



ARTICLE

Mobile Touch Dynamics–Based User Classification Using Machine Learning and Fusion Techniques

Animaw Kerie Aseres^{1,2,*}, Asrat Mulatu Beyene^{3,2} and Lemlem Kassa Tegegne^{1,2}

¹Department of Software Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

²High Performance Computing and Big Data Analytics Center of Excellence, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

³Department of Electrical and Computer Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

*Corresponding Author: Animaw Kerie Aseres. Email: animaw.kerie@kue.edu.et

Received: 02 June 2026; Accepted: 23 July 2026; Published: 21 August 2026

ABSTRACT: Conventional multi-factor and one-time authentication approaches, such as passwords and one-time passwords (OTPs), have become increasingly vulnerable to advanced attack methods, motivating the need for continuous authentication (CA) systems that can verify user identity throughout an active session rather than only at login. For such a system to be effective, it must analyze user behavior reliably and in real time. This paper presents a novel approach to implementing CA on mobile devices using tap and swipe behavioral biometrics combined with machine learning (ML) and multimodal fusion. The dataset was collected from 400 volunteer participants using BahriApp, a custom-developed Android application (Bahri meaning “behavior” in Amharic). Several models were then trained to classify user identity, including Logistic Regression, SVM, Random Forest, and deep learning models (1D-CNN and LSTM), evaluated with feature-level and decision-level fusion of the tap and swipe datasets. Feature-level fusion using LSTM performed best, achieving 97.1% accuracy and a 4.4% Equal Error Rate (EER), balancing usability and security; weighted decision-level fusion achieved 96.5% accuracy and a 4.8% EER. The main contributions of this paper are: a custom behavioral biometrics dataset constructed from African participants; comparative benchmarks across machine learning models and fusion strategies; and insights that inform the design of an adaptive continuous authentication scheme. The study’s main limitation is dataset size and session coverage: of the 400 participants, only 29.75% and 20.75% provided 15 or more sessions in the tap and swipe datasets, respectively. Future work will focus on larger and more diverse multimodal datasets.

KEYWORDS: Continuous authentication; behavioral biometrics; gait; fusion techniques

1 Introduction

Authentication systems can be broadly categorized into static and dynamic approaches. Static authentication, such as passwords, PINs, and one-time passwords (OTPs), verifies a user’s identity only once at login and provides no further protection once a session has begun. This is the central problem addressed in this paper: because static, one-time verification cannot detect a change of user within an already-authenticated session, systems remain exposed to session hijacking, device theft, and insider threats for as long as the session stays open. Continuous authentication (CA) addresses this gap by monitoring user behavior throughout the session and re-verifying identity in real time, thereby reducing the window of vulnerability after the initial login [1]. Behavioral biometrics is particularly well suited to this task because, unlike physiological biometrics, it can be captured passively throughout an active session [2]. The core feature of behavioral biometrics is its non-intrusive nature. Meaning, users are not required to remember additional

credentials such as OTPs, as the system passively monitors their natural interactions. However, there is also an inherent challenge in the data collection process because, in physical biometrics, it is easy to capture a natural trait. On the other hand, in behavioral biometrics, users are required to do some activity to gather their continuous behavior, specifically those related to their human-computer interaction.

As technology advances, the integration of behavioral biometrics into authentication frameworks is becoming increasingly feasible and impactful [3]. These traditional, one-time credentials have formed the foundation of authentication in the cybersecurity industry for decades, but the rapid advancement and sophistication of attacks have exposed their inherent vulnerabilities [4]. As a result, an urgent demand for strong authentication techniques has been introduced in recent years. For example, Microsoft has introduced the concept of passkeys in this regard to offer a frictionless sign-in experience by eliminating passwords¹.

Fusion strategies have greatly enhanced the potential of behavioral biometric systems by combining multiple types of data or techniques to improve accuracy and robustness. These strategies typically involve integrating different modalities, such as keystroke dynamics (the rhythm and speed of typing) and mouse movements (how a user moves and clicks on the mouse), to create a more comprehensive picture of a user's behavior. This combination is valuable because each data type has its own strengths and using them together helps mitigate the effects of noise or variability in individual data sources. Fusion can occur at different stages of the process: at the raw data level, where the unprocessed data from different sources are combined before analysis; at the feature level, where specific features from each modality are extracted and merged for further analysis; or at the decision level, where separate models for each data type make decisions that are then combined into a final authentication verdict [5].

In this research, a custom dataset is developed combining the tap and swipe datasets using a self-developed mobile application named BahriApp. Bahri or “ባሕሪ” in Amharic means behavior. Mainly, users provide their tapping patterns (custom behavior) based on gamification through identification of a picture of mammals and swiping patterns top or bottom and left or right based on identification of African countries' flags on their smartphones. This paper aims to answer the following research questions:

- Which machine learning and deep learning models provide the most reliable real-time biometric fusion?
- Which fusion techniques will result in better accuracy in the user's identity classification?

2 Related Works

Touch dynamics such as swipe velocity, touch pressure, and screen contact area have shown high accuracy in user identification. Implicit authentication systems often utilize these features for transparent, continuous verification using built-in sensors like accelerometers and gyroscopes [6,7]. However, to improve accuracy and resilience, multimodal continuous authentication (CA) systems integrate features at the feature level or decision level. This fusion leverages the strengths of each modality to reduce false positives and increase spoofing resistance [8]. For instance, combining keystroke dynamics with gait allows compensation when one modality is affected by environmental factors [7,9]. Ongoing research also emphasizes feature selection, dimensionality reduction, and adaptive weighting to improve system performance across heterogeneous environments [10]. However, much of this literature evaluates fusion strategies on datasets collected in controlled laboratory settings or from populations concentrated in North America, Europe, and East Asia, and rarely reports latency or memory figures needed to judge real-world, on-device feasibility; this leaves an open gap regarding how such approaches transfer to naturalistic, BYOD usage and to underrepresented populations, which the present study specifically targets.

¹<https://learn.microsoft.com/en-us/entra/identity/authentication/concept-authentication-passkeys-fido2>.

Machine learning (ML) has become a cornerstone of continuous authentication (CA) systems. Classical ML methods, such as Support Vector Machines (SVMs) and Random Forests, remain foundational in CA due to their interpretability and efficiency in handling structured datasets. For instance, template-based ML pipelines are experimented with to simplify CA system design by auto-generating feature-extraction code from sensor specifications [11]. Compared to deep learning, SVMs struggle with high-dimensional and unstructured behavioral data, showing degraded accuracy as the feature count grows [12]. Random Forests are more robust to noisy, high-dimensional inputs, but require substantial feature engineering; they also lack native support for temporal dependencies common in CA time-series data [13,14]. In real-world applications, SVM and Random Forest classifiers achieve low-latency decisions ideal for device access control and desktop-security scenarios [15]. A common limitation across this body of classical ML work is that reported gains in interpretability and low latency are typically demonstrated on relatively small, single-context datasets, and comparisons across SVM, Random Forest, and Logistic Regression are rarely benchmarked side by side under one consistent evaluation protocol, which motivates the unified classical-ML benchmark presented later in this paper.

On the other hand, deep learning models, especially Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have significantly advanced the continuous authentication innovation by capturing complex spatial and temporal patterns in user behavior. For example, a recent evaluation shows CNN-LSTM hybrids outperform traditional models on raw motion-sensor and touch data, yielding over 97% accuracy on benchmarking datasets [16]. In real-World applications, smartphone facial-authentication modules have routinely employed lightweight CNNs for continuous re-verification [17].

