model generalization. In an ideal scenario, intra-session windows are expected to show low feature distance due to consistent environmental and behavioral conditions, whereas inter-session windows may reflect natural variations in user posture, context, or interaction style.

To explore this, Euclidean distances were computed between window-level feature vectors (not model embeddings) extracted from the sensor data. Specifically, for each user, distances were calculated between randomly selected pairs of windows from within the same session (G1–G1) and across sessions (G1–G3), based on the hand-

crafted statistical features described in Section 3.3. As reported in Table 4.4, most users exhibited substantially lower intra-session distances compared to their inter-session distances. This pattern confirms that user behavior tends to be temporally consistent within a session but more variable across sessions, reinforcing the need to diversify training pairs using cross-session windows in order to avoid overfitting to trivially similar patterns.

However, certain users (e.g., User 7, User 19, User 25) displayed extremely low intra-session distances (e.g., 0.013–0.022), while maintaining high inter-session distances (e.g., above 1.2). These cases suggest an overly homogeneous G1 session, which, under session-balanced sampling, can lead to anchor-positive pairs that are excessively similar. As a result, the model may overfit to these near-duplicate patterns, learning to assign high similarity scores regardless of broader behavioral variability.

To complement the feature-level analysis presented earlier, a second evaluation was conducted at the embedding level to examine how the two sampling strategies influence the learned representation space. Table 4.5 reports the average distances between anchor-positive and anchor-negative pairs in the embedding space, computed after model training. This analysis was performed using the same users examined in Table 4.4, allowing a direct comparison of how sampling strategies affect similarity learning for individuals with highly homogeneous session data.

For example, in User 34, the session-balanced model produced an average distance of 0.069 between anchor and positive samples, while the distance to negative samples (i.e., from different users) was 0.103—indicating only a small separation between same-user and different-user embeddings. In contrast, the session-weighted model increased the anchor-positive distance to 0.352 and the anchor-negative distance to 0.491, thereby enlarging the margin between classes. This improvement suggests that incorporating cross-session variation during training enables the model to better generalize and avoids overfitting to redundant intra-session patterns.

These results demonstrate that increasing the representation of cross-session windows in the positive pool improves embedding separation. By exposing the model to more diverse intra-user patterns, session-weighted sampling mitigates the tendency to overfit to highly similar positive pairs and produces representations that better reflect the

underlying variability in user behavior.

Table 4.4: Mean and standard deviation of Euclidean distances between intra-session (G1–G1) and inter-session (G1–G3) feature vectors for each user.

| **User ID** | **Intra-Session Dist. (Mean ± SD)** | **Inter-Session Dist. (Mean ± SD)** |
|---|---|---|
| 7 | 0.013 ± 0.005 | 1.201 ± 0.126 |
| 19 | 0.018 ± 0.007 | 1.118 ± 0.108 |
| 25 | 0.022 ± 0.006 | 1.250 ± 0.137 |
| 34 | 0.016 ± 0.004 | 1.301 ± 0.120 |
| 48 | 0.015 ± 0.005 | 1.289 ± 0.115 |

Table 4.5: Mean and standard deviation of embedding distances between anchor-positive and anchor-negative pairs for selected users under two sampling strategies: Session-Balanced Random Window Sampling (SBRWS) and Session-Weighted Random Window Sampling (SWRWS). Anchor-positive pairs consist of two windows from the same user, while anchor-negative pairs include one window from the anchor user and one from a different user.

| **User ID** | **Siamese (SBRWS)** | | **Siamese (SWRWS)** | |
|---|---|---|---|---|
| | **Pos. Dist. (Mean ± SD)** | **Neg. Dist. (Mean ± SD)** | **Pos. Dist. (Mean ± SD)** | **Neg. Dist. (Mean ± SD)** |
| 7 | 0.093 ± 0.015 | 0.115 ± 0.022 | 0.232 ± 0.034 | 0.328 ± 0.048 |
| 19 | 0.088 ± 0.013 | 0.117 ± 0.018 | 0.273 ± 0.039 | 0.392 ± 0.052 |
| 25 | 0.067 ± 0.010 | 0.181 ± 0.025 | 0.364 ± 0.041 | 0.460 ± 0.045 |
| 34 | 0.069 ± 0.012 | 0.103 ± 0.019 | 0.352 ± 0.037 | 0.491 ± 0.051 |
| 48 | 0.081 ± 0.014 | 0.105 ± 0.020 | 0.358 ± 0.033 | 0.470 ± 0.047 |

In conclusion, these findings support the effectiveness of the session-weighted random window sampling approach in addressing overfitting and improving generalization. By strategically increasing session diversity in anchor-positive pairing, the model is encouraged to learn more robust and semantically meaningful representations of user-specific behaviors.

### 4.2.3 Comparison of Session-Weighted Sampling (SWRWS) and Entropy-Based Filtering with Session-Weighted Sampling (EBFSWS)

To assess the contribution of entropy-based filtering in the context of sampling strategies, the performance of two variants of the Siamese model—trained with identical Session-Weighted Sampling logic—was compared. The first model utilizes all touch-aligned windows, while the second operates on a subset of entropy-filtered windows that were selected based on the method described in Section 3.6. As shown in Table 4.6, although the filtered version was trained on significantly fewer windows (22% of the original set), it consistently demonstrated improved generalization across several metrics.

The filtered model achieved a higher test AUC (0.693 vs. 0.688), and precision increased from 0.565 to 0.580, reflecting more reliable discrimination between true and false matches. Additionally, the F1-score improved slightly (from 0.580 to 0.590), indicating a more favorable balance between precision and recall. While the recall of the filtered model (0.591) is marginally lower than that of the unfiltered version (0.595), this small trade-off suggests that overly permissive positive classification was tempered, yielding better-calibrated outputs. The EER was also reduced from 0.349 to 0.327, further supporting the robustness of the entropy-filtered pipeline.

