JCSDA, Vol. 02, No. 01, 83–118
DOI: 10.69660/jcsda.02012505
ISSN 2959-6912

BehFayda: A Comprehensive Review and Framework Proposal for Adaptive Authentication in National Identity Systems Using Multi-Modal Biometric Fusion

Animaw Kerie Aseres*, Asrat Mulatu Beyene, and Lemlem Kassa Tegegne
College of Engineering,
HPC and Big Data Analytics Center of Excellence,
Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
**corresponding author: animaw.kerie@gmail.com*

The proliferation of digital services necessitates robust identity verification mechanisms. The Ethiopian digital national ID, Fayda, built on the Modular Open-Source Identity Platform (MOSIP), aims to offer a secure and scalable solution for national identity management. However, MOSIP lacks explicit support for adaptive continuous authentication—a crucial aspect of ensuring security and usability. This paper introduces BehFayda, a comprehensive architecture for a privacy-enhanced multi-modal biometric fusion system for adaptive continuous authentication tailored to digital identity systems. The framework integrates behavioral biometrics, such as keystroke dynamics in two languages, swipe gestures, motion data, and contextual data as a candidate for the proposed fusion strategy. We propose the Multi-Modal Deep Residual Fusion (MM-DRF) algorithm, which incorporates feature-level fusion with adaptive attention mechanisms to dynamically adjust the contribution of different biometric modalities based on their relevance. Our approach provides a new insight to enhance authentication accuracy which mainly aims to guide future research in advancing adaptive authentication in national digital identity systems, with a focus on privacy-preserving techniques and real-time behavioral analysis.

Keywords: Fayda identification number, adaptive authentication, continuous authentication, multi-modal biometrics, user impersonation, privacy preservation, Modular Open-Source Identity Platform.

1. Introduction

The initial interaction with mobile devices and applications often begins with the authentication process. Authentication is the process of verifying that the user accessing the system is indeed legitimate, ensuring their identity is confirmed. Conversely, identification involves determining the specific user accessing the system without necessarily verifying their legitimacy. Smartphones are frequently utilized for storing and accessing confidential and sensitive data. In scenarios where a user possesses multiple financial accounts, they may find themselves managing a multitude of login credentials. While single-sign-on (SSO) technologies can streamline access to less sensitive cloud services, they fall short when it comes to highly critical services such as banking. Recognizing this challenge, the Ethiopian government launched the Ethiopian National ID Program[a] as part of the Digital Ethiopia 2025

[a] https://id.et/

Strategy, marking a significant milestone in 2022. The National Digital ID, dubbed "Fayda," was established under Proc. 1284/2023 and serves as Ethiopia's foundational legal identification. The term "Fayda" not only denotes a unique national identity but also conveys vitality, a concept emphasized by the Prime Minister of Ethiopia and consistent across various languages[b]. Throughout this paper, the term "Fayda" refers to a Digital Public Good (DPG) for identity management. Digital identity systems play a pivotal role in facilitating access to essential services, financial transactions, and government programs. MOSIP[c], an open-source initiative, seeks to address identity challenges by providing a modular and adaptable platform. Fayda is constructed upon MOSIP and currently integrates three physiological biometrics: fingerprints, iris scans, and facial recognition.

Using the traditional physiological biometric feature for authentication is always challenging. For example, for a facial, the user might adjust their head position in a way that obstructs the camera's view of their face. Alternatively, the user might enter an area with poor lighting conditions, presenting a challenge for face recognition.[1]. During the usage of software applications, particularly FinTech apps, users frequently input various texts such as credentials, account numbers, and transfer amounts. Additionally, they engage in tapping, pressing buttons, and scrolling activities for larger text and windows. In addition to static biometrics, these activities can be used as constructs of behavioral biometrics (BB). BB is getting high attention across the developed nations. For example, the US Government has launched a pilot project using behavioral biometrics for the authentication systems of the Internal Revenue Service [2]. Behavioral biometrics, like typing rhythm, screen touch gestures, and mouse click patterns offer convenience, leveraging everyday devices for authentication, and enabling continuous monitoring against unauthorized access. Alongside familiar biometrics, we advocate for behavior as a distinct identity factor in this research.

Identifying a user with unimodal biometrics is challenging. Therefore, the current state-of-the-art research involves the fusion of multiple sources of biometric data to identify a subject. Our research mission is to construct a tailored dataset derived from built-in smartphone sensors while users perform different activities, aligning them with the Amharic textual content they input, with the ultimate goal of developing a user behavioral model using deep learning algorithms. Our research also aims to investigate the impact of typing language, specifically English versus Amharic, on authentication methods. Additionally, we seek to explore low-cost authentication techniques that do not necessitate the installation of extra hardware, focusing primarily on touch gestures and keystroke data. This is crucial because acquiring standard hardware security module-enabled devices for authentication among citizens in developing nations may pose challenges as the Fayda system becomes more integrated across digital services. In essence, our goal is to improve

[b] https://www.youtube.com/watch?v=1ZdSMbcXbRY. [Accessed January].
[c] https://mosip.io

the Ethiopian digital national ID system by integrating behavior-based identity verification alongside conventional biometrics.

1.1. *Background and motivation*

Although the exact number of smartphone users in Ethiopia is not known, one can predict that 6.40 million social media users as of January 2023, equating to 5.1 percent of the total population are expected to own smartphones [3]. And today smartphone subscriptions have exceeded 6 billion worldwide [4]. In addition, Mobile and Internet Banking users also rapidly growing. For example, CBE reached more than 6.6 million mobile banking users[d]. From this population, there is no distinct study dictating how many of them employ any form of authentication to safeguard their smartphones, however, in the other world, it is reported that 30% of the sample do not use any security method on their smartphone, although they have a strong belief in their device's protection [5]. From simple observation, most users use pattern-based and password-based authentication. Yet, these static authentication methods are vulnerable to various attacks such as brute force attacks and spoofing. Moreover, these security measures only safeguard the device during login or unlocking. If an unauthorized individual gains physical access to an unlocked device, they could potentially access all the stored data. Consequently, it's crucial to address this security threat.

In the realm of authentication security, the defense-in-depth strategy entails implementing multiple layers to safeguard user access. The conceptualized defense-in-depth authentication is designed here having three layers. These layers collaborate to bolster security by offering redundancy and resilience. Initially, strong and complex passwords are pivotal. Users ought to create robust passwords combining various character types. Secondly, one-time multi-factor authentication (MFA) adds a layer. It encompasses something the user knows (like a password), something they have (such as a one-time code), and something they are (biometric factors). Identifying users through lengthy passcodes and passwords can mitigate the usability issue. An eager user needing to urgently pay or transfer money shouldn't waste time entering passwords and re-authenticating several times to conduct the transaction. Moreover, user interruptions may lead to unconscious repeated attempts and account lockouts. Therefore, due to the usability problems in multifactor authentication, it's necessary to verify user identity on smartphones without active user involvement to either replace or complement the existing login process, thereby enhancing user experience. Therefore, the first two methods have a drawback: if an unknown user successfully forges authentication, the system will continue to operate without any resistance, posing a significant risk to the entire system. The third layer is the continuous authentication layer which facilitates seamless real-time identity verification for authorized users, ensuring smooth device access while

[d] https://factcheck.afp.com/doc.afp.com.34H96ZW

safeguarding against unauthorized attempts. Using behavioral biometrics presents a promising approach to smartphone security. Continuous authentication based on behavioral biometrics presents several advantages over typical one-time authentication methods. Firstly, it eliminates the need for additional hardware support, as it leverages existing sensors in devices to capture biometric data representing user behavioral patterns. Secondly, it doesn't require root access privileges for sensor data acquisition, ensuring security and accessibility. Thirdly, users are not actively involved in the authentication process, as their behavioral patterns are continuously monitored passively in the background. Lastly, it offers session-long identity authentication, verifying the user's identity persistently throughout the session without the need for repeated logins, thereby enhancing overall system security and user convenience. Hence, continuous authentication is an implicit authentication technique that monitors user behavior, contextual factors, and risk levels. While still in its developmental stages, limited industries are providing commercial products integrating behavioral continuous authentication for smartphones, as referenced in [6]. The defense-in-depth authentication discussed above is depicted diagrammatically in Figure 1.

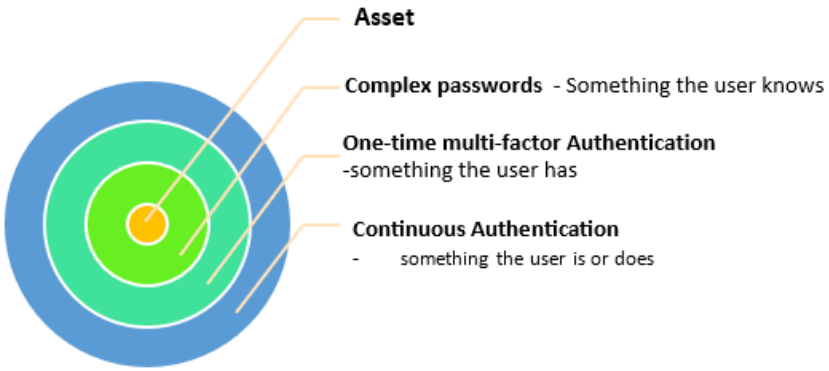

Fig. 1. Defense-in-depth authentication

In addition, our motivation is to explore whether a user should be identified as an imposter or an attacker when switching between English and Amharic typing languages during authentication. This paper remains at a framework-level proposal, with experimentation left for future work.

1.2. *Problem statement*

The utilization of physiological biometric authentication presents challenges for certain demographics, particularly Ethiopian farmers who engage in manual labor,

resulting in fingerprints that may vary over time. This variation decreases recognition accuracy due to environmental factors and the aging of samples, leading to intra-class variability. Additionally, prevalent mobile applications for financial transactions, like TeleBirr, CBE mobile banking, Amole, AwashBIRR, etc. rely on re-authentication methods. While this approach bolsters security, it compromises usability. To address this, smartphones should implement implicit and continuous re-authentication through behavioral biometrics.

Moreover, authentication device to use Fayda Digital Public Good (DPG) requires costly authenticator devices. These devices require robust security features, such as Foundational Trust Modules (FTMs), to ensure the protection of cryptographic keys and secure application execution. However, the expense and limited availability of such devices may hinder widespread adoption, particularly in resource-constrained areas. To address these issues, we propose leveraging behavioral biometrics as an alternative authentication method. BB utilizes unique patterns in user behavior, such as typing rhythms, touchscreen gestures, and voice characteristics, for identity verification. Unlike hardware-based solutions, behavioral biometrics can be implemented on existing mobile devices without the need for specialized hardware, making them more accessible and cost-effective. Integrating behavioral biometrics into authentication systems not only improves accessibility but also enhances security by adding a layer of verification based on individual behavioral traits. Furthermore, our research aims to explore privacy-preserving techniques for implicit authentication during this process. This research will seek to answer the following questions:

- RQ1: What are the prevalent gaps and promising opportunities in the existing literature concerning continuous authentication?
- RQ2: What are the most effective methods and optimal parameters for collecting datasets to comprehensively capture user behaviors essential for continuous authentication purposes?
- RQ3: How does the BehFayda system architecture significantly contribute to the overall security stance of Fayda's National Digital Identity System, ensuring robust authentication mechanisms?
- RQ4: How does the MM-DRF algorithm improve multi-modal fusion for dynamic authentication through feature-level fusion and adaptive attention mechanisms?