Omnidirectional modeling is a recently introduced technique that removes directional preference during the aggregation of touchscreen strokes, thus allowing for accurate continuous verification processes by simulating natural swipe actions. The feed-forward approach for stroke aggregation allows for low-latency CA² by processing features of touches, such as pressure and coordinates, into a unified model that enables low-latency CA on a mobile device. These approaches improve the original touch dynamics by considering variable-length sessions [18]. In another recent work, contrastive learning formulations reimplemented CA by learning similarities using Siamese architectures that were pretrained with Temporal Masked Autoencoders on raw multi-sensor data. Self-supervised learning on pretraining-derived temporal patterns from sequences with no labels was successfully transferred for authentication tasks with over 98% accuracy [19].

Taken together, the reviewed deep learning and contrastive-learning approaches report strong accuracy, but three limitations recur across this body of work: (i) benchmark datasets are predominantly drawn from North American, European, or East Asian participant pools, leaving open how well these models generalize to other populations; (ii) on-device latency and memory costs of CNN-LSTM and contrastive/Siamese architectures are seldom reported, making real-world mobile deployment difficult to assess from the published results alone; and (iii) evaluation is typically performed under a single, unspecified usage context, without examining whether accuracy is stable across postures or physical activities such as sitting, standing, or walking. The present study is designed to directly address the first of these gaps, and explicitly discusses the second and third as limitations and directions for future work.

3 Materials and Methods

This study uses an experimental design to systematically evaluate the performance of various machine learning models and fusion strategies for continuous authentication. The primary objectives are to enhance system reliability and optimize models for real-time authentication with minimal computational overhead.

²CA refers to continuous authentication.

Key steps in the proposed research pipeline follow a well-known experiment flow in machine learning, including data collection, preprocessing, model training, and performance evaluation tasks, as outlined in the flowchart depicted in Fig. 1.

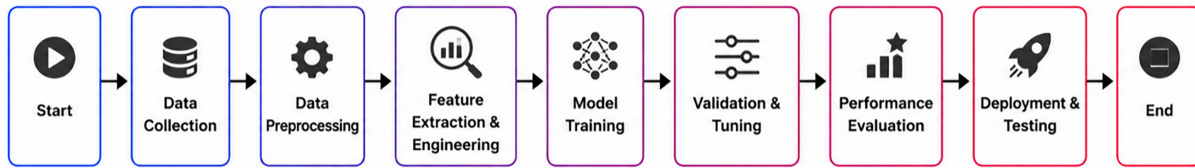


Figure 1: Pipeline of the research design.

Data Collection

A. Selection of participants and dissemination of the data collection tool: A multimodal dataset incorporating touch and tap gestures was collected from different users, mainly senior university students learning at HiLCoE School of Computer Science³ and Kombolcha Institute of Technology⁴ in Ethiopia, but also academic and administrative staff in other universities. A total of 400 participants were recruited to participate for eight months from January 2025 to Augst 2025. The cohort was selected based on diversity in age, gender, and smartphone proficiency. Participants' demographics are illustrated in Fig. 2 below. The native Android APK file of the mobile application developed for this purpose was distributed to them via a dedicated Telegram group. Users were allowed to install the app and play different games in natural settings with informed consent and without intervention after continuous preliminary awareness was created via phone calls and a Telegram group. The Bring Your Own Device (BYOD) approach was used so that users could use their smartphone devices anywhere, in any setup. No specific venue or controlled environment was introduced to minimize smartphone usage disruption and restrictions so that the real behavior of users was captured. This diversification supported us in exploring variability to keep the natural patterns that users feel, including their variability of emotions while they play the game, variability of environment, usage of both hands while playing the game, and phone grips.

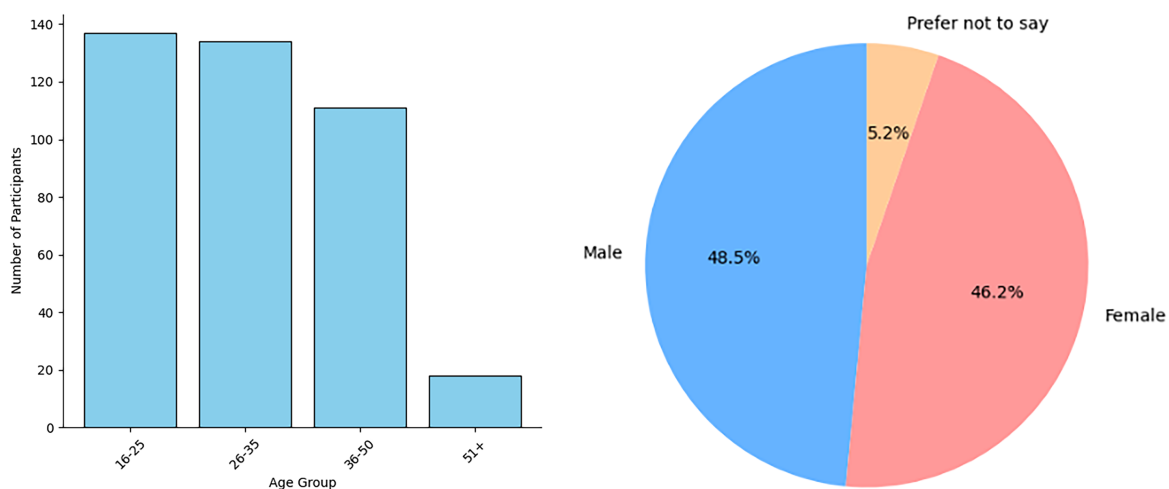


Figure 2: Age demography of participants.

³<https://hilcoe.net>.

⁴<https://wu.edu.et/kiot>.

- B. The data collection method:** The data collection process required the installation of a pre-defined mobile app specifically designed for this research purpose. There were two structured interaction sessions designed for data collection purposes. The first was based on the tap interaction task to collect the discrete features of user interactions. The second was based on the swipe task to collect continuous features. Records of each interaction were encrypted on the device, locally stored, and instantly transmitted to a secure cloud-based storage solution provided by Firebase. The end-to-end pipeline is outlined in Fig. 3.

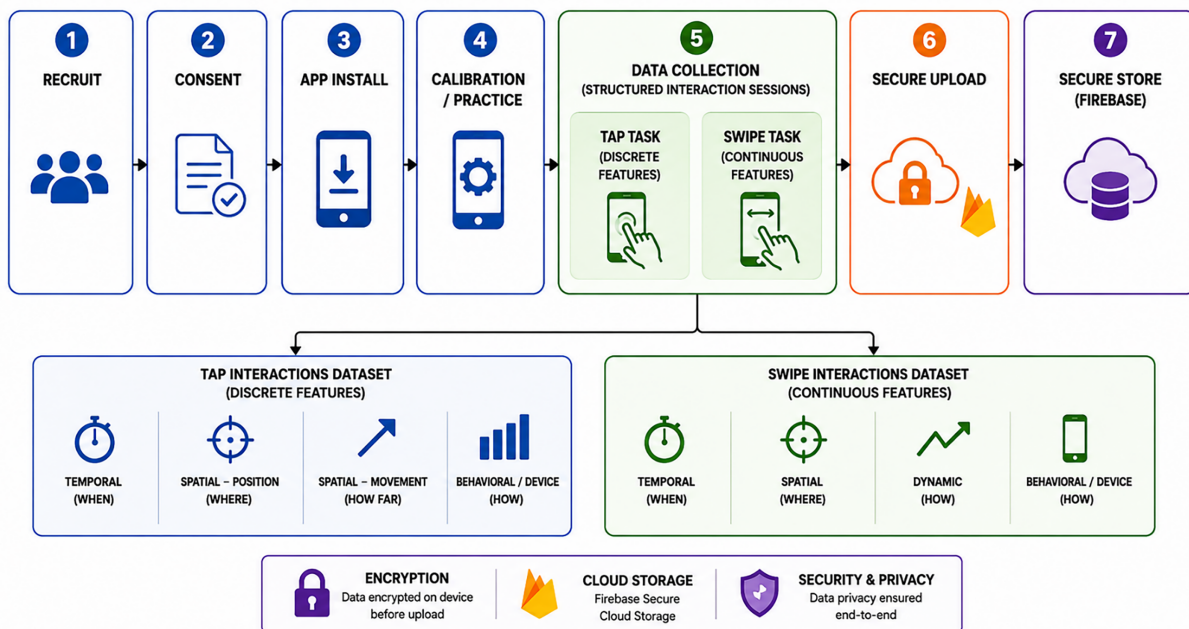


Figure 3: Data collection pipeline and datasets infographic.