These findings support the notion that quality, rather than quantity, of input windows can drive better verification performance. By filtering out non-informative segments and focusing model capacity on semantically rich data, entropy-based preprocessing not only reduces computational burden but also improves generalization in session-level matching tasks. This strategy ensures that the model attends primarily to the portions of the interaction that exhibit meaningful behavioral signals, thereby increasing the likelihood of capturing user-specific traits. Furthermore, by reducing exposure to redundant or idle gestures, the model becomes less prone to overfitting on non-discriminative patterns. As a result, the training process becomes more efficient, both in terms of computation and convergence stability. The filtering mechanism thus acts as a form of regularization, constraining the learning process to more informative subsequences, which ultimately enhances robustness in real-world deployment scenarios where noise and variability are common.

Table 4.6: Comparison of Session Weighted Sampling: With vs. Without Entropy Filtering

| **Metric** | **Session-Weighted** | | | **Filtered** | | |
|---|---|---|---|---|---|---|
| | **Train** | **Val** | **Test** | **Train** | **Val** | **Test** |
| AUC | 0.886 | 0.739 | 0.688 | 0.862 | 0.748 | 0.693 |
| F1-score | 0.784 | 0.704 | 0.580 | 0.772 | 0.716 | 0.593 |
| Precision | 0.687 | 0.632 | 0.565 | 0.701 | 0.658 | 0.580 |
| Recall | 0.860 | 0.775 | 0.595 | 0.850 | 0.783 | 0.591 |
| EER | 0.318 | 0.324 | 0.349 | 0.309 | 0.313 | 0.327 |

### 4.2.4 Comprehensive Comparison of Sampling Strategies Integrating Temporal Semantics

In this final subsection, a comparative evaluation is conducted among the three most effective triplet construction strategies developed in this study: *EBFSWS*, *Index-Aligned Triplet Sampling*, and *Relative Position-Based Triplet Sampling*. These configurations have been selected based on their empirical performance and conceptual robustness, each reflecting a different facet of temporal structure-aware window pairing.

The quantitative results are presented in Table 4.7. All three strategies exhibit high performance in terms of AUC and low EER. However, notable differences are observed in the precision-recall balance, which reflects the ability of the model to mitigate bias toward one class in the binary classification setup.

The entropy-filtered configuration yielded the best F1-score on the test set ($0.593$), along with superior precision ($0.580$), suggesting that eliminating uninformative or noisy segments significantly improves discriminability without sacrificing generalization. This is achieved with substantially fewer samples, as noted in Table 3.4. Index-Aligned Triplet Sampling follows closely, achieving the highest test AUC ($0.702$) and a well-balanced recall ($0.617$) and precision ($0.581$) profile. This consistency can be attributed to its temporal alignment mechanism, which ensures that paired windows

represent similar temporal contexts, thereby improving inter-session discrimination.

Relative Position-Based Sampling, while comparable to index-aligned sampling, demonstrates slightly more stable generalization across all three splits, as evidenced by marginally higher validation performance in AUC ($0.753$), precision ($0.583$), and recall ($0.619$). This indicates that relative temporal matching may offer slightly improved robustness to intra-session variability compared to absolute index matching.

Overall, all three strategies substantially outperform earlier baseline model, and the results affirm the hypothesis that incorporating temporal semantics—whether through entropy, index alignment, or relative position—plays a important role in improving biometric verification accuracy from sensor data.

Table 4.7: Comparison of Sampling Strategies Integrating Temporal Semantics

| **Metric** | **Filtered** | | | **Index-Aligned** | | | **Relative Position-Based** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Train** | **Val** | **Test** | **Train** | **Val** | **Test** | **Train** | **Val** | **Test** |
| AUC | 0.862 | 0.748 | 0.693 | 0.901 | 0.752 | 0.702 | 0.902 | 0.753 | 0.703 |
| F1-score | 0.772 | 0.716 | 0.593 | 0.795 | 0.729 | 0.599 | 0.797 | 0.731 | 0.601 |
| Precision | 0.701 | 0.658 | 0.580 | 0.712 | 0.667 | 0.581 | 0.714 | 0.669 | 0.583 |
| Recall | 0.850 | 0.783 | 0.591 | 0.878 | 0.791 | 0.617 | 0.879 | 0.793 | 0.619 |
| EER | 0.309 | 0.313 | 0.327 | 0.312 | 0.319 | 0.344 | 0.311 | 0.317 | 0.343 |

# CHAPTER 5

# DISCUSSIONS & LIMITATIONS

This chapter discusses the study's findings in relation to the two central research questions posed in Chapter 1. By analyzing the outcomes of the proposed sampling and filtering strategies, and their impact on model performance and generalization, this section aims to contextualize the results within the broader scope of continuous behavioral authentication. Each research question is revisited to assess how effectively the proposed methods address the associated challenges, while highlighting key insights and methodological implications. The discussion is followed by an examination of the study's limitations to clarify its boundaries and suggest directions for future research.