1.3. *Objective of the study*

The specific objectives of this study are to:

- Research the most recent studies in behavioral biometrics for adaptive authentication and identify research gaps.
- Conceptualize the core topics of the research domain by synthesizing insights from the literature review.

- Investigate the effectiveness of different combinations of keystroke dynamics, textual features, and swipe gestures for robust authentication through critical review.
- To evaluate various fusion methods to determine the most effective approach for combining behavioral features and propose a mathematical model for fusion strategy.
- Establishing a pioneering conceptual framework for Multi-Modal Biometric Fusion, advancing understanding within the field.
- Develop an integrated BehFayda architecture to enable continuous authentication in Fayda's Digital Public Good.

The rest of this paper is organized as follows. Related research is reviewed and critically summarized in Section 2. In Section 3, we present the design of the conceptual framework. Section 4 details our proposed system architecture, while Section 5 concludes our work and recommends potential future enhancements and research directions.

2. Literature review

2.1. *Adaptation techniques*

Adaptation techniques within the realm of biometric authentication can be broadly categorized into two types: structural and parametric adaptation [7]. Structural adaptation involves altering the system's structure or components, which may entail adding or removing modules based on changing conditions. On the other hand, parametric adaptation involves adjusting system parameters without altering the overall architecture. This could include modifying templates, selecting different features for sampling, or choosing algorithms for multi-modal authentication.

Updating templates associated with biometric features emerges as a prevalent adaptation method in behavioral biometrics. For instance, patterns in keystroke dynamics may evolve as users become accustomed to repetitive password entry. Therefore, updating templates becomes crucial to ensuring accurate identification across diverse scenarios.

MOSIP's modular design allows customization, but it does not explicitly address adaptive methods. Adaptive continuous authentication adapts security thresholds based on real-time context, user behavior, and risk factors. Our motivation lies in bridging this gap within MOSIP while ensuring robust privacy protection.

2.2. *Continuous authentication*

The motivation for Continuous authentication is "Your device may know you better than You know yourself" [8]. Continuous authentication for smartphones is essential due to the shortcomings of traditional methods like PINs, passwords, fingerprints, and face recognition in thwarting physical access by adversaries. According

to Heather et al. [5] , 90% of participants favor behavioral biometrics-based authentication over physiological-based authentication, signifying a rising interest in this field. Behavioral signals such as touchscreen interactions, gait patterns, eye movements, and hand gestures are employed for continuous authentication without interrupting user tasks. However, current approaches suffer from fixed authentication intervals, disregarding external cues and leading to security vulnerabilities and usability challenges. In [6] , SMARTCOPE is introduced to address these issues by utilizing smartphone movement signals to detect instances when the device is no longer in the owner's possession. Improved authentication outcomes were observed in a separate study that utilized keystroke dynamics with the sensors within smartphones known as the Inertial Measurement Unit (IMU), including accelerometers, gyroscopes, and magnetometers. These sensors were employed to collect data about users' behavioral tendencies, such as their manner of holding their phones [9]. In the paper [10], a two-stage framework process was proposed, comprising static authentication at login using vein biometrics and continuous authentication using keystroke dynamics throughout the user session. However, in both stages, the proposed methodology includes data collection, feature extraction, classification phases, and evaluation. While the approach demonstrates an effort to implement multimodal biometrics, it falls short in two key aspects: it does not integrate different levels of fusion to potentially enhance results, and the framework itself lacks comprehensive evaluation. Another work [11] , proposed a two-factor authentication for multi-site large enterprises using mouse click-based behavioral authentication. The limitation of this work is that, it is dedicated for only employees in geographically constrained branch locations without considering the public group. Moreover, they have employed mouse-based desktop computers but not smartphones.

2.2.1. *Touch-based continuous authentication*

Biometric touch data is collected from user interactions with the device. It includes both Swipe Dynamics (SD) (with attributes of speed, direction, length, and acceleration) and Touch Dynamics (with attributes of pressure, duration, multi-touch, accuracy, etc.). There are two steps:

(1) Compare Against Known Behavior: The captured touch data is compared to a pre-established biometric user template, which represents expected touch behavior.
(2) Evaluate and Decide: Based on the comparison results, a decision is made:
    - If the touch behavior closely matches the biometric user template, access to the device is granted (Step 3.1).
    - If there's a significant deviation from the expected behavior, the device is locked (Step 3.2).

This process forms a continuous loop, ensuring the device's security over time by continuously capturing, comparing, and evaluating touch data for access control.

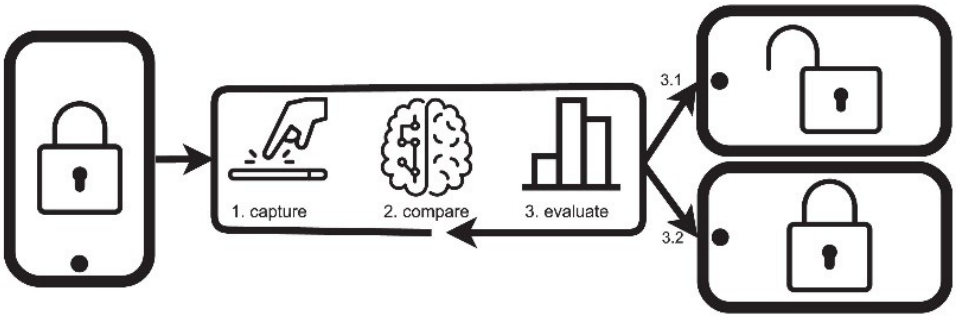

Fig. 2. Touch-based continuous authentication concept [12]

One of the earliest studies [13], conducted around the introduction of touch smartphones, created a proof-of-concept classification framework. It extracted 30 behavioral features from raw touchscreen data and utilized -nearest neighbor and Gaussian RBF kernel SVM classifiers for training. Authentication results demonstrated equal error rates between 0% and 4% across various scenarios, highlighting strong performance for continuous authentication based on natural navigation gestures. Smith-Creasey et al. [14] , introduced a continuous authentication scheme for mobile devices that dynamically adjusts touchscreen interaction thresholds based on trust derived from passive sensor data. They have employed probabilistic methods. Another closely related study to our proposal [15] Introduced enhancements to Biotouch, a supervised machine-learning framework designed for continuous user authentication. However, their focus was solely on touch dynamics without employing any fusion techniques. Additionally, their research does not pertain to national identity verification frameworks.

2.2.2. *Keystroke-based continuous authentication*

Our interest in this modality is because keystroke dynamics can be used for both identity and verification with its inherent challenges. The keystroke dynamics feature extraction process involves organizing interactions into KeyDown and KeyUp events. [16]. These events are paired to form di-graphs using a sliding window of size 2. Six features are typically extracted from these di-graphs:

- H1: Time elapsed for the first interaction.
- H2: Time elapsed for the second interaction.
- RP: Time elapsed between the release of the first interaction and initiation of the second.
- PP: Time elapsed between the initiation of the two interactions.
- RR: Time elapsed between the release of the two interactions.
- PR: Time elapsed between initiation of the first interaction and release of the second.

This methodology is particularly pertinent for typing most Amharic letters as it employs digraphs. Because, among many other features set, key down–key-down latency, which is the time taken for a user to press two consecutive keys is the most widely used feature [17].

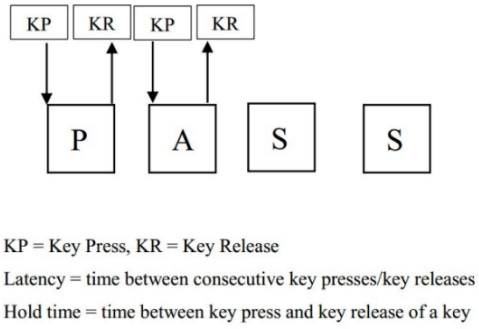

Fig. 3. Timing Features of Keystrokes[18]

Paper [19] , addresses a noted gap in existing research by proposing contents and keystroke dual attention networks, which integrate pre-trained models for continuous authentication. Unlike previous studies predominantly centered on keystroke dynamics alone, this approach acknowledges the textual content entered alongside keystrokes. In our research, we employ both English and Amharic scripts for authentication credentials. We consider both the content of the text and the keystroke dynamics to model users' behavior towards system usage. Our methodology involves leveraging a pre-trained language model adopted from [20], to facilitate the continuous authentication process. There is very few amounts of work in continuous authentication which combines both physiological and behavioral biometrics. Maria et al. [10] , proposed the use of a free-password authentication scheme using vein recognition at the login stage and keystroke dynamics as a continuous authentication during the user session.

Using stylometric features it is also possible to develop an authorship verification model applicable for continuous authentication so that it is possible to identify anonymous authors on a specific topic [21]. So it will be helpful to identify long text hate speeches for instance. This is another interesting area of study for future research. Moreover, an experiment in keystroke research is conducted to evaluate the performance when touchscreen and physical keyboards are used [16].

2.3. *Machine learning for continuous authentication*

The latest trend in behavioral biometrics-based mobile CA systems includes the adoption of deep learning approaches such as MLPs, RNNs, LSTMs, and possibly transformer-based architectures, aiming to improve authentication accuracy

and user experience while reducing the manual effort required in feature extraction. Among these: The results of the experiments [22] reveal that MLP and the convolutional-LSTM algorithms achieve the best performance on raw data from both motion sensors and touch screens. In [23], a framework that leverages dynamic selection of classifiers (DS) is proposed. Rather than employing a uniform classifier for all touch strokes, their proposed approach categorizes each touch sample using the most effective classifier(s) selected from a pool of classifiers. Paper [24] assessed the feature extraction capability of a proposed convolutional transformer for continuous authentication. Various advanced deep learning models were evaluated, including ResNet, ResNeXt, MobileNetV2, MobileNetV3, ShuffleNetV2, MnasNet, EfficientNet, and RegNet. Results demonstrated the effectiveness of the proposed convolutional transformer. Authors in [24] , pinpointed two limitations of behavioral continuous authentication: 1) weak capability of capturing smartphone user's behavioral patterns from biometric data sequences with time correlation because these methods don't consider the long-range dependencies between the behavioral biometric data sequences; 2) poor performance under a high authentication frequency. To solve these issues, they have presented AuthConFormer, a novel continuous authentication system based on a proposed convolutional transformer. Another compelling study was brought to our attention. [25] , employed typing and screen sliding behavior patterns in conjunction with GPS location for authentication purposes. They assessed the framework by applying over six different machine learning algorithms, with the primary findings indicating an average accuracy ranging from 78% to 91%. This suggests that there is still room for further research in this domain. Instead of discussing each paper's findings, from recent literature, we have distilled five crucial pieces of information from papers for a comprehensive overview. The utilized biometric and behavioral modalities, any fusion techniques employed (if applicable), the tools utilized for data collection, the machine learning algorithms experimented with, and the evaluation metrics employed.