- C. The data collection protocol:** As shown in Figs. 4 and 5, only participants who completed **15 or more sessions per game** were included in this study. This inclusion criterion was adopted to ensure that each participant had sufficient behavioral data to capture stable and representative touch interaction patterns. Having an adequate number of sessions improves the reliability of feature extraction and model training, thereby supporting more robust continuous authentication based on behavioral biometrics.
- D. Shape of the dataset:** The tap dataset, as outlined in Fig. 3, is composed of a set of variables that describe tap interactions based on the user's temporal and spatial behavior. The variables include tap duration, press duration, and release duration, which describe the temporal behavior of tap interactions. The other variables used are initial and final tap positions defined on a global screen and local component coordinates on the X and Y axes. The variables additionally comprised displacement on global and local axes with respect to horizontal and vertical components. The dataset additionally includes variables describing behavior, given as tap speed, tap drift, and tap repeat rate, with a two-second window. The dataset is made device-independent by including variables such as a normalized tap position and a target size area. Conversely, the swipe interactions dataset includes all information about swipe interactions, such as their time dimensions, including swipe start and finish time, swipe duration, and inter-swipe interval impacts. It includes information about swipe interactions along with their spatial dimensions, such as XY coordinates for swipe start and finish positions with a swipe interaction, swipe distance, swipe path length, proportionate coverage for a screen with swipe length, and swipe length, etc. The

dataset also includes information about swipe interactions along with their dynamic dimensions, such as swipe speed, swipe acceleration, swipe deceleration, swipe jerk, and swipe angle descriptions.

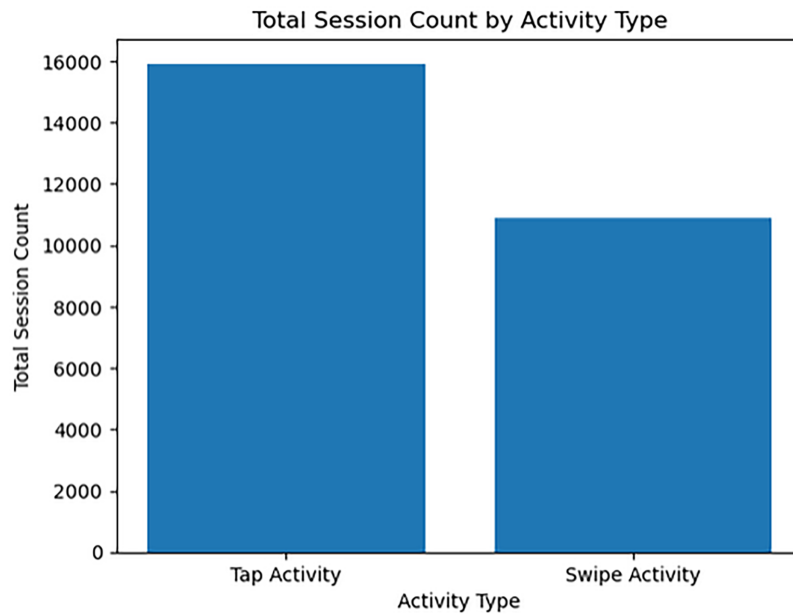


Figure 4: Total number of sessions gathered by all users that fulfill the best minimum criteria.

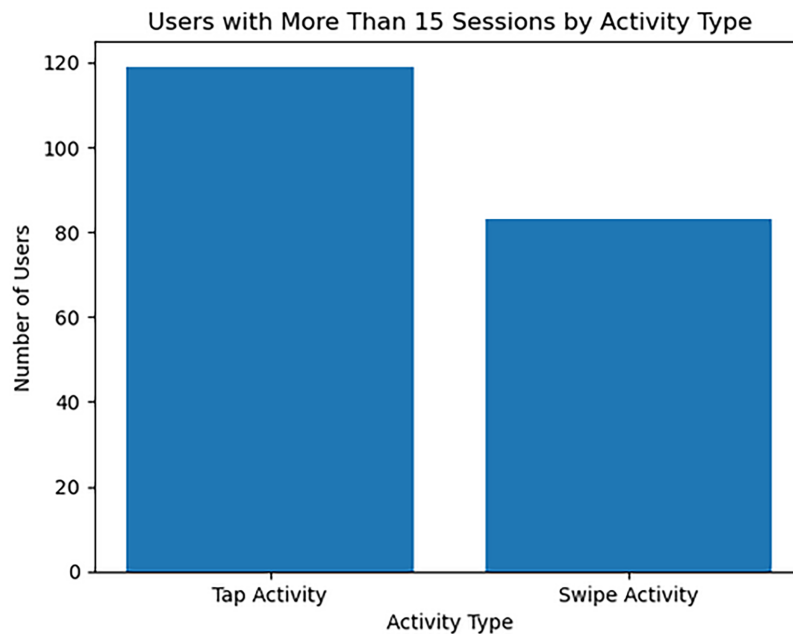


Figure 5: Total number of users who provided the minimum best session amount.

- E. Data preprocessing:** Before analysis, systematic data cleaning and preparation were performed. This step ensured consistency and cleaned up raw tap and swipe interaction data collected to avoid issues related to temporal, spatial, and behavioral consistency. To eliminate noise and non-behavioral observations

from sensors, outlier observations were removed at a session and user level based on predefined and static thresholds and statistics to eliminate observations with improbable latency, speed, and acceleration. As a heterogeneous set of devices was recruited, coordinates were normalized against screen resolutions to provide a device-agnostic representation, and generated kinesthetic attributes like speed and displacement were checked for consistency. Moreover, all timestamps were converted into Unix format. Mainly, users who played games for fewer than 15 sessions were discarded to get more convenient, stable user behavior patterns. In Fig. 6, a histogram of Tap duration is depicted, and in Fig. 7, a box plot of Swipe Velocity is shown to demonstrate the sample dataset exploration.