## 5.1 Discussions on the Research Questions

**How can behavioral biometric data be selectively sampled to maximize its discriminative value while minimizing redundancy?**

This study explored the challenge of extracting meaningful behavioral information from high-dimensional and often noisy mobile sensor streams by devising strategies to selectively sample and retain the most informative data segments. Given the continuous and uncontrolled nature of user interaction data, not all segments contribute equally to identity discrimination. In this context, fixed-length windowing served as the foundation to impose temporal consistency, yet it also introduced a risk of incorporating uninformative or redundant signal parts. To address this, an entropy-based filtering approach was employed, wherein each window's discrete Shannon entropy was computed and those falling below a task-specific threshold were discarded. This

step effectively eliminated low-amplitude, repetitive patterns, which are often found in prolonged idle periods, thereby filtering out structurally uninformative content.

After filtering, only 22% of the original data was retained without degrading model performance—and in some cases, it even improved. This suggests that much of the raw signal data lacked identity-relevant information. Using signal entropy to guide data reduction proved effective, confirming that intrinsic signal features can support selective sampling for compact and discriminative input representations.

**How can comparison pairs in biometric authentication models be constructed to meaningfully capture behavioral similarities and differences within the same user and between distinct user profiles?**

The formation of comparison pairs lies at the core of learning effective distance metrics in behavioral biometric systems, especially within Siamese architectures that rely on relative similarity between input pairs or triplets. This thesis introduced a sequence of novel sampling strategies to improve the informativeness of comparison pairs, each rooted in different assumptions about behavioral consistency across time and sessions. The initial approach, Session-Balanced Random Window Sampling (SBRWS), evenly sampled anchor, positive, and negative windows from across sessions, aiming to expose the model to session-wise variation. However, this strategy often yielded overly similar anchor-positive pairs, particularly when both were drawn from the same session, resulting in an inflated recall and reduced generalization due to the model's bias toward outputting positive predictions.

To mitigate this issue, a Session-Weighted sampling strategy was implemented, emphasizing positive samples from different sessions. This enforced a greater diversity of intra-user comparisons, which encouraged the model to focus on more robust and generalizable behavioral traits. Beyond this, two additional strategies—Index-Aligned Sampling and Relative Position-Based Sampling—were developed to use temporal structure more explicitly. The former aligned windows based on their ordinal session positions, while the latter matched windows using their relative progression within a session. These designs introduced temporal coherence into the comparison logic, producing anchor-positive pairs with similar contextual placement but varied sensor content, and more challenging anchor-negative pairs due to improved

alignment across users.

This alignment proved especially beneficial in improving model generalization and embedding space separability, as shown by both quantitative metrics and intra/inter-user distance analyses. Altogether, these sampling strategies demonstrate that embedding temporal logic into comparison pair construction can significantly improve a model's ability to learn meaningful distinctions between users, thereby providing a strong and comprehensive answer to the second research question.

## 5.2 Limitations

While the proposed methodology demonstrates strong performance across multiple sampling strategies and model configurations, several limitations should be acknowledged. First, the analysis is constrained to a specific subset of behavioral modalities—primarily keystroke-related accelerometer and touch data—within a controlled dataset. Although this setting enables focused experimentation, it may not fully generalize to more diverse behavioral signals or unconstrained mobile environments where user interaction patterns exhibit greater variability.

In addition, the dataset used in this thesis was collected in a controlled setting with a limited number of users and sessions. Although the design attempted to simulate real-world usage by varying keyboard layouts and interaction contexts, the behavioral variability observed in uncontrolled environments may be broader and harder to model. The dataset also lacks metadata on user input behaviors such as corrections, deletions, or hesitation patterns—factors that could significantly enrich feature representations in keystroke-based biometrics. For example, pauses, backspaces, or erratic typing may reveal important individual traits that were not captured due to the structure of the data. Future work should consider collecting richer interaction logs that include such behavioral signals.

Another limitation stems from the nature of the input tasks. Users were asked to type fixed content including their name, surname, and answers to specific questions. While this improves label consistency, it may also bias the data collection process. Typing one's own name or familiar words may lead to highly practiced and auto-

mated movements, reducing the variability that would be present in more natural text entry scenarios. In contrast, real-world applications such as messaging or form filling would involve more diverse and cognitively demanding input, which could affect the model's generalizability.

Second, the fixed-length windowing and sampling logic, while effective, rely on predefined parameters such as window size, entropy thresholds, and positional matching tolerances. These hyperparameters, although empirically validated, may not be optimal for all users or interaction types. The approach also presumes temporal consistency across sessions, particularly in the index-aligned and relative position-based strategies, which may not hold in real-world usage where behavioral drift is more prominent.

Third, while entropy-based filtering provides a principled mechanism to eliminate low-information windows, it may inadvertently discard subtle but discriminative patterns if they occur within low-variance regions. Similarly, pairing strategies, although carefully designed to encourage difficult and informative comparisons, still rely on heuristic definitions of session diversity and position alignment, which may not universally capture the most relevant distinctions between users.

Lastly, the evaluation in this thesis is based on a single test protocol employing window-ensembled inference for session-to-session verification. Alternative verification paradigms—such as session-to-profile matching, where a new session is compared against an aggregated representation of a user's past behavioral history—are not explored. Such approaches more closely resemble real-world biometric systems, enabling faster inference and greater robustness by leveraging long-term user-specific templates. Furthermore, incremental user modeling, in which user profiles are continuously updated with new data over time, offers potential for handling behavioral drift and improving adaptability in longitudinal deployments. Incorporating these dynamic and scalable evaluation strategies could substantially enhance the practical applicability of the proposed framework and remains a compelling direction for future research.

# CHAPTER 6

# CONCLUSION AND FUTURE WORK

This thesis proposed a comprehensive framework for sensor-based biometric authentication, emphasizing the discriminative power of time-aligned and frequency-aware features extracted from raw mobile sensor data. Beginning with a robust preprocessing pipeline, the study filtered and segmented noisy sensor streams—most notably from the accelerometer modality—into fixed-length windows, synchronizing them with touch-based user interaction to ensure contextual informativeness.

