Neural Computing and Applications
https://doi.org/10.1007/s00521-024-10403-y

ORIGINAL ARTICLE

# M2auth: A multimodal behavioral biometric authentication using feature-level fusion

Ahmed Mahfouz[1,2] · Hebatollah Mostafa[3] · Tarek M. Mahmoud[4] · Ahmed Sharaf Eldin[5,6]

Received: 1 March 2024 / Accepted: 20 August 2024

## Abstract

Conventional authentication methods, such as passwords and PINs, are vulnerable to multiple threats, from sophisticated hacking attempts to the inherent weaknesses of human memory. This highlights a critical need for a more secure, convenient, and user-friendly approach to authentication. This paper introduces M2auth, a novel multimodal behavioral biometric authentication framework for smartphones. M2auth leverages a combination of multiple authentication modalities, including touch gestures, keystrokes, and accelerometer data, with a focus on capturing high-quality, intervention-free data. To validate the efficacy of M2auth, we conducted a large-scale field study involving 52 participants over two months, collecting data from touch gestures, keystrokes, and smartphone sensors. The resulting dataset, comprising over 5.5 million action points, serves as a valuable resource for behavioral biometric research. Our evaluation involved two fusion scenarios, feature-level fusion and decision-level fusion, that play a pivotal role in elevating authentication performance. These fusion approaches effectively mitigate challenges associated with noise and variability in behavioral data, enhancing the robustness of the system. We found that the decision-level fusion outperforms the feature level, reaching a 99.98% authentication success rate and an EER reduced to 0.84%, highlighting the robustness of M2auth in real-world scenarios.

**Keywords** Biometrics authentication · User authentication · Information fusion · Behavioral analysis

✉ Ahmed Mahfouz
ahmed.m@aou.edu.om; e.ahmedmahfouz@mu.edu.eg

Hebatollah Mostafa
e.hebaelkaiaty@gmail.com

Tarek M. Mahmoud
tarek@fcai.usc.edu.eg

Ahmed Sharaf Eldin
profase2000@yahoo.com

[1] Faculty of Computer Studies, Arab Open University, Muscat, Oman

[2] Computer Science Department, Minia University, Minya, Egypt

[3] Faculty of Computers and Information, Minia University, Minya, Egypt

[4] Faculty of Computers and Artificial Intelligence, Sadat City University, Sadat City, Egypt

[5] Faculty of Information Technology and Computer Science, Sinai University, Arish, Egypt

[6] Information Systems Department, Helwan University, Helwan, Egypt

## 1 Introduction

In an era where digital interactions shape our daily lives, securing personal devices, especially smartphones, has become a critical priority. As smartphones seamlessly integrate into various aspects of our daily routines, from communication to financial transactions, ensuring the confidentiality and integrity of personal data has emerged as a substantial concern. The increasing reliance on smartphones for sensitive tasks, such as mobile banking and accessing confidential information, has heightened the need for robust security measures. Traditional authentication methods, such as passwords and PINs, are vulnerable to diverse threats, ranging from sophisticated hacking attempts to the limitations of human memory [1–6]. Consequently, many smartphone users choose not to secure their devices with locks, primarily due to perceived motivational shortcomings and the inconvenience associated with implementing locking mechanisms [7, 8].

Previous research was conducted to address these vulnerabilities by exploring various approaches, including

Published online: 14 September 2024

physiological biometrics like Iris [9] and Face [10]), as well as behavioral traits like gesture [11], sensors [5], and keystrokes [12]. However, the common behavioral biometric systems in the existing literature adhere to a **unimodal**-based framework, relying on a single source of information for user authentication. Unfortunately, these proposed unimodal systems encounter different challenges such as non-universality and noisy data [13].

In response to these challenges, novel approaches are emerging, leveraging cutting-edge technologies to enhance security without compromising user convenience [14, 15]. Behavioral biometrics, encompassing a range of user actions from touch gestures to keystrokes, is at the forefront of this evolution [16, 17]. However, this field faces several significant challenges that necessitate innovative solutions [4, 18]. One challenge lies in the dynamic and context-dependent nature of user behaviors on smartphones, posing difficulties in developing authentication systems that can effectively adapt to diverse scenarios [19]. Moreover, ensuring the accuracy and reliability of behavioral biometrics in real-world settings is a persistent challenge, as the presence of noise and variability in user actions can impact authentication performance [20]. Additionally, there is a need for comprehensive datasets that reflect authentic user behavior across various application contexts, addressing a crucial gap in current research [21, 22].