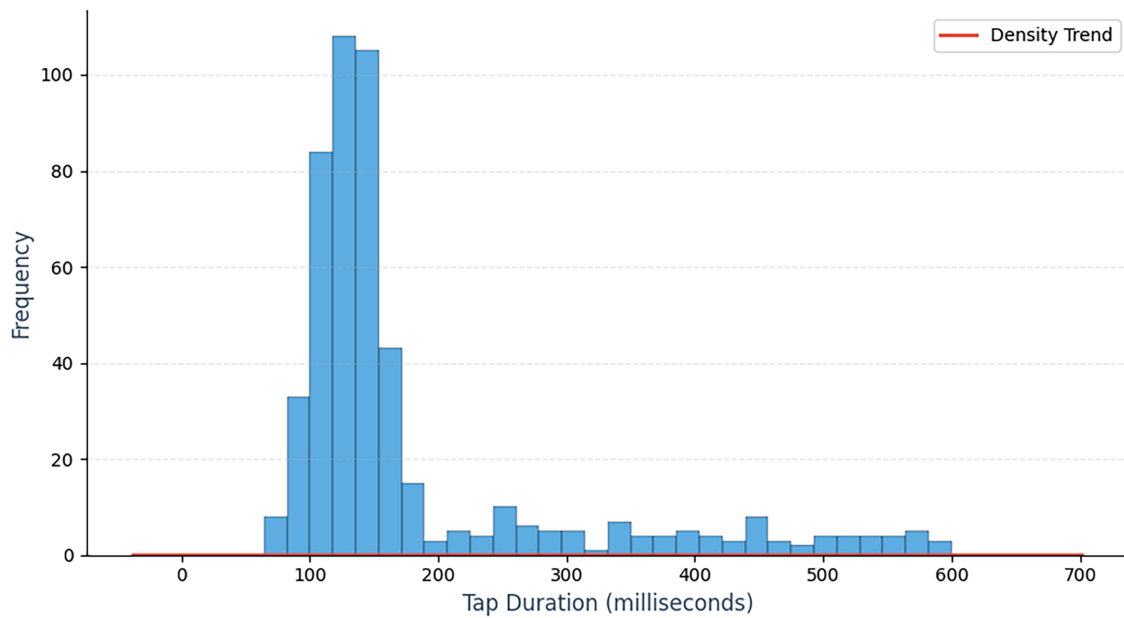


Figure 6: Histogram of tap duration in ms.

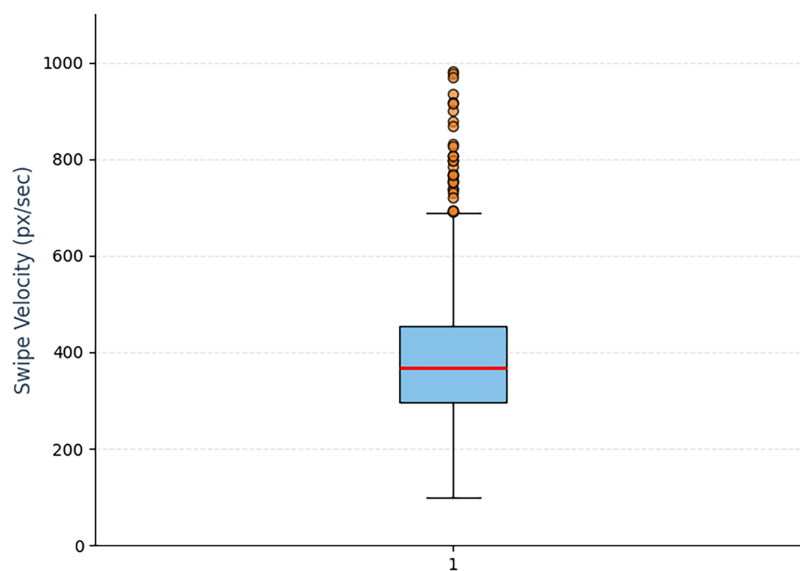


Figure 7: Boxplot of swipe velocity (pixels/sec).

- F. Feature descriptions and visualizations:** The raw interaction data were transformed into meaningful behavioral features. For example, taps' features included tap duration, inter-tap interval, and spatial error, etc. For swipes, features included velocity, trajectory curvature, path length, pressure dynamics, etc.
- G. Experimental setup:** The experiments were conducted on a workstation equipped with the hardware and software specifications summarized in [Table 1](#):
- H. Hyperparameter configuration:** Each model was tuned using grid search optimization for optimal performance based on the key hyperparameters listed in [Table 2](#).
- I. Evaluation metrics:** Model performance was evaluated using standard metrics widely adopted in biometric authentication research, as listed in [Table 3](#).

Table 1: Hardware and software configuration.

Component	Specification
Processor	Intel Core i7-12700K (12 Cores, 3.6 GHz)
GPU	NVIDIA RTX 3060 (12 GB VRAM)
RAM	32 GB DDR4
Operating System	Windows 11 Pro, 64-bit
Programming Language	Python 3.10
ML/DL Libraries	Scikit-learn 1.3, TensorFlow 2.12, PyTorch 2.0
Mobile Deployment Tool	TensorFlow Lite (for Android implementation)

Table 2: Model hyperparameters.

Model	Key Hyperparameters
Logistic Regression	Regularization (L2, C = 1.0)
SVM (RBF)	Kernel = RBF, C = 10, Gamma = 0.01
Random Forest	Trees = 200, Max Depth = 20, Min Samples = 5
CNN (1D)	3 Conv layers, Kernel size = 5, Dropout = 0.3
LSTM	Hidden units = 128, Dropout = 0.2, Optimizer = Adam
Autoencoder	Layers: [64–32–64], Activation = ReLU, Epochs = 50
One-Class SVM	Kernel = RBF, ν = 0.1, Gamma = 0.05

Table 3: Evaluation metrics.

Metric	Description
Accuracy	Percentage of correct classifications (genuine vs. impostor)
False Acceptance Rate (FAR)	The rate at which impostors are incorrectly accepted
False Rejection Rate (FRR)	The rate at which genuine users are incorrectly rejected
Equal Error Rate (EER)	Point where FAR = FRR, balancing security & usability
Latency	Average decision-making time (ms)
Memory Footprint	Model storage size on the device

J. Experimental protocol: To critically analyze the proposed system, a comprehensive procedure for an experiment has been designed. This includes partitioning and validation of the model, fusion approaches, and evaluation at the application level. The dataset has been divided into three distinct subsets: 70% for the training process, 15% for validation, and 15% for the test process. The overall system architecture showing the research process is depicted in Fig. 8. However, for this preliminary research, only two datasets were used among the eleven datasets constructed and shown on the data visualisation dashboard in Fig. 8.

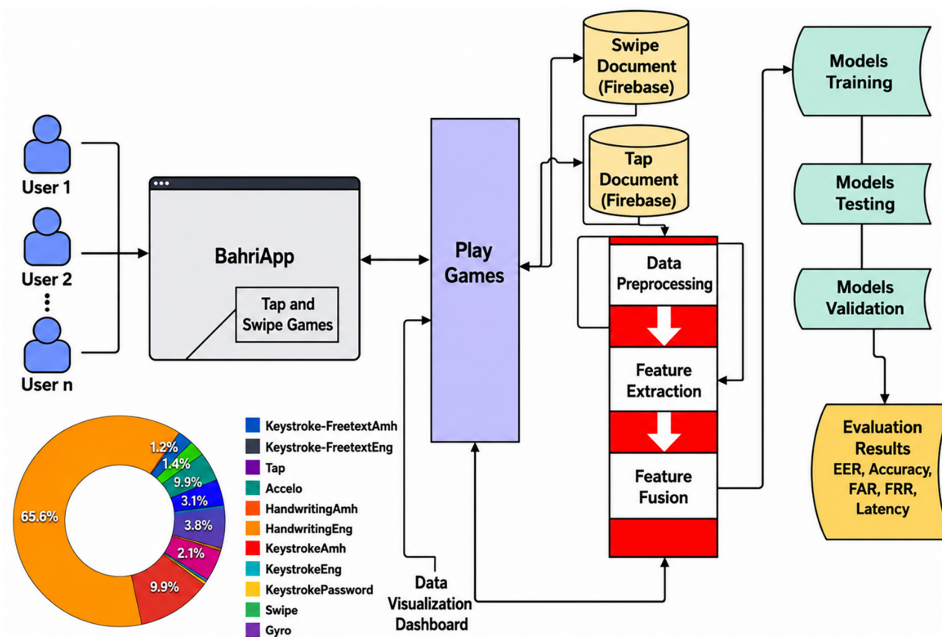


Figure 8: End-to-end experiment workflow.