2.4. *Fusion Techniques*

The integration of behavioral data can occur at two distinct levels [16]. Initially, at the decision level, separate ML models are constructed for each source of information. Subsequently, the predictions from these models are aggregated to yield a single prediction at any given moment. Alternatively, at the feature level, the behavioral traits derived from each information source are amalgamated to input a singular ML model. Consequently, the transitions between different information sources by a user can be concurrently accounted for. However, combining information from multiple sources and making it ready for training data is challenging. Authors are cite16 , to combine mouse and keystroke data, proposed temporal information representation based on Symbolic Aggregate approximation (SAX) implemented through Random Trees Embeddings (RTEs). Then they used DNA sequence alignment techniques to compare them accurately. Then, they extracted behavioral

cores using a density-based clustering model to discard outliers' samples.

In [26], authors introduced Context Weighted Majority Algorithm (CWMA), which is an extension of the well-known Weighted Majority Algorithm (WMA) [27]. The WMA combines predictions from multiple experts by assigning weights based on their past performance. Correct predictions increase expert weights, while incorrect predictions decrease them. The final decision is made by weighing expert predictions according to their weights.

Nuttapong et al. [1] Proposed two-dimensional dynamic fusion considering multibiometric continuous authentication calculates two-dimensional matching scores over classifiers and over time. Based on this, they dynamically select a set of classifiers based on the context in which authentication is taking place, and fuse matching scores by multi-classifier fusion and multi-sample fusion.

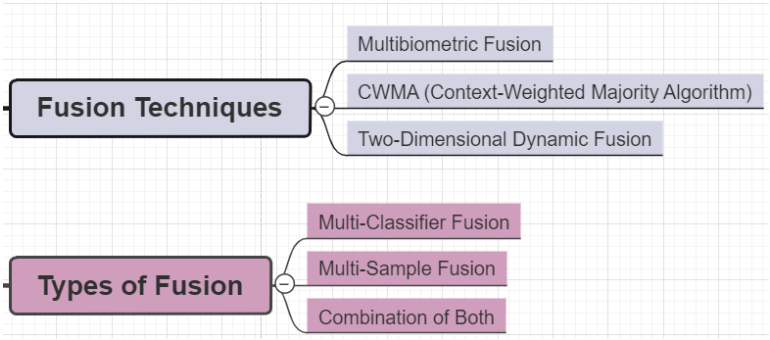

Fig. 4. Comprehensive overview of state-of-the-art multi-modal fusion techniques and fusion types

2.5. *Summary of the previous Research*

A recent survey on keystroke dynamics, referenced in [18], offers a comprehensive comparison of various keystroke dynamics surveys across multiple parameters. These parameters encompass a range of critical aspects, including the scope of covered areas, publication year, citation count, dataset availability, utilization of algorithms, compatibility with mobile devices, employed data processing techniques, and potential applications. Unlike specifying particular attributes, the comparison primarily employs binary indicators (yes/no) to denote the presence or absence of certain features within each surveyed study. In our critical review, we scrutinize algorithmic specifications, behavioral dynamics, data collection tools, fusion techniques, and model evaluation metrics. In the first paper summary findings, we aim to identify the modalities that exhibit a high frequency of usage across all research papers, then we seek to discern which modalities are predominantly utilized in the context of multimodal behavioral dynamics.

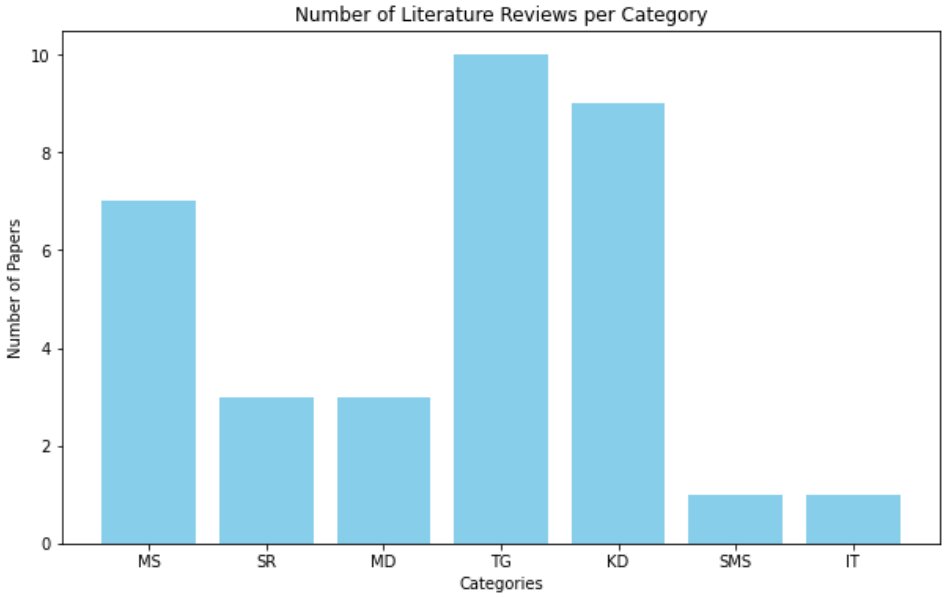

Fig. 5. Distinguishing behavioral biometric modalities

Legend: MS= Motion sensors (they are also called Inertial Measurement Unit (IMU) sensors- accelerometer, gyroscope, magnetometer are common) MD: Mouse dynamics TG=Touch gestures KD=Keystroke dynamics SMS=Smartphone movement signals IT= Input data; ST=stylometry

Among the behavioral biometric modalities analyzed, touch gesture emerges as the most frequently utilized, followed by keystroke dynamics and motion sensors. Sensor readings encompass a wide range of modalities, including data from the accelerometer, Bluetooth, GPS, gravity, gyroscope, light, magnetometer, noise, cellular tower, proximity, Wi-Fi, activity, and pressure sensors. Some authors employed individual sensor readings, while others employed combinations of two or more modalities.

Five out of 24 papers incorporate two modalities, while three papers incorporate three modalities. Notably, there are no papers that integrate four or more modalities concurrently. This observation underscores researchers' inclination toward exploring multiple modalities to augment their findings. Such combinations can yield deeper insights and more thorough outcomes. The absence of papers featuring four or more modalities suggests that achieving extensive combinations might be less prevalent or more challenging.

In the summary of the second paper's findings on the usage patterns of fusion techniques, feature fusion emerges as the predominant technique, followed by score-level fusion. However, the majority of the papers do not utilize any fusion technique

as illustrated in Figure 6.

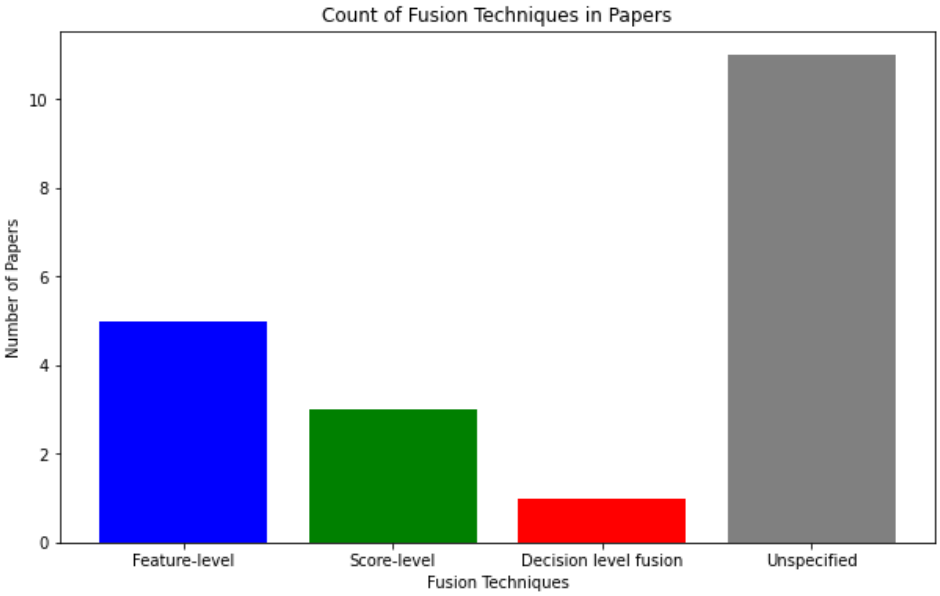

Fig. 6. Adoption of fusion techniques

Thirdly, regarding the dataset, half of the research experiments utilized public datasets, while the remainder collected their datasets. Within the subset utilizing custom datasets, 45.45% were dedicated to developing game apps, 18.18% focused on Chat Applications, and 36.36% involved developing unnamed custom apps. Consequently, game apps were the most prevalent, followed by custom apps and chat applications.

In our fourth observation, the pie chart (figure 8) reveals that the majority of papers (52%) do not use any specific data collection application. Among the papers that do use data collection tools, game apps are the most popular, accounting for 20% of the total. Custom apps follow at 16%, indicating a significant reliance on tailored solutions for data collection. Mobile banking applications and chat applications are less commonly used, representing 8% and 4% of the papers, respectively. This distribution suggests a diverse approach to data collection, with a notable preference for using game apps and custom applications over more conventional tools like mobile banking and chat applications.

In the fifth paper summary of findings, a critical comparison of the most widely used keystroke public datasets is presented. Notably, there is a gap in existing datasets as very few have used language variation other than English, and none are based on typing patterns of the Amharic script or an Amharic keyboard. This