A Siamese network model, integrating an LSTM-based encoder for temporal sequences and an MLP for FFT features, was trained using multiple triplet sampling strategies designed to expose the model to both intra-user and inter-user variations. Among these strategies, Session-Balanced Random Window Sampling (SBRWS) was shown to suffer from overfitting due to the high similarity of windows within the same session, leading to inflated recall rates. Session-Weighted sampling improved the balance between recall and precision by increasing the diversity of positive pairs. Further filtering via entropy-based criteria enabled removal of non-informative idle windows, reducing the data size while improving generalizability. Index-based and relative position-based sampling schemes introduced temporal consistency into anchor-positive-negative selection, aligning semantically similar segments across sessions and thereby enhancing training quality.

The final evaluation was conducted through a post-inference mechanism using window-ensembled matching. This strategy computed average distances between aligned window embeddings, providing robustness against noise and intra-session variability. Across experiments, models trained with position-aware and entropy-filtered triplets outperformed baselines in balanced metrics, supporting the thesis' core hypothesis:

that informed sampling and representation strategies can enhance biometric discriminability in real-world, session-variant sensor data.

From a deployment perspective, several challenges arise when translating these results into operational systems. Real-time sensor-based biometric authentication requires continuous monitoring and fast inference, which may not be feasible on all mobile devices due to hardware and energy constraints. Privacy concerns also arise, especially when raw sensor data must be stored or transmitted for analysis. Moreover, behavioral biometrics may exhibit high intra-user variability over time, requiring models that can adapt incrementally as user behavior evolves. Another critical consideration is determining the optimal point during the session to make a reliable authentication decision. Making a decision too early may lead to conclusions drawn from insufficient data, whereas deferring the decision until the session's end can delay response time and affect user experience. Balancing accuracy and responsiveness in such temporal decision-making remains a key deployment challenge for continuous authentication systems.

Despite these limitations, the results of this thesis have important implications for real-world applications. The proposed methodology is particularly well-suited for passive authentication systems, where the user's identity is verified in the background based on ongoing interactions. Such systems can enhance security in mobile banking, personalized healthcare apps, or secure workplace environments without requiring explicit user input. Additionally, embedding position-aware triplet selection strategies in training pipelines can help improve robustness in scenarios where session structures vary, such as shared device contexts or multi-session authentications.

In terms of future research directions, several promising avenues remain. Firstly, the size of the positive pool could be expanded by collecting additional sessions per user or incorporating more users into the dataset. A denser session matrix would facilitate a broader range of meaningful anchor-positive pairings, improving model generalization.

Secondly, the concept of temporal alignment can be extended by incorporating more sophisticated positional encoding strategies. Inspired by the success of transformers in sequential modeling tasks, future implementations could make use of transformer-

based encoders to model not only local temporal dependencies but also global sequence dynamics. Such architectures could naturally embed window indices or relative positions into the model input, obviating the need for handcrafted triplet sampling schemes.

Moreover, the integration of adaptive triplet selection algorithms—such as semi-hard or distance-weighted sampling—could further enhance training by selecting more informative and challenging negative examples dynamically. These approaches could replace or complement the fixed sampling strategies introduced in this thesis.

Finally, broader verification paradigms such as session-to-profile matching or incremental user modeling should be investigated. While this study focused on pairwise session verification, real-world deployments often require user authentication against aggregated behavioral profiles or evolving user representations. Extending the proposed architecture to support such scenarios may involve combining metric learning with memory-augmented or continual learning models.

In sum, the findings of this thesis provide an effective foundation for behavior-based mobile authentication systems, while also highlighting critical avenues for improving robustness, adaptability, and scalability in future research.

# REFERENCES


[1] I. M. Alsaadi, "Study on Most Popular Behavioral Biometrics, Advantages, Disadvantages and Recent Applications: A Review," *International Journal of Scientific & Technology Research*, vol. 10, no. 1, pp. 15–21, 2021.

[2] G. Pahuja and T. N. Nagabhushan, "Biometric Authentication & Identification through Behavioral Biometrics: A Survey," *JSS Academy of Technical Education*, 2019.

[3] D. Bhattacharyya, R. Ranjan, F. A. Alisherov, and M. Choi, "Biometric Authentication: A Review," *International Journal of u- and e- Service, Science and Technology*, vol. 2, no. 3, pp. 13–28, 2009.

[4] M. Abuhamad, A. Abusnaina, D. Nyang, and D. Mohaisen, "Sensor-Based Continuous Authentication of Smartphones' Users Using Behavioral Biometrics: A Contemporary Survey," *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 65–84, 2021.

[5] R. Zhang and Z. Yan, "A Survey on Biometric Authentication: Toward Secure and Privacy-Preserving Identification," *IEEE Access*, vol. 7, pp. 5994–6009, 2019.

[6] G. Stragapede, R. Vera-Rodriguez, R. Tolosana, and A. Morales, "BehavePassDB: Public Database for Mobile Behavioral Biometrics and Benchmark Evaluation," *Pattern Recognition*, vol. 134, 2023, Art. no. 109089.

[7] P. Bermejo, J. A. Gámez, and J. M. Puerta, "Speeding up incremental wrapper feature subset selection with Naïve Bayes classifier," *Knowledge-Based Systems*, vol. 55, pp. 140–147, 2014.

[8] Stragapede, G., Vera-Rodriguez, R., Tolosana, R., Morales, A., Acien, A., & Le Lan, G. (2022). Mobile behavioral biometrics for passive authentication. *arXiv preprint arXiv:2203.07300*.