To avoid the leakage of identity information and to make the system more robust against impostors, the strategy of non-user-disjoint data partitioning was used. All the data samples of the same user were distributed across all subsets. A global model was then built using data from all users.

In terms of hyperparameter tuning, a 5-fold cross-validation method was employed on a combination of both train and validation data, with a constraint to avoid user-disjoint divisions. This method allowed for effective evaluation of model performance and helped prevent overfitting concerns.

Two methods of fusion were explored to evaluate the efficacy of multimodal fusion. In feature-level fusion, the features were merged by stacking the feature vectors obtained from each mode before classification. In decision-level fusion, a weighted average of the results of each single-modal classifier was used.

4 Results and Discussion

4.1 Classical Machine Learning Models

Classical machine learning models serve as strong baselines due to their efficiency and interpretability. Logistic Regression, Support Vector Machines (SVM), and Random Forest were trained and evaluated on the extracted tap and swipe features. As a result, as shown in Table 4, Random Forest produced the best results, with an overall accuracy of 92.7 percent and the lowest EER of 8.2 percent. Its robustness against

the non-linear aspects of behavioral biometric information was exemplary. Logistic Regression performed poorly, thereby emphasizing the inadequacy of its linear model structure, which, despite its efficiency, was not sufficient for handling behavioral biometric information. Moreover, SVM offered favorable results, but its high memory requirement made it not very desirable for resource-constrained systems.

Table 4: Performance of classical ML models.

Model	Accuracy (%)	FAR (%)	FRR (%)	EER (%)	Latency (ms)	Memory (MB)
Logistic Regression	86.5	12.1	15.0	13.5	25	2
SVM (RBF)	90.2	9.8	10.5	10.1	42	12
Random Forest	92.7	7.5	8.9	8.2	55	18

4.2 Deep Learning Models

Deep learning models were also evaluated to capture the sequential and temporal dependencies in swipe and tap interactions. A one-dimensional Convolutional Neural Network (1D-CNN) and a Long Short-Term Memory (LSTM) network were implemented. As a result of the experiment, it was ensured that the deep learning models performed much better compared to the traditional machine learning models in terms of user identification through their behavioral biometrics. Further, the LSTM model provided the best results with an accuracy of 96.3% and an Equal Error Rate of 4.8%, indicating its effectiveness in encoding complicated behavioral sequences. The capability to encode temporal relationships in the behavioral data made it an excellent choice for encoding the swipe velocity and gesture pace features. Though it had a slightly lower accuracy level compared to the LSTM model results, the CNN model had the best latency and memory efficiency, and therefore the best choice for the mobile environment for system deployment. The complete performance figures for both deep learning models are reported in Table 5. Fig. 9 presents the ROC curves of the classical ML models for comparison, while Figs. 10 and 11 show the training-vs.-validation accuracy and loss curves of the deep learning models, respectively; both models converge steadily without evidence of overfitting, with validation curves closely tracking training curves throughout training.

Table 5: Performance of deep learning models.

Model	Acc (%)	FAR (%)	FRR (%)	EER (%)	Latency (ms)	Memory (MB)
1D-CNN	95.1	6.2	5.8	6.0	68	25
LSTM	96.3	5.0	4.7	4.8	85	32

4.3 Unsupervised Models

Since impostor behavior is often unpredictable, unsupervised methods such as Autoencoders and One-Class SVM were also tested. These models learn only the “genuine” user’s behavior and detect deviations as impostor attempts. The autoencoder-based method performed moderately with an accuracy of 89.4% and proved its effectiveness in recognizing impostor attacks; yet, its performance was suboptimal compared with other methods based on supervised learning. One-Class SVM performed poorly and proved that unsupervised learning alone is inadequate in highly secure systems like biometric authentication. Despite this, both techniques can be useful in systems with minimal labeled training data. The DET curves comparing the two unsupervised models are shown in Fig. 12.

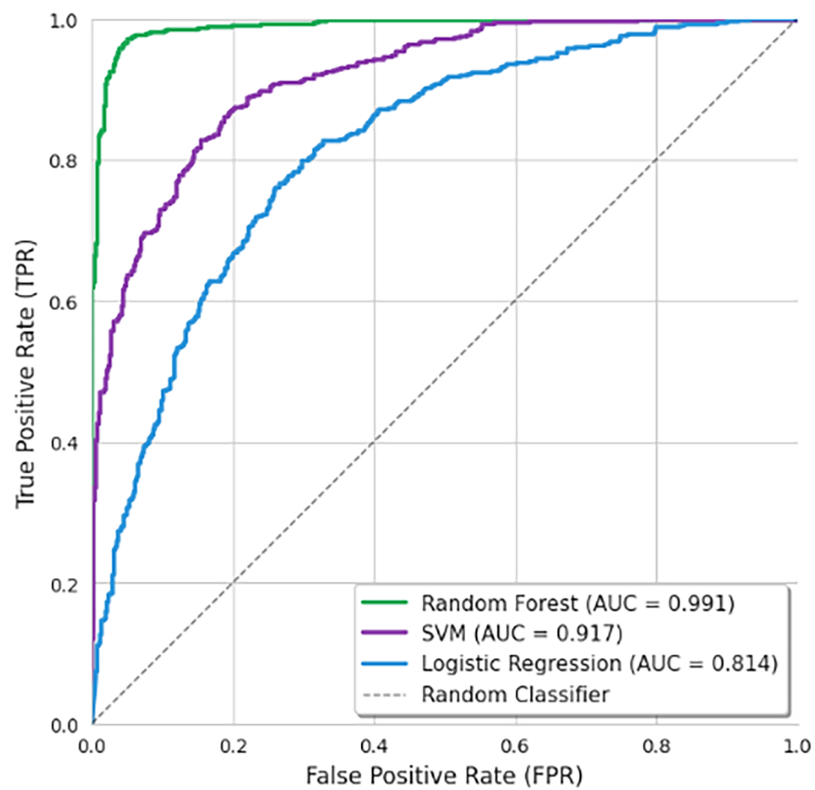


Figure 9: ROC curves of classical ML models.

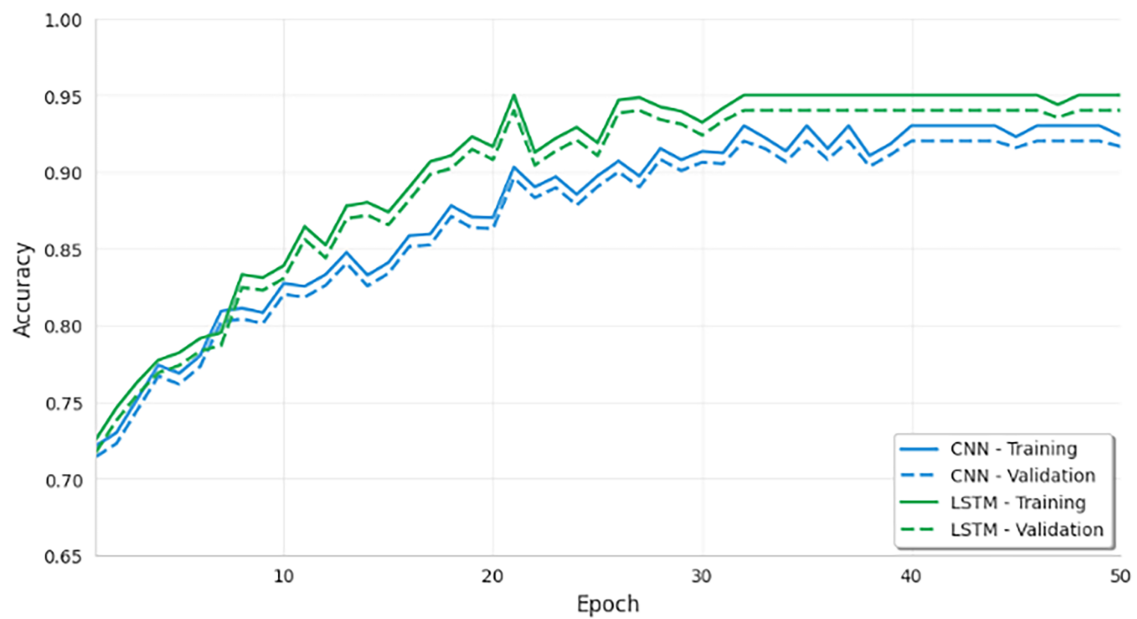


Figure 10: Training vs. validation accuracy for deep models.