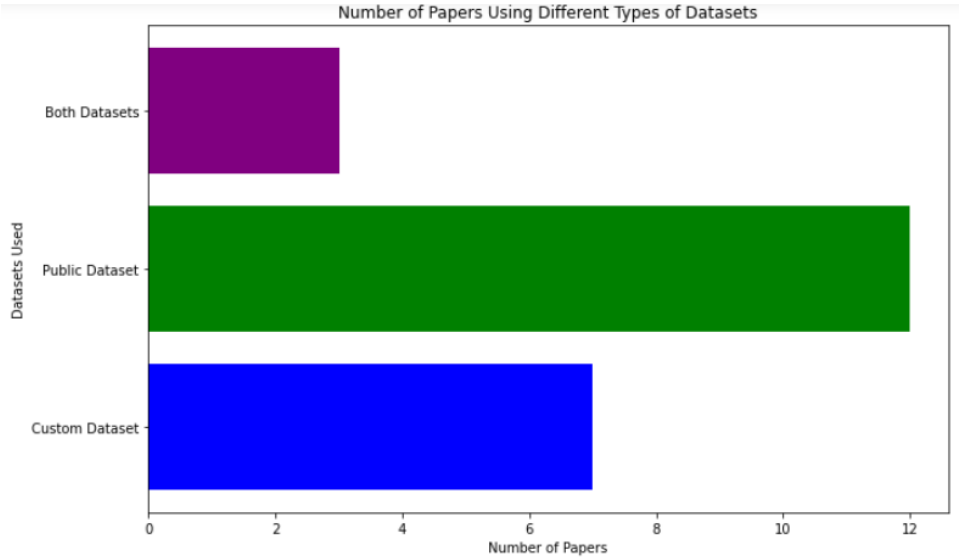

Fig. 7. Dataset variation in selected previous research

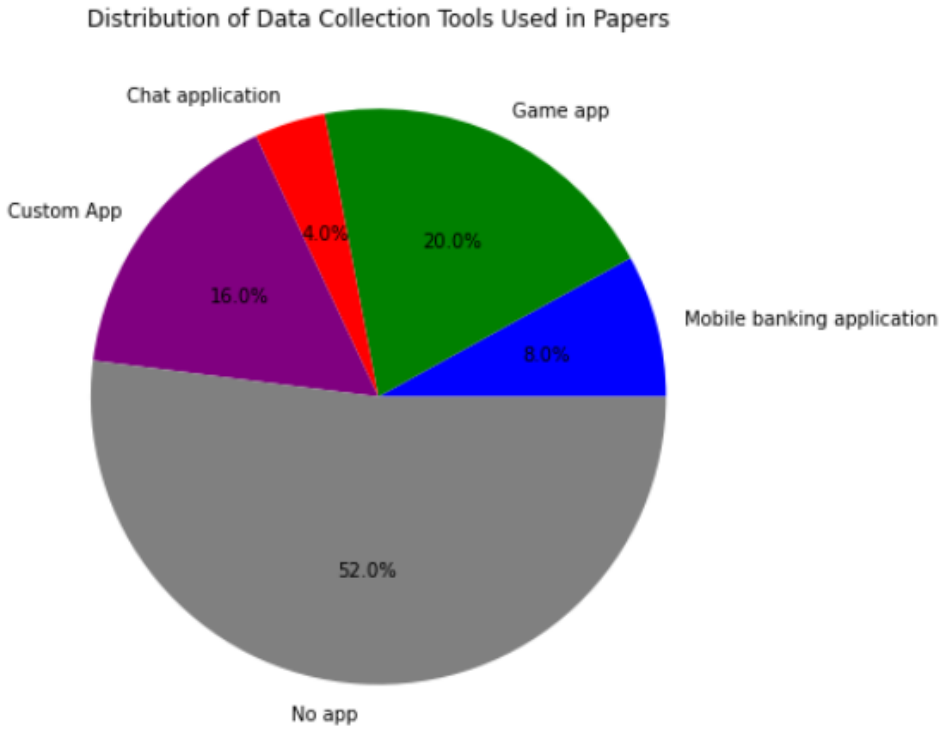

Fig. 8. Data collection tools

void represents our contribution to the field. Table 1 provides a summary of the keystroke datasets most commonly utilized in continuous authentication research and development.

Table 1: A summary of major keystroke datasets

Year	Dataset Name	Text Type	No. of Subjects	Language Variation	Data collection tool	Total No. of Keystrokes	Platform Used
2012	GreyC-B [e]Dataset	Free	-	English	Web	-	Web
2009	[f]CMU Dataset	Fixed Text (Password)	51	English	Microsoft Windows	20,400	Microsoft Windows
2009	GreyC-A Dataset [18]	Fixed Text (Password)	133	English	GREYC-Keystroke Software	7,555	SQLite
2012	GreyC-B Dataset [18]	Free	83	English	Web	16,378	Web
2013	[g]Pace Dataset	Fixed-Text (Numeric)	30	English	Third-Party Keylogger	1800	-
2014	Clarkson I Dataset [18]	Fixed Text, Free Text	39	English	Browser-based Keylogger	21,533	Browser
2015	OhKBIC Dataset [18]	Free Text	64	English	JavaScript Keylogger	6400	Server
2016	Buffalo Dataset [18]	Fixed Text, Free Text	157	English	System Logger	51,000	Windows
2017	Clarkson II Dataset [18]	Free Text	103	English	Keylogger on Personal PCs	12.9M	Personal PCs

[e]https://downloads.greyc.fr/Greyc3DColoredMeshDatabase/
[f] https://github.com/CMU-MultiComp-Lab/CMU-MultimodalSDK
[g] https://www.earthdata.nasa.gov/learn/articles/preparing-for-pace-data

2018	Aalto Desktop Dataset [18]	Controlled Free Text	168,000	English	HTML, CSS, JavaScript	136M	MySQL
2019	Aalto Mobile Dataset [18]	Controlled Free Text	37,370	English	Browser logger	-	Mobile Devices
2021	AR Dataset [18]	Semi-Fixed	44	English	Web Forms	500,000	Web Forms
2022	Multi-K Dataset [18]	Free Text	86	English, Chinese	Web Keylogger	86,000	Desktop and laptop
2024	**Ours (BahriApp Keystroke Dataset)**	**Fixed Text and free text**	**100**	**Anglish Amharic**	**Keylogger On mobile app**	**203,546**	**Smartphone**

In the sixth paper summary findings, we have critically reviewed machine learning model availability across the referenced papers in multimodal behavioral fusion research. As a result, Convolutional Neural Networks (CNNs) emerge as the most prevalent, mentioned in 67% of the papers, followed by Random Forest (RF) and Long Short-Term Memory (LSTM) models, each mentioned in 56% of the papers, indicating their effectiveness and widespread adoption. Specialized architectures like Bidirectional LSTM (BLSTM) and Convolutional LSTM (CLSTM) are also noted, mentioned in 33% and 11% of the papers, respectively, highlighting tailored approaches for sequential data processing tasks. However, certain techniques such as Transfer Learning (TR) are underutilized, mentioned in only 11% of the papers, suggesting potential avenues for future exploration. Overall, while some models enjoy broad usage, the varied selection of techniques underscores the importance of adaptability and selecting appropriate methodologies tailored to the specific challenges of multimodal behavioral fusion research.

Lastly, the analysis of metric usage in behavioral biometrics for continuous authentication research reveals a diverse landscape of evaluation criteria employed across the referenced papers. For authentication, True Positives (TP) represent genuine users who are correctly authenticated, while True Negatives (TN) indicate impostor users correctly denied access. False Positives (FP) occur when impostor users are incorrectly authenticated as genuine, and False Negatives (FN) denote genuine users incorrectly flagged as impostors and denied access. Specificity is the proportion of impostors correctly identified, that is, the recall for the group of impostors. NPV is the effectiveness of the method when predicting impostors, which is the precision for the group of impostors. The f-1 score is a trade-off metric between NPV and Specificity [16].

As a result, the Equal Error Rate (EER) stands out as the most universally

Table 2. ML models and evaluation metrics (Legend: GB=Gradient Boosting Classifier; EL= Ensemble Learning; CLSTM= Convolutional; LSTM BLSTM= Bi-directional LSTM; SA= Self-attention; DT=decision tree; AB=AdaBoost; NN=neural networks; TR= transfer Learning; FS=Few-shot learning; kNN=K-Nearest Neighbor)

Referenced papers	Machine learning or deep learning methods	Evaluation Metrics
[4]	SVM, GB, NN	Accuracy, Precision, False negative, False Positive,
[8]	RF, KNN	EER, Accuracy, Precision, Recall, FAR, FRR, F1− score
[9]	LSTM, TR	EER
[11]	RF, NN	EER
[12]	GB	EER
[16]	SVM	EER, Accuracy, Specificity, FAR, NPV, FRR, oF1-score
[17]	SVM, RF, DT	EER, Accuracy, TAR, Recall, FAR, FRR
[22]	MLP, LSTM, BLSTM, CLSTM	EER, Accuracy, FAR, FRR,
[23]	EL	EER
[28]	MLP	EER
[29]	RF	EER, Accuracy
[31]	NN, FS	EER, FAR, FRR, F1− score
[32]	CNN	Accuracy, TAR, FAR
[33]	LSTM, CNN	Accuracy
[34]	CNN, SA, one-class SVM	EER
[34]	CNN	EER, FAR, FRR
[36]	SVM, RF	EER, Accuracy, TAR, Recall, F1− score
[38]	LSTM	EER

utilized metric, present in 100% of the papers, underscoring its significance in providing a holistic view of system accuracy. Accuracy, True Accept Rate (TAR), and False Reject Rate (FRR) are also prominently featured, utilized in 45%, 35%, and 30% of the papers, respectively, reflecting their importance in assessing overall correctness, user identification, and system reliability. While Precision, Recall, False Accept Rate (FAR), and F1-score are less frequently utilized, appearing in 15%, 10%, 20%, and 20% of the papers, respectively, they nonetheless provide valuable insights into minimizing false positives, false negatives, and achieving a balance between precision and recall. Collectively, these metrics play a pivotal role in guiding researchers toward informed decisions in system design and deployment, ensuring

the effectiveness and reliability of behavioral biometric systems.

2.6. *Gap analysis*

Our literature review reveals significant gaps in evaluating adaptive biometric authentication systems on mobile devices, particularly regarding accommodating evolving hardware limitations. While some research focuses on constructing robust networks using deep learning models for processing user activity data like gestures and human motions [31] , there's a noticeable lack of effort in applying these hybrid patterns to low-resource-constrained user behaviors such as keystrokes and screen touch data. Dataset availability and shortage is another critical gap, exacerbated by the digital divide influencing users' device usage patterns due to IT skill and accessibility disparities. To solve the dataset, challenge an attempt by Hossein et al. [31] utilized few-shot learning for rapid large-scale user authentication, requiring minimal training data. Their system has employed a dynamic Siamese neural network architecture solely based on motion patterns from accelerometer, gyroscope, and magnetometer data. Currently, there is considerable interest in Siamese neural networks. For example, James et al. [35] introduced the Multi-Modal Siamese Convolutional Neural Network (mmSNN). This model was designed to learn spatial and temporal information independently and then fuse sensor data within a Siamese framework. Its primary goal is to predict a person's identity. Hence, the limitations inherent in keystroke data, owing to its dependence on specific applications, underscore a significant research gap in utilizing Inertial Measurement Unit (IMU) sensors alongside fundamental behavioral biometrics.

Additionally, to develop high-performance authentication models, it is crucial to know the distribution of negative training data stemming from a variety of attackers. However, getting imposters or attackers' data is challenging. We don't know how they behave. To solve this, an attempt to reference [24], authors propose a relative attention-based one-class adversarial autoencoder architecture for continuous authentication of smartphone users. This architecture comprises four key components. Firstly, the One-Class Adversarial Autoencoder learns the legitimate user's behavioral patterns solely from positive samples in an unsupervised manner. The Latent Discriminator ensures that the latent representations of legitimate user samples adhere to a uniform spatial distribution. Meanwhile, the Sample Discriminator distinguishes between positive and negative samples generated by the decoder, aiding the autoencoder in reconstructing higher-quality positive samples during training. Lastly, the Relative Attention Mechanism, utilizing convolution projection, enhances the model's capability to capture contextual semantic information from behavioral biometrics, particularly beneficial in scenarios with limited computing power such as smartphones.