[9] T. Feng, X. Zhao, B. Carbunar, and W. Shi, "Continuous Mobile Authentication Using Virtual Key Typing Biometrics," *Proceedings of the IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, 2013.

[10] F. Li, N. Clarke, M. Papadaki, and P. Dowland, "Active Authentication for Mobile Devices Utilising Behaviour Profiling," *Proceedings of the IEEE Security and Privacy Workshops*, 2013.

[11] Y. Song, M. Ben Salem, S. Hershkop, and S. J. Stolfo, "System-Level User Behavior Biometrics Using Fisher Features and Gaussian Mixture Models," *IEEE Security and Privacy Workshops*, 2013.

[12] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 136–148, 2013.

[13] G. Gündüz, M. Kaplan, and T. T. Temizel, "Feature Distribution-Based Touch Biometrics Using CNN and Siamese Networks," in *Proc. of the International Conference on Machine Learning and Data Engineering (iCMLDE)*, 2024.

[14] M. Nerini, E. Favarelli, and M. Chiani, "Augmented PIN Authentication through Behavioral Biometrics," *Sensors*, vol. 22, no. 13, 2022.

[15] Li, Y., Yang, J., Xie, M., Carlson, D., Jang, H., & Bian, J. (2022). Comparison of PIN- and pattern-based behavioral biometric authentication on mobile devices. *MILCOM*.

[16] Meng, W., Wong, D.S., Furnell, S., & Zhou, J. (2015). Surveying the development of biometric user authentication on mobile phones. *IEEE Communications Surveys & Tutorials*, 17(3), 1268–1293.

[17] C. Wang, Y. Xiao, X. Gao, L. Li, and J. Wang, "A Framework for Behavioral Biometric Authentication using Deep Metric Learning on Mobile Devices," *IEEE Transactions on Mobile Computing*, Early Access, 2023, doi:10.1109/TMC.2023.3241234. arora2023deepmetric

[18] A. Mahfouz, T. M. Mahmoud, and A. S. Eldin, "A Behavioral Biometric Authentication Framework on Smartphones," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS '17)*, pp. 923–925, 2017, doi:10.1145/3052973.3055160.

[19] Lis, K., Niewiadomska-Szynkiewicz, E., & Dziewulska, K. (2023). Siamese neural network for keystroke dynamics-based authentication on partial passwords. *Sensors*, 23(15), 6685.

[20] Choi, H.-S. (2023). Siamese Neural Network for User Authentication in Field-Programmable Gate Arrays (FPGAs) for Wearable Applications. *Electronics*, 12(19), 4030.

[21] Bah, C.U., Seyal, A.H., & Yahya, U. (2022). Combining PIN and biometric identifications as enhancement to user authentication in internet banking. *IET*.

[22] Sundararajan, K., & Woodard, D.L. (2018). Deep learning for biometrics: A survey. *ACM Computing Surveys*, 51(3), 1–34.

[23] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Biometrics Recognition Using Deep Learning: A Survey," *arXiv preprint arXiv:1912.00271*, 2021. [Online]. Available: https://arxiv.org/abs/1912.00271

[24] Gruschina, D., Schröder, L., & Busch, C. (2023). Surveying the Development of Biometric User Authentication on Mobile Phones. *ACM Computing Surveys*, 55(8), 1–37.

[25] K.-N. Nguyen, S. Rasnayaka, S. Wickramanayake, D. Meedeniya, S. Saha, and T. Sim, "Spatio-Temporal Dual-Attention Transformer for Time-Series Behavioral Biometrics," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 6, no. 4, pp. 591–601, 2024.

[26] M. Hu, K. Zhang, R. You, and B. Tu, "AuthConFormer: Sensor-based Continuous Authentication of Smartphone Users Using A Convolutional Transformer," *Computers & Security*, vol. 127, 2023, Art. no. 103122.

[27] X. Zhao, X. Xu, and J. Liu, "Fusion of Behavioral Biometrics for Continuous Authentication: A Review," *IEEE Access*, vol. 11, pp. 55267–55282, 2024.

[28] L. Fridman, S. Weber, R. Greenstadt, and M. Kam, "Active Authentication on Mobile Devices via Stylometry, Application Usage, Web Browsing, and GPS Location," *IEEE Systems Journal*, vol. 11, no. 2, pp. 513–521, 2017, doi:10.1109/JSYST.2015.2472579.

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

[30] J. Shen and N. Robertson, "BBAS: Towards large scale effective ensemble adversarial attacks against deep neural network learning," *Information Sciences*, vol. 569, pp. 469–478, 2021.

[31] F. Monrose and A. Rubin, "Keystroke dynamics as a biometric for authentication," *Future Generation Computer Systems*, vol. 16, no. 4, pp. 351–359, 2000.

[32] A. Acien, R. Tolosana, R. Vera-Rodriguez, A. Morales, and J. Fierrez, "TypeNet: Deep learning keystroke biometrics," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 1, pp. 57–70, 2022.

[33] D. Buschek, A. De Luca, and F. Alt, "Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI)*, 2015, pp. 1393–1402.

[34] G. Stragapede, R. Vera-Rodriguez, R. Tolosana, A. Morales, and J. Fierrez, "IJCB 2022 Mobile Behavioral Biometrics Competition (MobileB2C)," in *Proceedings of the IEEE International Joint Conference on Biometrics (IJCB)*, 2022.

[35] C. Wu and Y. Chen, "An Implicit Identity Authentication System Considering Changes of Gesture Based on Keystroke Behaviors," *Mathematical Problems in Engineering*, vol. 2015, pp. 1–8, 2015.