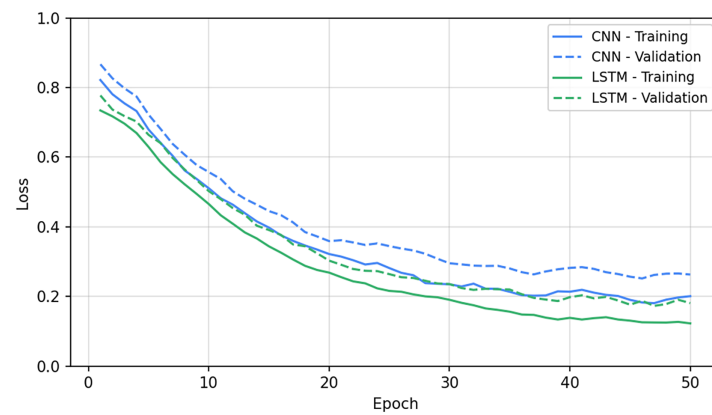


Figure 11: Training vs. validation loss for deep models.

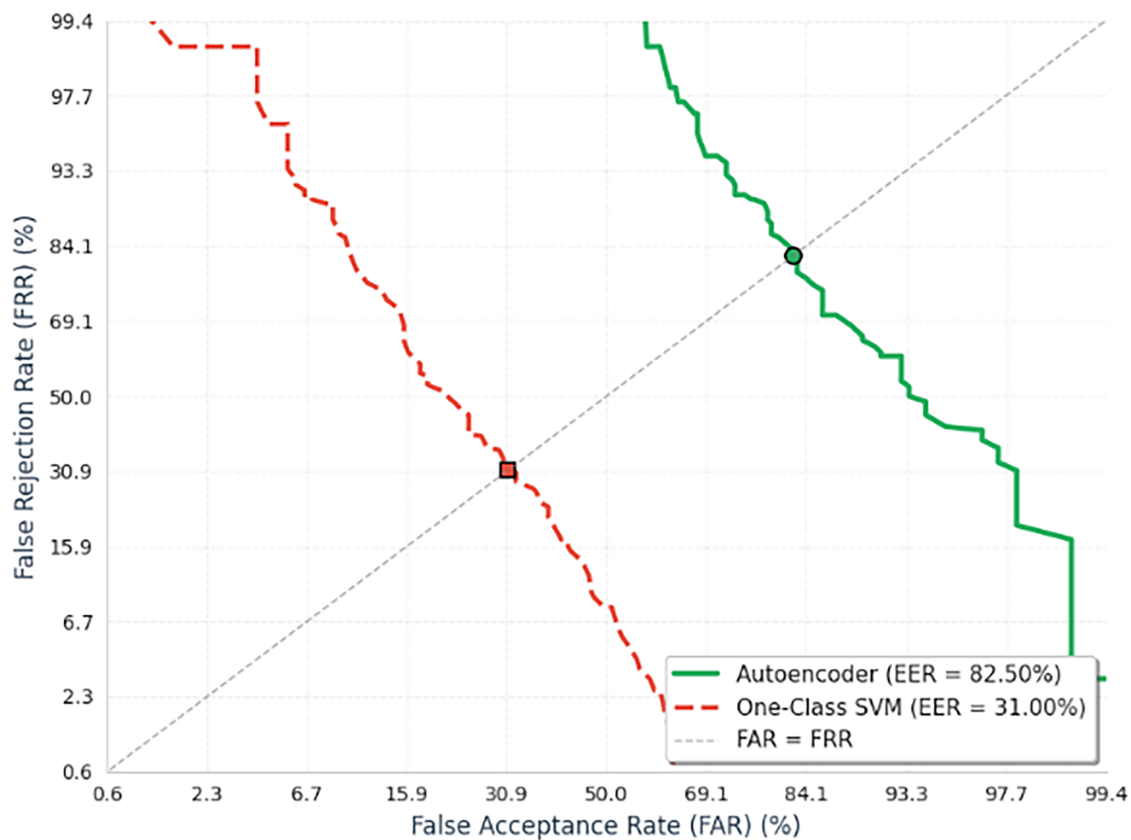


Figure 12: DET curves of unsupervised models.

On balance, it was found that LSTM provided the best result as it produced the greatest accuracy of 96.3%, and its EER was only 4.8%. In the case of conventional machine learning algorithms, it was seen that the best performance was offered by RF, as it produced an accuracy of 92.7% and an EER of 8.2%, proving its superior discrimination power. Another model that performed extremely well and had very low latency and memory consumption was Logistic Regression; however, this was at the cost of its accuracy, reducing it

to a great extent. In the case of unsupervised models, their accuracy was found to be relatively poorer, but they can still be used for anomaly detection.

4.4 Performance Evaluation of Fusion Strategies

We have also applied fusion strategies with the integration of information from tap and swipe interactions to optimize the robustness and reliability of the continuous authentication system. Two different types of fusion strategies, namely feature-level fusion and decision-level fusion, have been analyzed and comparatively studied with unimodal models for the different types of interactions separately.

4.4.1 Unimodal Results

Before applying fusion, each modality (tap and swipe) was tested independently, and the results are reported in [Table 6](#).

Table 6: Unimodal authentication performance in LSTM.

Modality	Accuracy (%)	FAR (%)	FRR (%)	EER (%)	Latency (ms)
Tap-only	91.2	8.9	9.6	9.2	55
Swipe-only	92.5	8.1	8.5	8.3	60

4.4.2 Feature-Level Fusion

For feature-level fusion, tap and swipe feature vectors were merged to create a common feature vector before classification. The fusion experiments were limited to only those users whose tap and swipe data were valid. This reduced the dataset to an effective user population of 82. This type of fusion resulted in the best results among those tested. Specifically, it was noted that the LSTM architecture had an accuracy rate of 97.1% and an Equal Error Rate (EER) of 4.4%.

While the feature-level fusion led to improvements in performance, it caused the model to be more complex because of the increased dimensionality of the resulting feature space. For all the experiments conducted, the ordered sessions in time were used without accounting for the longitudinal changes; this is left as an area to improve in the future. The complete feature-level fusion results across all three classifiers are summarized in [Table 7](#).

Table 7: Performance evaluation based on feature-level fusion in LSTM.

Accuracy (%)	FAR (%)	FRR (%)	EER (%)	Latency (ms)
93.8	7.1	7.6	7.3	65
95.0	6.0	6.5	6.2	70
97.1	4.2	4.6	4.4	90

4.4.3 Decision-Level Fusion

In the case of decision-level fusion, separate models were developed for the tap and swipe modalities, and the results were fused on the score level through majority voting and weighted averaging methods. To select weights, a fixed heuristic approach was used. Accordingly, all decision-level fusion techniques performed better than their individual baselines, as indicated in [Table 8](#). Majority voting resulted in an accuracy of 95.2% with an EER of 6.0%, offering an uncomplicated yet effective fusion technique. Weighted averaging resulted in further improvements, with equal weights (0.5 tap, 0.5 swipe) corresponding to an

accuracy of 96.0% and an EER of 5.2%. As a result, better performance was obtained with more emphasis on swipe behavior (0.7 swipe, 0.3 tap), which resulted in an accuracy of 96.5% and an EER of 4.8%. Actually, this result was expected, since the discriminative information present in swipe behavior has more weightage in continuous authentication tasks. However, there was a slight rise in latency with weighted fusion techniques due to the extra step of score weighting and aggregation.