Finally, socioeconomic factors, demographics, and technological adoption patterns further compound the digital divide, emphasizing the need for custom behavioral biometric datasets tailored to specific demographics, such as third-world

countries of certain age and literacy levels, to develop effective biometric models and identity solutions. For example, Pedro et al. [8] , used a small sample size comprising only 15 subjects to collect the touch gesture dataset. Moreover, there's a dearth of research on the influence of the Amharic language on keystroke dynamics-based behavioral biometric authentication systems. Among the existing studies, the most recently published research [17] focuses on a Semitic language i.e. Arabic keystroke, which serves as a baseline for investigating Amharic, a related language within the same language family.

3. Conceptual Framework Towards Continuous Authentication

After conducting a thorough literature review, we formulated our research questions and devised a comprehensive conceptual framework centered around three key concepts: domain knowledge of selected features within our dataset, the learning process involved in continuous authentication tasks, and the flow of a multi-modal authentication system. To illustrate this framework, we utilized three distinct visualization tools: a mind mapping tool, a flowchart, and UML sequence diagrams, sequentially. We opted for the sequence diagram due to its widespread usage in UML modeling. Moreover, it effectively illustrates the behavioral interaction and message exchange among various objects.

3.1. *Authentication phases*

The learning process in continuous authentication has three phases. i.e. Enrollment Phase, Continuous Authentication Phase, and verification phase.

(1) **Enrollment Phase**: Initially, the system undergoes training during the enrollment phase. This involves using the fictitious mobile application with the keylogger capability. During this phase, we identify two specific user actions, typing Amharic and English script which is copy-locked for some amount of time t, and navigation gestures such as sliding and scrolling such as selecting the bank to transfer and button press. Other miscellanies activities may be also required such as number entry (bank account number, OTP entry, transfer reason, etc.).
(2) **Continuous Authentication Phase**: Following the training of classifiers, the system progresses into the continuous authentication phase, where it continuously monitors all user input, including keystrokes and other interactions. During this phase, the trained classifier assesses whether these actions align with patterns characteristic of the legitimate user. Should consecutive negative classification results occur, suggesting potential unauthorized access, the system initiates a fallback to the initial entry-point authentication method. Consequently, the user is prompted to reauthenticate using the original authentication method.

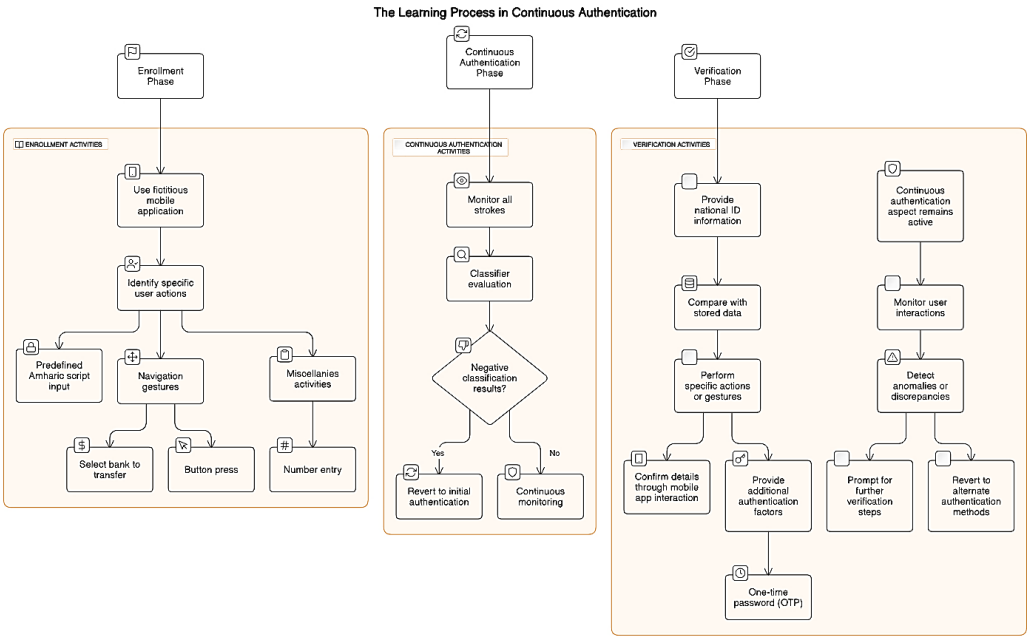

Fig. 9. Phases in continuous authentication

(3) **Verification Phase**: During the Verification Phase, which follows successful enrollment and system training, users are prompted to provide their national ID information, optionally supplemented with biometric data like fingerprints or facial recognition. The system then compares this provided information with stored data obtained during enrollment. Additional verification steps may involve user actions or gestures, such as confirming details through interaction with a mobile application or providing a one-time password (OTP) as an extra authentication factor. Throughout this phase, continuous authentication remains active, monitoring user interactions for anomalies indicating potential unauthorized access. If the provided information aligns with the user's profile and continuous authentication validates legitimacy, access is granted; otherwise, further verification steps may be required, or alternate authentication methods may be invoked to ensure security and user integrity. This comprehensive verification process underscores system security and fosters user trust.

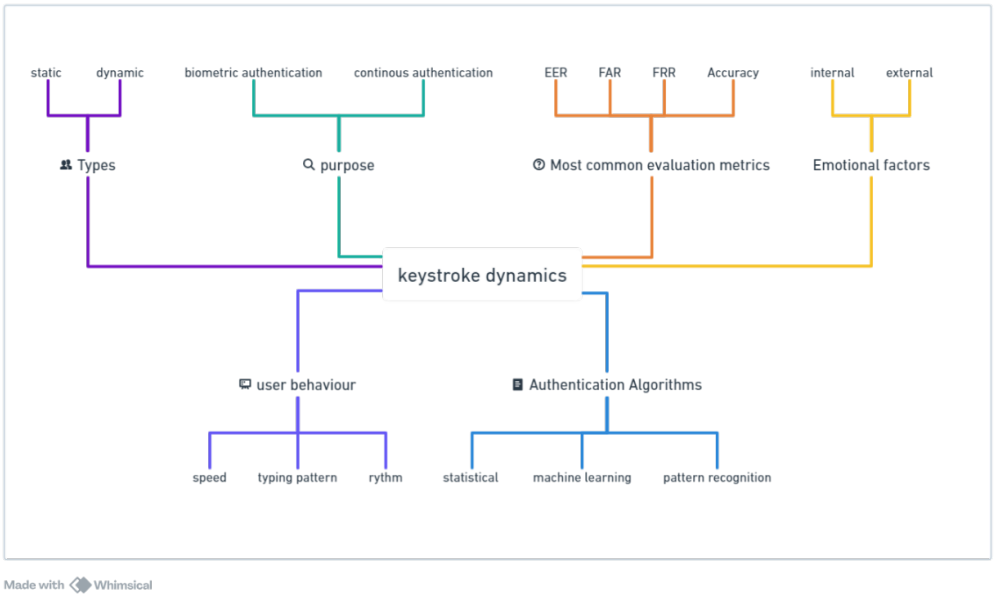

Fig. 10. Mind Map Illustration of the Domain of Keystroke Dynamics

3.2. *Conceptualizing Behavioral Biometric Modalities*

3.2.1. *Keystroke Dynamics in Focus*

Keystroke is a biometric modality utilized in both traditional authentication methods, such as one-time login and trust, as well as in continuous adaptive authentication systems. However, its implementation poses challenges due to various internal and external factors that can cause deviations from a user's normal typing behavior. Figure 8 illustrates a comprehensive conceptualization of Keystroke dynamics.

Considering keystroke dynamics for authentication without adaptive mechanisms poses a significant challenge due to substantial intra-class variation attributable to aging [7]. The aging process introduces degradation in the accuracy of keystroke dynamic authentication. To improve keystroke authentication, we introduce a multi-modal authenticator that leverages user-entered Amharic text during keystrokes. Therefore, both the content input and the dynamics of typing will be used as complementary attributes. Our approach capitalizes on the inherent linguistic characteristics of Amharic, a low-resource language, to enhance authentication.

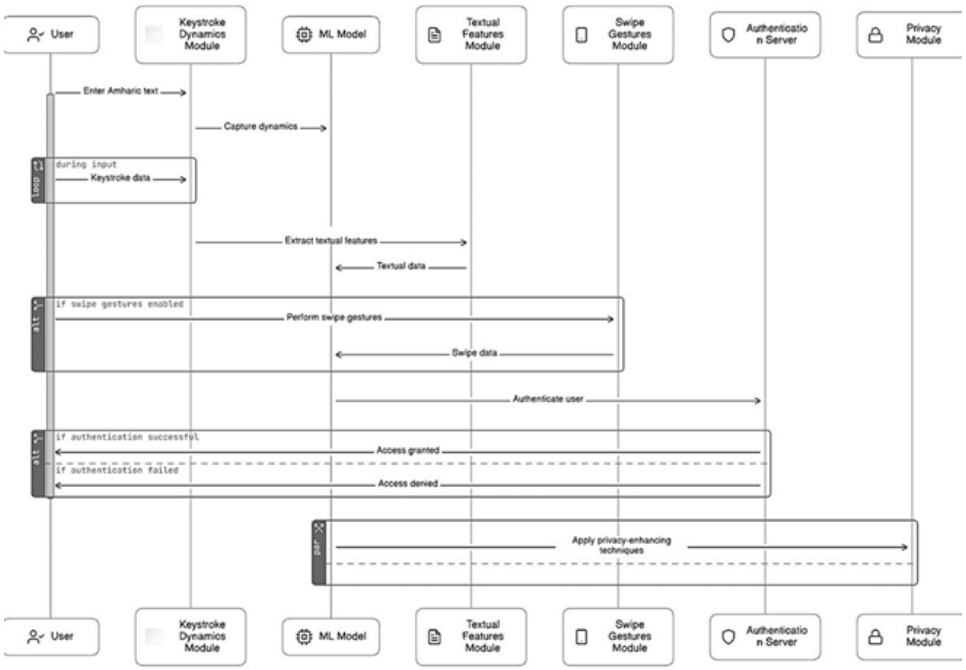

Fig. 11. Proposed Multi-Modal Biometric Authentication Procedure

3.2.2. *Proposed Framework for multi-modal Authentication System Flow*

The interaction among the Machine Learning model (ML model), authentication server (part of the system architecture of which is detailed in Section 4, and all data collection modules is illustrated in Figure 10.

4. Proposed system architecture

In this section, we present our system architecture named "**BehFayda**" to indicate a behavioral biometric-based Identity system dedicated to continuous authentication. We provide an overview of the concept; detail of every component including the functionality of each engine and explain the modeling process for the various collected data types. The BehFayda Architecture encompasses multiple hardware devices and components, designed to operate synergistically within a secure framework. These components include the user's smartphone, the Fayda authentication server, and the relying party (RP) server. Figure 10 illustrates the integrated operation of these hardware devices across the secured channel. This architecture ensures robust functionality and secure communication between the user, the authentication server, and the relying party server, facilitating seamless operation and enhanced security measures. These three servers are geographically dispersed and connected over the secured channel.