[36] M. J. Coakley, J. V. Monaco, and C. C. Tappert, "Keystroke biometric studies with short numeric input on smartphones," in *2016 IEEE 8th International*

*Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–6.

[37] E. P. Kukula, M. J. Sutton, and S. J. Elliott, "The Human–Biometric–Sensor Interaction Evaluation Method: Biometric Performance and Usability Measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 4, pp. 784–791, Apr. 2010.

[38] L. Gao, A. K. Bourke, and J. Nelson, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems," *Medical Engineering & Physics*, vol. 36, no. 6, pp. 779–785, Jun. 2014.

[39] S. Mekruksavanich and A. Jitpattanakul, "Wearable Sensor-Based Behavioral User Authentication Using a Hybrid Deep Learning Approach with Squeeze-and-Excitation Mechanism," *Computers*, vol. 13, no. 12, Art. no. 337, 2024.

[40] M. Goldstein and S. Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data," *PLOS One*, vol. 11, no. 4, pp. e0152173, Apr. 2016.

[41] K. S. Killourhy and R. A. Maxion, "Comparing Anomaly-Detection Algorithms for Keystroke Dynamics," in *Proceedings of the IEEE/IFIP International Conference on Dependable Systems & Networks (DSN)*, 2010, pp. 125–134.

[42] E. Al-Mahadeen, M. Alghamdi, A. S. Tarawneh, M. A. Alrowaily, M. Alrashidi, I. S. Alkhazi, A. Mbaidin, A. A. Alkasasbeh, M. A. Abbadi, and A. B. Hassanat, "Smartphone User Identification/Authentication Using Accelerometer and Gyroscope Data," *Sustainability*, vol. 15, no. 13, p. 10456, Jul. 2023.

[43] S. Sprager and M. B. Juric, "Inertial Sensor-Based Gait Recognition: A Review," *Sensors*, vol. 15, no. 9, pp. 22089–22127, Sep. 2015, doi:10.3390/s150922089.

[44] J. L. Wayman, "Digital signal processing in biometric identification: a review," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, 1998, pp. 2009–2012, doi:10.1109/ICASSP.1998.681737.

[45] S. Yang and F. Deravi, "Wavelet-Based EEG Preprocessing for Biometric Applications," in *Proc. 2013 Fourth International Conference on Emerging Security Technologies (EST)*, 2013, pp. 43–46, doi:10.1109/EST.2013.18.

[46] B. Bala and S. Behal, "A Brief Survey of Data Preprocessing in Machine Learning and Deep Learning Techniques," in *Proc. 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Oct. 2024, pp. 1–6, doi:10.1109/I-SMAC61858.2024.10714767.

[47] D. Ravi, C. Wong, B. Lo, and G.-Z. Yang, "A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, Dec. 2016, doi:10.1109/JBHI.2016.2633287.

[48] S. A. Abdulrahman and B. Alhayani, "A Comprehensive Survey on the Biometric Systems Based on Physiological and Behavioural Characteristics," *Materials Today: Proceedings*, 2021. https://doi.org/10.1016/j.matpr.2021.07.005

[49] J. Fierrez, R. Tolosana, and R. Vera-Rodriguez, "Multiple classifiers in biometrics. Part 1: Fundamentals and review," *Information Fusion*, vol. 44, pp. 57–64, 2018.

[50] L. Nanni, A. Lumini, M. Ferrara, and R. Cappelli, "Combining biometric matchers by means of machine learning and statistical approaches," *Neurocomputing*, vol. 149, Part B, pp. 526–535, Feb. 2015, doi:10.1016/j.neucom.2014.08.021.

[51] P. P. Jena, K. N. Kattigenahally, S. Nikitha, S. Sarda, and H. Y., "Multimodal biometric authentication: Deep learning approach," in *Proc. 2021 Int. Conf. Commun. Electron. Syst. (ICCES)*, pp. 1172–1178, 2021, doi:10.1109/ICCES51350.2021.9702724.

[52] X. Cheng, S. Tulyakov, and V. Govindaraju, "Multiple-sample fusion of matching scores in biometric systems," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, pp. 100–105, 2011, doi:10.1109/CVPRW.2011.5981827.

[53] H. Alshazly, C. Linse, E. Barth, and T. Martinetz, "Ensembles of deep learn-

ing models and transfer learning for ear recognition," *Sensors*, vol. 19, no. 19, Art. no. 4139, 2019, doi:10.3390/s19194139.

[54] T. Zhao, Y. Wang, J. Liu, and Y. Chen, "TrueHeart: Continuous authentication on wrist-worn wearables using PPG-based biometrics," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, Jul. 2020, pp. 1–10, doi:10.1109/INFOCOM41043.2020.9155526.

[55] W. Neiswanger and E. P. Xing, "Post-Inference Prior Swapping," in *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017.

[56] M. F. Wahid, R. Tafreshi, and R. Langari, "A multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 2, pp. 451–458, Feb. 2020, doi:10.1109/TNSRE.2019.2961072.

[57] M. Gadaleta and M. Rossi, "IDNet: Smartphone-based Gait Recognition with Convolutional Neural Networks," *arXiv preprint* arXiv:1606.03238, 2016.

[58] L. Yang, C. Li, R. You, B. Tu, and L. Li, "TKCA: a timely keystroke-based continuous user authentication with short keystroke sequence in uncontrolled settings," *Cybersecurity*, vol. 4, art. 13, 2021, doi:10.1186/s42400-021-00075-9.