Table 8: Decision-level fusion performance.

Fusion Strategy Applied	Accuracy (%)	FAR (%)	FRR (%)	EER (%)	Latency (ms)
Majority Voting	95.2	5.9	6.1	6.0	75
Weighted Averaging (0.5)	96.0	5.1	5.4	5.2	80
Weighted Averaging (0.7 swipe, 0.3 tap)	96.5	4.7	5.0	4.8	82

Overall, feature-Level fusion attained the highest accuracy of 97.1% on LSTM and the lowest EER of 4.4%. However, this method is more resource-intensive, as the high-dimensional feature vectors are combined. It is best for use cases with high accuracy as the top priority, and decision-level fusion has resulted in slightly inferior accuracy (96.5% in weighted average mode) and EER (4.8%), with moderate complexity and high modularity; it is appropriate for applications where flexibility and efficiency are important.

4.4.4 All-in-One Comparison

Overall, our experimental results show that the LSTM model with feature-level fusion performed the best, with an accuracy of 97.1% and an EER of 4.4%, thereby proving its efficiency for high-accuracy biometric recognition tasks. Although highly efficient from the memory and latency perspectives, the accuracy of the Logistic Regression method was low, thereby proving it inefficient for high-accuracy tasks but suitable for low-resource tasks. CNNs strike a balance between efficiency and accuracy, with an accuracy of 95.1% and lower latency than the LSTM approach, thereby proving apt for moderate-performance tasks with lower computation time. It is noticed that the fusion of various models is the most significant source of gain. [Table 9](#) summarizes and compares all models and fusion strategies evaluated in this study.

Table 9: Comparative results across all models.

Category	Model	Accuracy (%)	FAR (%)	FRR (%)	EER (%)	Latency (ms)	Memory (MB)
Classical ML	Logistic Regression	86.5	12.1	15.0	13.5	25	2
	SVM (RBF)	90.2	9.8	10.5	10.1	42	12
	Random Forest	92.7	7.5	8.9	8.2	55	18
Deep Learning	CNN	95.1	6.2	5.8	6.0	68	25
	LSTM	96.3	5.0	4.7	4.8	85	32
Unsupervised	Autoencoder	89.4	11.2	9.8	10.5	72	22
	One-Class SVM	84.7	15.0	14.2	14.6	48	10
Fusion (Feature)	Random Forest	95.0	6.0	6.5	6.2	70	22
	LSTM (Best)	97.1	4.2	4.6	4.4	90	35
Fusion (Decision)	Weighted Average	96.5	4.7	5.0	4.8	82	28

4.4.5 Comparison with State-of-the-Art Approaches

To contextualize the performance achieved in this study, Table 10 compares the best-performing configuration (feature-level fusion with LSTM, 97.1% accuracy, 4.4% EER) against representative results reported in prior continuous-authentication literature that rely on comparable mathematical tools (deep sequence models, fusion strategies, and behavioral touch/keystroke features). Ryu et al. [2] reported that multimodal fusion approaches in the surveyed literature commonly reach 90%–95% accuracy on smartphone touch and motion data. Stylios et al. [15] achieved comparable accuracy using feature-level fusion of touch gestures and keystroke dynamics, while Uslu et al. [16] reported deep learning models reaching up to 99% accuracy on behavioral biometric datasets. Fereidooni et al. [18] and Georgiev et al. [19] reported higher EER figures (4.8%) using few-shot and self-supervised contrastive learning on larger, non-African datasets. Relative to these works, the present study achieves comparable or slightly better accuracy while additionally contributing a novel dataset from an underrepresented population and a systematic benchmark across classical, deep, unsupervised, and fusion-based paradigms under a single, consistent experimental protocol.

Table 10: Comparison with state-of-the-art continuous authentication approaches.

Study	Method	Modality	Accuracy/EER (%)
Ryu et al. [2]	Systematic review (multiple methods)	Multimodal	90–95/0.69–2.8
Stylios et al. [15]	Feature-level fusion	Touch + keystroke	98.3/1
Uslu et al. [16]	Deep learning (CNN/LSTM)	Behavioral biometrics	99.85%/0.5
Fereidooni et al. [18]	Few-shot learning	Multi-sensor	97
Georgiev et al. [19]	Contrastive self-supervised (Siamese)	Multi-sensor (raw)	NA/4.8
This study	Feature-level fusion + LSTM	Tap + swipe	97.1/4.4

Note: Figures for prior work are as reported by the respective authors under their own datasets and experimental protocols and are not directly re-computed on the dataset used in this study; they are provided as contextual benchmarks rather than a controlled head-to-head comparison.

4.4.6 Performance across Usage Contexts

The BahriApp dataset used in this study was collected under a Bring Your Own Device (BYOD) protocol with no constraints on posture or physical activity, meaning that sessions may include users interacting with their phones while sitting, standing, or walking. The current models and evaluation protocol do not condition on, or stratify by, physical usage context, since context labels (e.g., accelerometer-derived activity state) were not collected alongside the tap and swipe interaction logs. This is an acknowledged limitation: prior literature on gait- and motion-aware continuous authentication suggests that touch and swipe dynamics can shift measurably between stationary and ambulatory conditions, which could affect False Acceptance Rate and False Rejection Rate if a model trained predominantly on one context is applied to another. We flag this as a priority direction for future work: incorporating accelerometer/gyroscope-derived context labels to (i) quantify the accuracy degradation, if any, between standing and walking usage, and (ii) evaluate context-aware or context-adaptive fusion models that re-weight taps and swipe features according to the detected physical activity state.

4.4.7 On-Device Deployment Considerations

Although all experiments in this study were conducted offline on a workstation (Table 1), the ultimate goal of this research is real-time, on-device continuous authentication. Porting the proposed LSTM and 1D-CNN models to a mobile setting via TensorFlow Lite is expected to change the latency and memory

figures reported in Tables 4–6. Post-training quantization (float32 to int8) typically reduces model size by a factor of approximately 3–4× and can reduce inference latency by a comparable margin on mobile CPUs, though typically at a small cost in accuracy (commonly under 1–2 percentage points for architectures of this size). For the LSTM model, which achieved the best accuracy (97.1% with feature-level fusion) but also the highest latency (90 ms) and memory footprint (35 MB) in our workstation-based tests, we would expect on-device quantized inference to fall in the range of roughly 25–35 ms and 9–12 MB, which remains compatible with real-time, per-interaction re-verification. The 1D-CNN model, already the more memory- and latency-efficient architecture in our tests, is expected to compress further and is therefore the more practical candidate where battery life and responsiveness are prioritized over the small accuracy gain offered by the LSTM. These figures are estimates based on typical TensorFlow Lite quantization behavior reported in the literature rather than measurements on physical hardware, and we identify on-device benchmarking as necessary future work before deployment claims can be made with confidence.