The smartphone serves as the primary interface for the continuous acquisition of user behavioral data. It hosts a data acquisition engine tasked with seamlessly gathering dynamic behavioral data as users interact with their smartphones. To optimize resource usage on smartphones, resource-intensive tasks such as feature extraction, normalization, and model training are offloaded to the authentication server. The smartphone's role is to collect and transmit data to the corresponding user model on the authentication server.

4.1. *Data Collection process*

To collect authentication data, we developed an Android-based application called BahriApp. Android was chosen for its wide adoption and robust API access to device sensors and functionalities. The app captures diverse typing patterns, including different words in Amharic and English and strong password entries. Totally, 450 users were exposed to install the app and play keystroke games. All captured were stored locally and synchronized with a Firebase database. In future implementations, sensor data will be integrated into the Authentication Engine (Auth Engine) and homomorphically encrypted using individual public keys to ensure privacy. Similarly, contextual data requested by the Risk Assessment Engine will be encrypted with distinct keys and functions for secure processing.
Table *4* Number of sessions applied to collect Keystroke data in three months period from 450 users

Table 3. Number of sessions applied to collect Keystroke data in three months period from 450 users

	Total Number of user sessions
Keystroke Free Amharic text typing session count	763
Keystroke Free English text typing session count	2120
Keystroke Fixed Amharic Text typing Session count	2291
Keystroke Fixed Text typing Session English	4905
Keystroke Strong Password typing Session English	2836

4.2. *Architecture Components and Interactions*

4.2.1. *Authentication Engine (Auth)*

The training model will classify the user as legitimate or imposter using our proposed classification machine learning methods (proposed in section 3). These ML models are authentication models deployed in the Auth engine. Therefore, the Auth engine makes the final decision about whether or not to authenticate a user. It does this by considering the output of the Adaptation Engine (AE) and risk assessment engine. If the authentication results indicate the user is benign, the Risk Assess-

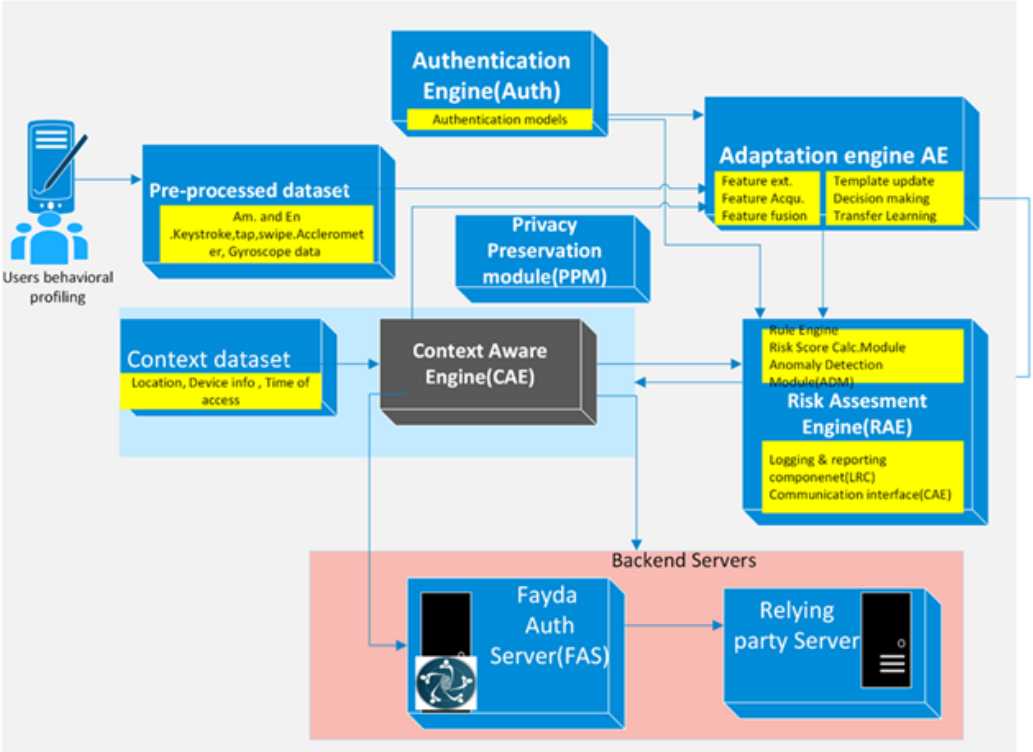

Fig. 12. Proposed System (**BehFayda**) Architecture

ment Engine (**RAE**) will verify the genuine identity of the user and will allow to use of the Relying party's app (mobile banking for example) to access transactions. This time the user starts to interact with the RP's server where the mobile app backend is deployed.

4.2.2. *Fayda authentication server (**FAS**)*

FAs is the central server of the system. It stores user data and authentication templates. The model will be deployed on the FAS for ease of security, trust, easy updates, and maintenance without requiring changes to be pushed to individual user devices. All the modules discussed in the upcoming sections including Auth engine, RAE, AE, and PPM are deployed on the FAS. We assume that the authentication model is trained at the FAS because FAS is trusted and we can avoid adversarial machine learning such as model poisoning, or privacy attacks.

4.2.3. *Risk Assessment Engine (RAE)*

This engine serves as the core component responsible for evaluating risk factors associated with each authentication request within the system. The following are its proposed sub-modules for different functionalities.

(A) **Communication interface with CAE**. This interface enables the RAE to learn various user context parameters including transaction type, user history, device information, geolocation, time of access, and any other relevant data to assess the risk level associated with a particular authentication request.

(B) **Risk Score Calculation Module**: Based on the evaluation of the above factors, the RAE calculates a risk score indicative of the level of risk associated with the authentication attempt. Then it will categorize risks as low, medium, and high risk.

Example scenario: Let's say we have a user, Daniel, who typically logs in to a mobile banking system from his home computer around 8:00 AM every weekday. The system records Daniel's login times, device information, IP address, and other relevant data. Now, Daniel attempts to log in from a different device (a smartphone) at 2:00 AM on a Saturday, which is unusual compared to his typical behavior. The Risk Score Calculation Module would evaluate this authentication attempt based on factors such as:

- User Behavior Logging in at an unusual time (2:00 AM)
- Device Information Using a different device
- Environmental Context Logging in from a different IP address

Based on these factors, the Risk Score Calculation Module may assign a higher risk score to this authentication attempt. This higher risk score indicates to the system that additional multi-factor authentication measures to continuously monitor him.

Although Risk-Based Authentication (RBA) models are widely adopted by prominent online platforms such as Amazon, Google, and LinkedIn, there is no standardization yet. Therefore, based on the success reported in previous studies, we have integrated the risk calculation formula outlined in the reference [24] into our proposal.

The model calculates the risk score. S for a user u and a set of feature values $\left(\mathrm{FV}^1, \ldots, \mathrm{FV}^\mathrm{d}\right)$ with features such as:

$$S_u(FV) = \left(\prod_{k=1}^{d} \frac{p(FV)^k}{p(FV^k|u, legit)}\right) \frac{p(u|attack)}{p(u|legit)} \tag{1}$$

S contains the probabilities $\mathrm{p}\left(\mathrm{FV}^\mathrm{k}\right)$ representing the likelihood of a feature value appearing in the overall login history of all users, $\mathrm{p}\left(\mathrm{FV}^\mathrm{k}|\mathrm{u}\right.$, legit) representing the likelihood that a legitimate user possesses this fea-

ture value in their login history. The probability p(u|attack) indicates the likelihood of the user being under attack, while p(u| legit) represents the likelihood of the legitimate user logging in.

(C) **Communication with Authorization Engine (Auth):** This communication involves transmitting the calculated risk score from the RAE to the Auth. This enables the Auth to make informed decisions about the appropriate level of authentication required for a given request. The risk score, along with any relevant contextual information, is securely transmitted using a predefined communication protocol. The Auth uses this information to dynamically adjust authentication requirements based on the perceived level of risk.

(D) **Rule Engine:** The RE allows for the definition of rules and policies that dictate how authentication requirements should be escalated based on risk scores and anomaly detection. These rules govern when and how additional authentication measures are triggered in response to identified risks.

(E) **Anomaly Detection Module (ADM):** This module utilizes previously proposed deep learning algorithms (in section 3) to detect anomalous behavior or suspicious activities that may indicate potential security threats. This module enhances the accuracy of risk assessment by identifying deviations from expected user behavior or system norms.

(F) **Logging and Reporting Component (LRC):** Records all activities related to risk assessment, including data inputs, calculated risk scores, and actions taken based on risk assessment outcomes. It provides auditing capabilities and generates reports for system administrators to monitor authentication activities and identify any security incidents or trends over time.

4.2.4. *The Adaptation Engine (AE)*

The adaptation engine (AE) is responsible for feature acquisition, feature extraction, feature fusion, template update, and decision-making.

(1) **Feature Acquisition**: is responsible for gathering relevant data from different modalities.

(2) **Feature Extraction:** Feature extraction is performed for each modality before the fusion step.

- Touchscreen gestures (taps, swipes, motion data): touch-related features both basic and derived as discussed in section 4 will be extracted.
- Keystroke Dynamics (Amharic and English Typing): All timing-related attributes of keystrokes are considered. Moreover, the linguistic patterns unique to each language are taken into consideration.
- Textual Input (Processed by AM-RoBERTa): The AM-RoBERTa [20] Pre-trained model processes textual input in the Amharic language

and extracts high-level linguistic features, including semantic information, language usage patterns, and individual writing styles. The feature extraction phase involves obtaining embeddings or representations from the output of AM-RoBERTa, capturing the learned linguistic features encoded in the contextualized word embeddings.

4.2.5. *Proposed Feature Fusion Strategy in AE*

We proposed the fusion algorithm called Multi-Modal Deep Residual Fusion (MM-DRF). Our strategy is a hybrid effect of several novel elements. First, to develop a better user behavioral profile, we integrated multiple behavioral biometric modalities, including touchscreen gestures (taps, swipes) from smartphones, keystroke dynamics during typing in Amharic and English languages, and textual input processed by the AM-RoBERTa pre-trained model for Amharic language. Second, we propose the use of residual learning principles for fusion which is novel in the context of behavioral biometrics. We introduce residual blocks to capture residual information between modalities, MM-DRF can effectively leverage the complementary nature of different behavioral cues while mitigating redundancy. This approach enhances the efficiency and effectiveness of multi-modal fusion, especially in capturing subtle variations in user behavior. Third, we proposed an adaptive attention mechanism because it helps to dynamically weigh the contribution of each modality. By adaptively adjusting fusion weights based on the relevance and discriminative power of each modality, MM-DRF can optimize performance in real time, thereby improving adaptability and robustness in dynamic authentication scenarios. Therefore, a combination of various deep learning techniques, including residual learning, attention mechanisms, and pre-trained language models will enhance and contribute to modeling complex behavioral biometric data.

4.2.5.1. Mathematical Modeling for MM-DRF

Let $\mathbf{X}_{\mathrm{t}}^{(1)}$ represent the feature vector extracted from touchscreen gestures, $\mathbf{X}_{\mathrm{t}}^{(2)}$represent the feature vector extracted from keystroke dynamics during typing in Amharic, and $\mathbf{X}^{(3)_{\mathrm{t}}}$represent the feature vector extracted from keystroke dynamics during typing in English at the time t.