[59] A. Ray, D. Hou, S. Schuckers, and A. Barbir, "Continuous Authentication based on Hand Micro-movement during Smartphone Form Filling by Seated Human Subjects," in *Proc. 7th Int. Conf. on Information Systems Security and Privacy (ICISSP)*, SCITEPRESS, 2021, pp. 424–431, doi:10.5220/0010225804240431.

[60] A. K. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[61] W. Neiswanger, C. Wang, and E. P. Xing, "Asymptotically Exact, Embarrassingly Parallel MCMC," in *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2014.

[62] T. S. Huang, G. J. Yang, and G. Y. Tang, "A fast two-dimensional median filtering algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 1, pp. 13–18, 1979.

[63] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, 1977.

[64] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, 1997.

[65] Q. Zhang, F. Lin, and H. S. Seah, "Detecting and imaging irregularities in time-series data," in *Proceedings of the 2018 International Workshop on Advanced Image Technology (IWAIT)*, Jan. 2018, pp. 1–4, doi:10.1109/IWAIT.2018.8369784.

[66] Farahani, B., Firouzi, F., & Chakrabarty, K. (2022). Repairing timestamps in wireless IoT data: Challenges, algorithms, and system-level evaluation. *Information Fusion*, 87, 40–53.

[67] R. M. Hidalgo, J. G. Fernandez, R. R. Rivera, and H. A. Larrondo, "A simple adjustable window algorithm to improve FFT measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 1, pp. 31–36, 2002.

[68] J. Walton, G. Cronin, and K. J. Stafford, "Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour," *Computers and Electronics in Agriculture*, vol. 156, pp. 283–291, 2019.

[69] N. E. Golyandina and A. A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Berlin, Heidelberg: Springer, 2013, doi:10.1007/978-3-642-34913-3.

[70] J. D'Arcy, "Introducing SSA for Time Series Decomposition," *Kaggle Notebook*, 2017. [Online]. Available: https://www.kaggle.com/code/jdarcy/introducing-ssa-for-time-series-decomposition

[71] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[72] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.

[73] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, "Signature verification using a 'Siamese' time delay neural network," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 6, 1993.

[74] R. Giot and A. Rocha, "Siamese Networks for Static Keystroke Dynamics Authentication," in *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, Delft, Netherlands, Dec. 2019, pp. 1–6, doi:10.1109/WIFS47025.2019.9035100.

[75] Y. Li, C. L. P. Chen, and T. Zhang, "A Survey on Siamese Network: Methodologies, Applications, and Opportunities," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 6, pp. 887–907, Dec. 2022, doi:10.1109/TAI.2021.3131554.

[76] Z. Zeng, J. Wan, and H. Liu, "An entropy-based filtering approach for airborne laser scanning data," *Infrared Physics & Technology*, vol. 75, pp. 87–92, 2016.

[77] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, 2017.

[78] H. Xuan, A. Stylianou, X. Liu, and R. Pless, "Hard negative examples are hard, but useful," *arXiv preprint arXiv:2007.12749*, 2021.

[79] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.

[80] P. S. Teh, A. B. J. Teoh, T. S. Ong, and H. F. Neo, "Statistical Fusion Approach on Keystroke Dynamics," in *Proc. 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based Systems*, Shanghai, China, Dec. 2007, pp. 918–923, doi:10.1109/SITIS.2007.151.