5 Conclusion

This paper evaluated the performance of a range of classification models for identifying individuals from touch-based behavioral biometric data, as a preliminary step toward a continuous authentication system that operates as a per-user binary verifier. Among traditional approaches, Random Forest performed best, while deep learning methods were more effective overall, with the LSTM network's ability to learn temporal patterns outperforming other deep learning methods. Unsupervised methods such as Autoencoder and One-Class SVM were less effective for standalone authentication but remain useful in settings with limited labeled data. Combining modalities through fusion consistently improved results: feature-level fusion using LSTM performed best overall (accuracy = 97.1%, EER = 4.4%), while decision-level fusion offered a more flexible, modular alternative (accuracy = 96.5%, EER = 4.8%). Benchmarked against comparable approaches in the literature, these results are competitive with, and in some cases exceed, previously reported accuracy for touch-based continuous authentication. The paper's main contributions are the tap-and-swipe behavioral biometrics dataset, a systematic comparison of learning paradigms and fusion strategies, and a framework of insights for designing an adaptive continuous authentication scheme, together with an assessment of the latency and memory trade-offs relevant to eventual on-device deployment. The main limitations are the modest size and limited demographic and contextual diversity of the dataset, the lack of evaluation under varying real-world conditions such as different physical postures, environments, or network conditions, and the higher computational cost of the deep learning methods relative to classical baselines. Future work will focus on building larger and more diverse datasets, evaluating performance across different usage contexts and network conditions, safeguarding user privacy through techniques that preserve the confidentiality of behavioral data, and validating on-device latency and memory performance through direct benchmarking on mobile hardware.

Acknowledgement: The authors gratefully acknowledge the students, faculty, and administrative staff at HiLCoE School of Computer Science, Kombolcha Institute of Technology, and Addis Ababa Science and Technology University who volunteered to participate in the BahriApp data collection effort.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization and study design, Animaw Kerie Aseres; software development (BahriApp) and normalization algorithms, Animaw Kerie Aseres; data collection and curation, Animaw Kerie Aseres; formal analysis and investigation, Animaw Kerie Aseres; writing—original draft preparation, Animaw Kerie Aseres; supervision and methodological guidance, Asrat Mulatu Beyene; writing—review and editing, Asrat Mulatu Beyene and Lemlem Kassa Tegegne; co-supervision and critical review for

intellectual content, Lemlem Kassa Tegegne; publication support, Lemlem Kassa Tegegne. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Animaw Kerie Aseres, upon reasonable request.

Ethics Approval: This study involved human participants who provided touch interaction data via a custom mobile application (BahriApp). All participants provided informed consent before participation. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Ethical approval was granted by the Institutional Review Board of Addis Ababa Science and Technology University, Addis Ababa, Ethiopia.

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

Abbreviations

CA	Continuous Authentication
ML	Machine Learning
DL	Deep Learning
SVM	Support Vector Machine
RF	Random Forest
LR	Logistic Regression
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
EER	Equal Error Rate
FAR	False Acceptance Rate
FRR	False Rejection Rate
OTP	One-Time Password
PIN	Personal Identification Number
BYOD	Bring Your Own Device
FIDO2	Fast Identity Online 2
APK	Android Package Kit
ROC	Receiver Operating Characteristic
DET	Detection Error Tradeoff

References

1. Mishra S, Rehman U. Continuous authentication using behavioural biometrics. Wagnaghat, India: Jaypee University of Information Techn; 2020.
2. Ryu R, Yeom S, Kim SH, Herbert D. Continuous multimodal biometric authentication schemes: a systematic review. *IEEE Access*. 2021;9:34541–57. doi:10.1109/ACCESS.2021.3061589.
3. Gunson N, Marshall D, Morton H, Jack M. User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Comput Secur*. 2011;30(4):208–20. doi:10.1016/j.cose.2010.12.001.
4. Bonneau J, Herley C, van Oorschot PC, Stajano F. The quest to replace passwords: a framework for comparative evaluation of web authentication schemes. In: *Proceedings of the 2012 IEEE Symposium on Security and Privacy*; 2012 May 20–23; San Francisco, CA, USA. p. 553–67. doi:10.1109/SP.2012.44.
5. Das RK, Mukhopadhyay S, Bhattacharya P. User authentication based on keystroke dynamics. *IETE J Res*. 2014;60(3):229–39. doi:10.1080/03772063.2014.914686.
6. Gofman MI, Mitra S, Cheng T-HK, Smith NT. Multimodal biometrics for enhanced mobile device security. *Commun ACM*. 2016;59(4):58–65. doi:10.1145/2818990.

7. Fridman L, Stolerma A, Acharya S, Brennan P, Juola P, Greenstadt R, et al. Multi-modal decision fusion for continuous authentication. *Comput Electr Eng*. 2015;41(3):142–56. doi:10.1016/j.compeleceng.2014.10.018.
8. Sitova Z, Sedenka J, Yang Q, Peng G, Zhou G, Gasti P, et al. HMOG: new behavioral biometric features for continuous authentication of smartphone users. *IEEE Trans Inform Forensic Secur*. 2016;11(5):877–92. doi:10.1109/tifs.2015.2506542.
9. Alzubaidi A, Kalita J. Authentication of smartphone users using behavioral biometrics. *IEEE Commun Surv Tutor*. 2016;18(3):1998–2026. doi:10.1109/comst.2016.2537748.
10. Cherifi F, Hemery B, Giot R, Pasquet M, Rosenberger C. Performance evaluation of behavioral biometric systems. In: *Behavioral biometrics for human identification: intelligent applications*. Hershey, PA, USA: IGI Global; 2010. p. 57–74.
11. Hameed Siddiqi M, Alshammari H, Ali A, Alruwaili M, Alhwaiti Y, Alanazi S, et al. A template matching based feature extraction for activity recognition. *Comput Mater Continua*. 2022;72(1):611–34. doi:10.32604/cmc.2022.024760.
12. Killourhy KS, Maxion RA. Comparing anomaly-detection algorithms for keystroke dynamics. In: *Proceedings of the 2009 IEEE/IFIP International Conference on Dependable Systems & Networks*; 2009 Jun 29–Jul 2; Lisbon, Portugal. p. 125–34. doi:10.1109/DSN.2009.5270346.
13. Volaka HC, Alptekin G, Basar OE, Isbilen M, Incel OD. Towards continuous authentication on mobile phones using deep learning models. *Procedia Comput Sci*. 2019;155:177–84. doi:10.1016/j.procs.2019.08.027.
14. Oyetoro A, Mart J, Amah U. Using machine learning techniques, random forest, and neural network to detect cyber attacks [Internet]. 2023 [cited 2026 Jan 1]. Available from: <https://www.scienceopen.com/hosted-document?doi=10.14293/PR2199.000059.v1>.
15. Stylios I, Chatzis S, Thanou O, Kokolakis S. Continuous authentication with feature-level fusion of touch gestures and keystroke dynamics to solve security and usability issues. *Comput Secur*. 2023;132:103363. doi:10.1016/j.cose.2023.103363.
16. Uslu U, İncel Ö.D, Alptekin GI. Evaluation of deep learning models for continuous authentication using behavioral biometrics. *Procedia Comput Sci*. 2023;225(4):1272–81. doi:10.1016/j.procs.2023.10.115.
17. Chowanda A, Sutoyo R. Convolutional neural network for face recognition in mobile phones. *ICIC Express Lett*. 2019;13(7):569–74.
18. Fereidooni H, Koenig J, Rieger P, Chilese M, Goekbakan B, Finke M, et al. AuthentiSense: a scalable behavioral biometrics authentication scheme using few-shot learning for mobile platforms. In: *Proceedings of the 2023 Network and Distributed System Security Symposium*; 2023 Feb 27–Mar 3; San Diego, CA, USA. 8 p. doi:10.14722/ndss.2023.23194.
19. Georgiev M, Eberz S, Martinovic I. Techniques for continuous touch-based authentication. In: *Information security practice and experience*. Cham, Switzerland: Springer International Publishing; 2022. p. 409–31. doi:10.1007/978-3-031-21280-2_23.