The feature vector $\mathbf{X}_{\mathrm{t}}$ is formed by concatenating the features from all modalities:

$$X_t = [x_t^{(1)}, x_t^{(2)}, x_t^{(3)}] \tag{2}$$

(A) Residual Learning Fusion

We introduce residual blocks specific to each modality to capture residual information. Let $\mathbf{F}_{\mathrm{t}}^{(\mathrm{k})}$note the feature representation obtained after processing the input from the modality k up to the residual block. The residual

block $\mathbf{R}_{\mathrm{t}}^{(\mathrm{k})}$ for each modality kIs defined as:

$$\boldsymbol{R}_t^{(k)} = \sigma\left(\boldsymbol{W}_r^{(k)} \cdot \boldsymbol{F}_t^{(k)} + \boldsymbol{b}_r^{(k)}\right) \tag{3}$$

where $\mathbf{W}_{\mathrm{r}}^{(\mathrm{k})}$ and $\mathbf{b}_{\mathrm{r}}^{(\mathrm{k})}$ are the weight matrix and bias vector of the residual block for modality k, respectively, and σ represents the activation function.

The output feature representation after the residual block for each modality is obtained by adding the residual to the original feature:

$$\boldsymbol{F}_{t+1}^{(k)} = \boldsymbol{F}_t^{(k)} + \boldsymbol{R}_t^{(k)} \tag{4}$$

(B) Multi-Modal Integration

In this work, we propose Feature-Level Fusion combined with Attention Mechanisms. This mechanism is employed to focus on specific regions or aspects of the input data that are relevant for authentication. It helps the model prioritize important features or behaviors while disregarding irrelevant ones. We employ attention mechanisms to dynamically weigh the contribution of each modality. Let $\alpha_{\mathrm{t}}^{(\mathrm{k})}$ denote the attention weight assigned to the modality k at time t. The attention weight is computed as:

$$\alpha_t^{(k)} = \frac{exp\left(e_t^{(k)}\right)}{\sum_{i=1}^{3} exp\left(e_t^{(i)}\right)} \tag{5}$$

where $\mathrm{e}_{\mathrm{t}}^{(\mathrm{k})}$ is the attention score for modality k at time t, calculated using a learnable parameter matrix $\mathbf{W}_{\mathrm{a}}$ and a non-linear activation function ϕ :

$$e_t^{(k)} = \phi\left(\boldsymbol{W}_a \cdot \boldsymbol{F}_t^{(k)}\right) \tag{6}$$

The attention-weighted feature representation $\tilde{\mathbf{F}}_{\mathrm{t}}$ is computed by combining the feature representations from all modalities:

$$\tilde{\boldsymbol{F}}_t = \sum_{k=1}^{3} \alpha_t^{(k)} \cdot \boldsymbol{F}_t^{(k)} \tag{7}$$

Generally, at each time step t, the multi-modal input feature vector $\mathbf{X}_{\mathrm{t}}$ is processed through the respective residual blocks to obtain the feature representations $\mathbf{F}_{\mathrm{t}}^{(\mathrm{k})}$. These feature representations are then combined using attention mechanisms to generate the final fused representation. $\tilde{\mathbf{F}}_{\mathrm{t}}$.This fused representation can be further processed through additional layers (e.g., fully connected layers or recurrent layers) for classification tasks.

During the training phase, the parameters of the residual blocks $\mathbf{W}_{\mathrm{r}}^{(\mathrm{k})}$ and $\mathbf{b}_{\mathrm{r}}^{(\mathrm{k})}$, as well as the attention mechanism parameters $\mathbf{W}_{\mathrm{a}}$, are learned using backpropagation and optimization techniques.

Additionally, the Amharic RoBERTa [20] model, a transformer-based language model, is integrated into the AE. Specifically, it belongs to the RoBERTa (Robustly optimized BERT approach) architecture, which is based on the transformer architecture. RoBERTa is a variant of BERT (Bidirectional Encoder Representations from Transformers), designed to improve the pretraining of deep bidirectional transformers for language understanding tasks. This pre-trained model enhances our limited dataset with its inherent ability to capture meaningful linguistic patterns in Amharic.[h].For example, a study referenced as [9], their use of transfer learning resulted in a significant 12.16% improvement in EER. All in one, we Fine-tune it on labeled data specific to our authentication task and the adaptation engine continuously analyzes successful and failed authentication attempts using the Feedback Loop mechanism and updates the model iteratively based on real-world feedback. It is also crucial in our Privacy Preservation module because of its ability to anonymize sensitive information in text inputs.

(C) Template Update:

This component updates the user's authentication template over time. This is important because a user's behavior can change over time.

(D) The decision-making:

This component within the Adaptation Engine (AE) is responsible for processing the fused representations generated by the template update component and making decisions regarding user authentication. First, it is responsible for receiving the fused representation, which contains integrated features from various behavioral biometrics. Second, its task is to classify the input data into legitimate or illegitimate user behavior. Third, it determines the Threshold by using either the predefined thresholds or dynamically adjusted thresholds based on the current context and the user's historical behavior. Thresholds define the level of similarity required between the current behavior and the stored template for authentication to be granted. Finally, it provides prompt responses to authentication requests. This component is also responsible for consulting the Risk assessment module to gather supplementary information for making authentication decisions. Finally, the authentication decision will be sent to the relying party server (bank server).

[h] https://iopscience.iop.org/article/10.1088/1742-6596/1529/2/022088

4.2.6. *Privacy-preservation Module (PPM)*

MOSIP has recently advised nations to implement its Digital Public Good (DPG) to address privacy issues both legally and technically [34]. While MOSIP adheres to the privacy by design principle, revisiting ID artifacts with the inclusion of behavioral biometrics, as suggested in our study, necessitates a separate set of redesign proposals. In addition, application authentication is typically performed using some form of secret credentials such as cryptographic keys, passwords, or API keys. Since clients are responsible for securely storing and managing the keys, the conventional approach is vulnerable to attacks on clients, such as Key Compromise, brute force attacks, and Keylogging attacks. Mihai et al. [35] Proposed an approach called Behavioral Application Authentication using a Fuzzy Extractor (user's biometric data to decommit a cryptographic key) to counteract vulnerabilities associated with traditional application authentication methods. They propose to establish application-to-application interaction by leveraging Behavior Monitors embedded within applications, such as web servers (e.g., IDS/IPS).

Behavioral biometrics like numerous other technological advancements, presents privacy concerns [31]. In authentication systems utilizing behavioral biometrics, data collection often occurs covertly or passively, leaving individuals without the ability to give consent or exert control over the collected information and its usage. In response to this challenge, we proposed the PPM. During the implementation phase, the proposed engine will utilize the three proposed mechanisms and algorithms. As per our knowledge, there is a lack of research in the context of continuous authentication utilizing our proposed behavioral biometrics features (keystroke, touch, motion), except for the mpsauth proposal outlined in [38]. However, they have exclusively employed homomorphic encryption, specifically applied to behavioral data before its transmission from the user's device and they didn't incorporate feedback from previous authentication processes. In our work, we proposed a feedback loop mechanism in the adaptation engine (AE) to adapt to users in the long run to attain better accuracy from time to time. Moreover, in BehFayda, we proposed three solutions to ensure strong privacy i.e.:

(i) Data anonymization: This component anonymizes the training data before transmission to the authentication server. By removing personally identifiable information (PII) or replacing it with pseudonyms, the user's identity is protected during data transfer.

(ii) Homomorphic encryption: In homomorphically encrypted data, various tools support operations such as neural networks, decision trees, and logistic regressions [35].So, our goal in this architecture is to conduct the model training on the encrypted data for better privacy. This component Encrypts communication channels. It is also responsible for performing feature extraction and model training operations on encrypted data. In action, the secrecy encoder unit of the PPM forwards the encrypted data from the

client device to the authentication server.

(iii) Differential Privacy: Implementing differential privacy techniques ensures that individual user contributions to the training dataset remain confidential. It adds noise to the data to prevent the reconstruction of sensitive information about individual users while still allowing meaningful aggregate analysis.

4.2.6.1. Context Aware Engine (CAE)

This component takes into account contextual factors such as location, device information, IP address, and time of access. This engine is significantly unique because this information is live-fed when the risk assessment engine is calculating a high risk.

5. Conclusion and future works

In conclusion, the comprehensive literature review on MOSIP Authentication Architecture and biometric authentication adaptation techniques underscores the pressing need for advancements in continuous, context-aware authentication systems within the MOSIP framework and beyond. While existing research has made strides in exploring touch-based and keystroke-based authentication methods and fusion techniques, significant gaps persist, particularly in the development of adaptive systems, dynamic authentication methods leveraging smartphone movement signals, and effective fusion methods for diverse user behaviors. Moreover, challenges related to dataset availability, socioeconomic factors, and language-specific considerations highlight the imperative for tailored solutions and inclusive approaches in biometric authentication research. Addressing these gaps will be paramount for advancing the field and ensuring the development of secure, user-friendly, and inclusive identity verification systems suited for diverse demographics and languages.

First, this paper has presented a comprehensive conceptual framework for a multi-modal authentication system, focusing on the learning process involved in continuous authentication tasks and the domain knowledge of selected features within our dataset. The framework was illustrated using three distinct visualization tools: a mind mapping tool, a flowchart, and UML sequence diagrams, providing a clear understanding of the behavioral interaction and message exchange among various objects. The learning process in continuous authentication was divided into three phases: the Enrollment Phase, the Continuous Authentication Phase, and the Verification Phase. Each phase plays a crucial role in ensuring the security and integrity of the user during the authentication process. Keystroke dynamics, a biometric modality utilized in both traditional and continuous adaptive authentication systems, was explored in depth. Despite the challenges posed by various factors that can cause deviations from a user's normal typing behavior, we proposed a multimodal authenticator that leverages user-entered Amharic text during keystrokes to

enhance authentication. Second, we proposed a conceptual map for a multi-modal authentication system flow, illustrating the interaction among the Machine Learning model, authentication server, and all data collection modules. By leveraging the strengths of multi-modal biometrics and continuous authentication, we aim to provide a robust and user-friendly solution for secure access control. Third, this paper has presented a comprehensive system architecture named "BehFayda" for a continuous authentication system based on behavioral biometrics. The proposed architecture encompasses multiple hardware devices and components, including the user's smartphone, the Fayda authentication server, and the relying party server, all designed to operate synergistically within a secure framework. The proposed system architecture leverages multiple behavioral biometric modalities, including touchscreen gestures (taps, swipes) from smartphones, keystroke dynamics during typing in Amharic and English languages, and textual input processed by the AM-RoBERTa pre-trained model for Amharic language. The proposed fusion algorithm, Multi-Modal Deep Residual Fusion (MM-DRF), effectively leverages the complementary nature of different behavioral cues while mitigating redundancy, enhancing the efficiency and effectiveness of multi-modal fusion. Furthermore, privacy concerns are addressed through mechanisms such as data anonymization, homomorphic encryption, and differential privacy techniques. The proposed system architecture and fusion algorithm present a promising approach to enhancing the security and reliability of authentication systems.