# APPENDIX A

## Sensor Length Distributions and Sampling Uniformity

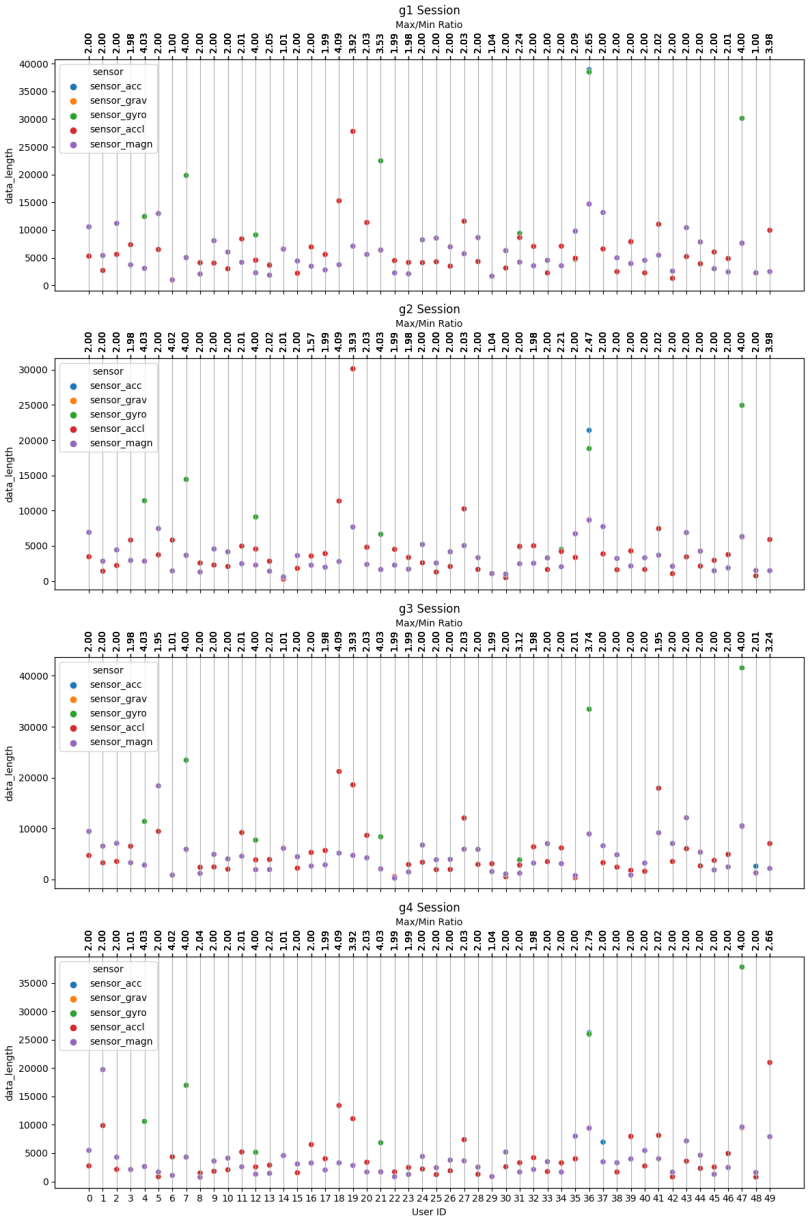



Figure A.1: Sensor sequence lengths before frequency downsampling

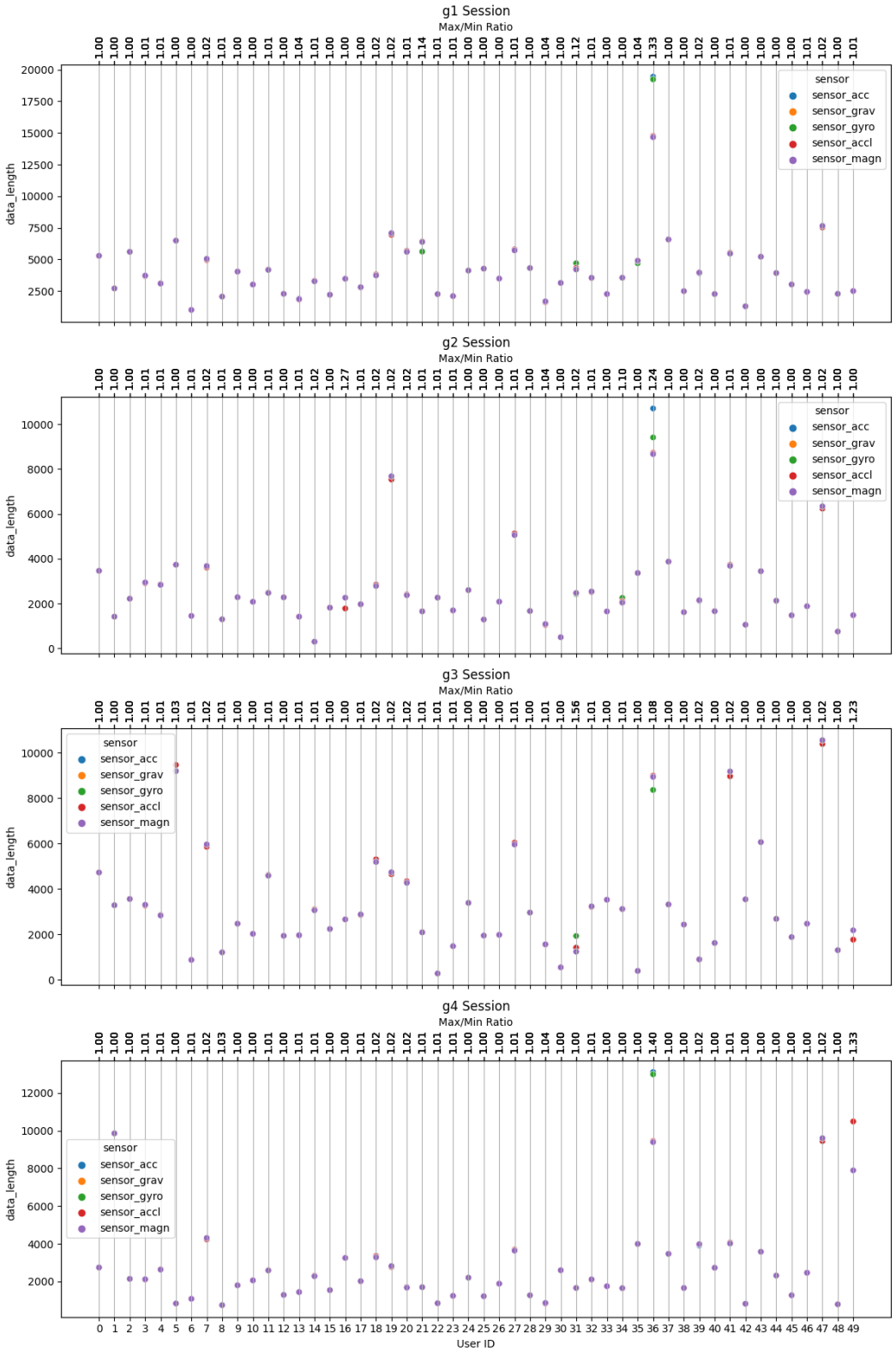



Figure A.2: Sensor sequence lengths after frequency downsampling

# APPENDIX B

## BASELINE MODEL HYPERPARAMETER TUNING

Table B.1: Bayesian optimization ranges and best configuration for XGBoost hyperparameters.

| Parameter | Range | Step | Best Value |
|---|---|---|---|
| `n_estimators` | [50, 300] | 25 | 150 |
| `max_depth` | [3, 10] | 1 | 6 |
| `learning_rate` | [0.01, 0.3] | - | 0.1 |
| `subsample` | [0.5, 1.0] | - | 0.85 |
| `colsample_bytree` | [0.5, 1.0] | - | 0.75 |
| `gamma` | [0, 5] | 0.5 | 1.0 |