Future work will focus on implementing and testing the proposed system in real-world scenarios to evaluate its performance and usability. This includes the development of a data collection application tailored for the Android mobile platform and the design of experiments to test the effectiveness of the proposed system. Furthermore, continuous refinement and optimization of privacy preservation techniques, particularly in response to evolving privacy regulations and concerns, should be prioritized. Future work should also include the details of the security analysis of the proposed system in terms of different characteristics such as Mutual authentication, forward security, resistance to replay attacks, resistance to man-in-the-middle attacks, and data confidentiality and integrity [32]. Lastly, exploring the integration of emerging technologies such as self–sovereign identity (SSI) and blockchain for enhancing data security and transparency in authentication processes could be an interesting direction for future research. By addressing these areas, BehFayda can evolve into a robust and adaptable solution for continuous authentication in various domains. Moreover, future research will also explore the potential of integrating additional behavioral biometric modalities and improving the fusion algorithm to further enhance the system's performance.

Data Availability The datasets generated and analyzed during this study are not publicly available due to privacy regulations. However, the data may be made available by the corresponding author upon reasonable request, subject to approval from the institutional review board and compliance with privacy regulations.

Acknowledgment We would like to express our gratitude to Addis Ababa Science and Technology University for supporting this work and for covering the presentation fee for the STII 2024 Conference.

References

1. A. Nuttapong, H. Goichiro, M. K.-X. Haochen and M. Takahiro, "Two-Dimensional Dynamic Fusion for Continuous Authentication," IEEE, 2023.W475W9648
2. B. Nandini, "Behavioural biometrics in action," Biometric Technology Today, vol. 2020, no. 10, 2021. W475W9648
3. K. SIMON, "DIGITAL 2023: ETHIOPIA," 13 FEBRUARY 2023. [Online]. Available: https://datareportal.com/reports/digital-2023-ethiopia.W475W9648
4. P. Brendan, V. Mounika and D. Rushit, "Your Identity is Your Behavior - Continuous User Authentication based on Machine Learning and Touch Dynamics," IEEE, Claire Eau Claire, 2022.W475W9648
5. C. Heather, R. Karen, and S. Tim, "A framework for continuous, transparent mobile device authentication," Computers & Security, vol. 39, pp. 127-136, 2013. W475W9648
6. C. Nicholas, L. Seth, Z. Gang, H. Blair, G. Paolo, and S. B. Kiran, "SMARTCOPE: Smartphone Change Of Possession Evaluation for continuous authentication," Pervasive and Mobile Computing, vol. 97, 2024. W475W9648
7. R. Riseul, Y. Sonja, D. Herber, and D. Julian, "The design and evaluation of adaptive biometric authentication systems: Current status, challenges, and future direction," ICT Express, vol. 9, no. 6, pp. 1183-1197, 2023. W475W9648
8. d. N. Pedro Gomes, W. Pidge, M. Tucker and Zachary, "YOUR DEVICE MAY KNOW YOU BETTER THAN YOU KNOW YOURSELF- CONTINUOUS AUTHENTICATION ON NOVEL DATASET USING MACHINE LEARNING," Minnesota State University, Winterfeldt.W475W9648
9. S. Dilshan, T. Sanuja, V. Maduka, R. Sanka and W. Sandareka, "BehaveFormer: A Framework with Spatio-Temporal Dual Attention Transformers for IMU enhanced Keystroke Dynamics," arXiv, Singapore, 2023.W475W9648
10. H. Maria and A. Ja'far, "A Proposed Password-free Authentication Scheme Based on a Hybrid Vein-Keystroke Approach," in International Conference on New Trends in Computing Sciences, Amman, 2017. W475W9648
11. D.-F. Lehel, E. Kail, and F. Rita, "Two-factor, continuous authentication framework for multi-site large enterprises," in IEEE 20th International Symposium on Computational Intelligence and Informatics (CINTI), 2020. W475W9648
12. A. Peter, V., Mario, J. B. William and T. Zhiyuan, "An omnidirectional approach to touch-based continuous authentication," Computers & Security, vol.

128, 2023. W475W9648

13. F. Mario, B. Ralf, E. M, I. Martinovic and S. Dawn, "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, 2013. W475W9648
14. S.-C. Max and R. Muttukrishnan, "Adaptive Threshold Scheme for Touchscreen Gesture Continuous Authentication using Sensor Trust," in 2017 IEEE Trustcom/BigDataSE/ICESS, London, 2017. W475W9648
15. M. A. ,. E. Priscila, R. d. O., and M. A. Dino, "A Framework for Continuous Authentication Based on Touch Dynamics Biometrics for Mobile Banking Applications," Sensors 2021, 2021. W475W9648
16. M. Alejandro G., M. d. D. Isaac, F.-I. Alberto, B. Marta and F. Rubén R., "Combining user behavioral information at the feature level to enhance continuous authentication systems," Knowledge-Based Systems, vol. 244, 2022. W475W9648
17. A. Najwa, "Authentication by Keystroke Dynamics: The Influence of Typing Language," Applied Sciences, 2023. W475W9648
18. S. Rashik, A. ,. Ahmed, M. Michael, L. Matthew, H. Daqing, and H. Faraz, "Keystroke Dynamics: Concepts, Techniques, and Applications," NY, 2023.W475W9648
19. H. Yang, M. Xiang, Z. Xuan, Y. Wang, L. Yuejun, K. Xiaoyu, S. Jiahui and H. Weiqing, "CKDAN: Content and keystroke dual attention networks with pre-trained models for continuous authentication," Computers & Security, vol. 128, 2023. W475W9648
20. S. M. Yimam, A. A. Ayele, G., G. Venkatesh and I. Biemann, "Introducing Various Semantic Models for Amharic: Experimentation and Evaluation with Multiple Tasks and Datasets," Future Internet, vol. 13, no. 11, 2021. W475W9648
21. L., I. ,. Marcelo and W. Isaac, "Toward a Framework for Continuous Authentication using Stylometry," in IEEE 28th International Conference on Advanced Information Networking and Applications, Toronto, 2014. W475W9648
22. U. Utku, D. İ. Özlem and I. ,. Gülfem, "Evaluation of Deep Learning Models for Continuous Authentication Using Behavioral Biometrics," Procedia Computer Science, vol. 225, pp. 1272-1281, 2023. W475W9648
23. Z., Ahmad, Y. C. Chun, R. P., and S., Ali, "A framework of dynamic selection method for user classification in touch-based continuous mobile device authentication," Journal of Information Security and Applications, vol. 67, 2022. W475W9648
24. H. Mingming, Z. Kun, Y. Ruibang and T. Bibo, "AuthConFormer: Sensor-based Continuous Authentication of Smartphone Users Using A Convolutional Transformer," Computers & Security, vol. 127, 2023. W475W9648
25. M. ,. B. ,. Priscila, d. O. ,. Robson and M. ,. Dino, "Biotouch: A Framework Based on Behavioral Biometrics and Location for Continuous Authentication

on Mobile Banking Applications," in 15th Iberian Conference on Information Systems and Technologies (CISTI), 2020. W475W9648

26. S. M. R. S. Divya, T. Sim, and Z. Yair, "Context-Aware Fusion for Continuous Biometric Authentication," 2018.W475W9648
27. L. Nick and K. W. Manfred, "The Weighted Majority Algorithm," in 30th Annual Symposium on Foundations of Computer Science, NC, 1989. W475W9648
28. S. Ioannis, C. Sotirios, T. Olga, and K. Spyros, "Continuous authentication with a feature-level fusion of touch gestures and keystroke dynamics to solve security and usability issues," Computers & Security, vol. 132, 2023, 103363. W475W9648
29. H. MINGMING, Z. KUN, Y. RUIBANG, and T. BIBO, "Relative Attention-based One-Class Adversarial Autoencoder for Continuous Authentication of Smartphone Users," ACM, 2022.W475W9648
30. A.-S. Jaafer and R., Mohammad, "Keystroke and swipe biometrics fusion to enhance smartphones authentication," Computers & Security, 2023. W475W9648
31. F. Hossein, J. König, R. Phillip, C. Marco, G. Bora, F. Moritz, D. Alexandra, and S. Ahmad-Reza, "AuthentiSense: A Scalable Behavioral Biometrics Authentication Scheme using Few-Shot Learning for Mobile Platforms," in Network and Distributed System Security (NDSS) Symposium 2023, San Diego, 2023. W475W9648
32. G. Aleksei, B. Konstantin and K. Konstantin, "Motion ID: Human Authentication Approach," 2023.W475W9648
33. O. James and A. U. A. Mohammad, "PhysioGait: Context-Aware Physiological Context Modeling for Person Re-identification Attack on Wearable Sensing," arXiv, 2022.W475W9648
34. M. I. David Monschein and Oliver P. Waldhorst, "mPSAuth: Privacy-Preserving and Scalable Authentication for Mobile Web Applications," IEEE, Karlsruhe, 2022.W475W9648
35. A. Alejandro, M. Aythami, V. M. John, V.-R. Ruben and F. Julian, "TypeNet: Deep Learning Keystroke Biometrics," JOURNAL OF LATEX CLASS FILES, vol. 14, no. 18, 2021. W475W9648
36. M. Jacob and P. Laura, "Hold On and Swipe: A Touch-Movement Based Continuous Authentication Schema based on Machine Learning," in 2022 Asia Conference on Algorithms, Computing and Machine Learning 2022, IEEE, 2022. W475W9648
37. M. Sakorn and J. Anuchit, "Deep Learning Approaches for Continuous Authentication Based on Activity Patterns Using Mobile Sensing," Sensors, 2021. W475W9648
38. v. Hiriharan, "Lessons Learned: Reflecting on MOSIP's Journey to Scale," DIGITAL IMPACT ALLANCE, 2024.W475W9648
39. C. Mihai and S. Maliheh, "Privacy-Preserving Application-to-Application Authentication Using Dynamic Runtime Behaviors," arXiv, 2022.W475W9648

40. Y. Bin, Z. Yongdong, L. Shanyun, and X. Tao, "AI-Oriented Two-Phase Multi-Factor Authentication in SAGINs: Prospects and Challenges," aRXrv, Hangzhou, 2023.W475W9648
41. S. Andreas, S. Ioannis, K. Maria, and K. Spyros, "Users' Privacy Attitudes towards the Use of Behavioral Biometrics Continuous Authentication (BBCA) Technologies: A Protection Motivation Theory Approach," Journal of Cyber security and Privacy, vol. 4, pp. 743-766, 2021. W475W9648
42. F. Lex and S. Ariel, "Multi-modal decision fusion for continuous authentication," Computers and Electrical Engineering, vol. 41, pp. 142-156, 2015. W475W9648
43. W. Stephan, T. Jan, and L. I. Luigi, "Privacy Considerations for Risk-Based Authentication Systems," in 2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), 2022. W475W9648
44. S. Giuseppe, V.-R. Ruben, and T. Ruben, "Mobile Behavioral Biometrics for Passive Authentication," Cesson-Sevigne, 2022.W475W9648