To address this critical need for enhanced security and user convenience, this paper introduces M2auth, a multimodal behavioral biometric authentication framework for smartphones. Unlike conventional methods, M2auth integrates three distinct modules: a data collection module, a data analysis and feature extraction module, and a decision fusion module. By fusing multiple authentication modalities, each capturing unique features from user interactions, M2auth creates a robust defense against potential threats [6, 20]. These interactions are diverse, influenced by contextual factors such as texting or connecting to specific networks. We report our findings from an extensive field study involving 52 participants over two months, collecting data from touch gestures, keystrokes, and smartphone sensors. The resulting dataset, comprising over 5.5 million action points, stands as an evaluation dataset to M2auth's reliability and relevance in real-world scenarios. Evaluation through feature-level and decision-level fusion scenarios demonstrates M2auth's exceptional performance, achieving an AUC of 99.98% with an equal error rate (EER) of 0.84%.

The core innovation in the proposed method, M2auth, lies in its comprehensive multimodal approach to behavioral biometric authentication, which integrates data from three distinct sources: touch gestures, keystroke dynamics, and sensor readings. Our contributions from this work are as follows:

- Our work presents a significant contribution through the creation of a comprehensive dataset encompassing 5.5 million events. This dataset, reflecting real-world user behavior across various application contexts, is a cornerstone in achieving M$^2$auth 's exceptional performance and ensuring both accuracy and relevance in authentication.
- M$^2$auth offers a dependable feature vector derived from genuine user-level activities. This ensures the accuracy and reliability of the behavioral traits considered for user authentication.
- M$^2$auth makes a significant contribution by introducing an advanced framework that combines gestures, keystrokes, and accelerometer data for smartphone user authentication. This innovative integration of diverse behavioral modalities sets a new benchmark for accuracy and efficiency in authentication systems.
- M$^2$auth presents a distinctive contribution with the introduction of two strategic fusion scenarios—feature-level fusion and decision-level fusion. These fusion approaches play a pivotal role in elevating authentication performance, effectively mitigating challenges associated with noise and variability in behavioral data.

The rest of the paper is organized as follows: In Sect. 2, we discuss the related works about authentication systems. Section 3 presents the threat model followed by a details description of the M$^2$auth framework modules in Sect. 4. We provide the evaluation and results of M$^2$auth in Sect. 5. Section 6 presents the discussion. Finally, Sect. 7 provides the conclusion and future work.

## 2 Related work

The landscape of behavioral biometric authentication has witnessed significant evolution, leveraging various traits such as touch gestures, keystroke dynamics, and sensor-based data [14]. In this section, we provide a more detailed review of related work in the field of behavioral biometrics, including recent advancements for each behavioral biometric traits.

### 2.1 Implicit authentication based on touch gesture

Behavioral biometric authentication methods leverages various behavioral traits for authentication, including gesture, keystroke, and sensors. The integration of touch gestures as a behavioral trait in implicit authentication represents a significant advancement in the biometric authentication [2, 23, 24]. Touch gestures exhibits two important characteristics: continuity and transparency. The acquisition process of touch gesture is unobtrusive, allowing for data collection during the regular device usage [15, 25, 26]. This noninvasive nature is one of the key advantages, as it enhances user experience by not interrupting normal device interactions, making the authentication process virtually invisible to the user.

Touch gesture-based authentication operates continuously in the background, monitoring user behavior without requiring explicit input, thereby offering a high level of convenience. This transparency is particularly valuable for maintaining a seamless user experience, as it does not require active user participation during authentication [15, 25, 26]. Features extracted from touch gestures, such as speed, pressure, and trajectory, have been shown to be highly discriminative. These features capture subtle distinctions in how different users interact with their devices, making touch gesture biometrics effective in accurately distinguishing individual users [5, 27, 28]. This level of specificity is a significant advantage in enhancing the security of the authentication system, as it reduces the likelihood of false acceptances. The non-intrusive nature of touch gesture data collection is a major benefit. Users are not required to perform specific actions solely for the purpose of authentication; instead, their routine interactions with the device are used for this purpose. This passive data collection reduces user burden and increases the likelihood of widespread adoption [25, 26].

Despite its strengths, touch gesture-based authentication can be susceptible to variations in environmental conditions. Factors such as screen moisture, temperature changes, or user stress levels can affect the accuracy of the biometric readings, potentially leading to increased false rejection rates [16]. These environmental sensitivities can limit the reliability of the system, especially in less controlled settings. Another challenge is the issue of template aging, where the user's biometric profile changes over time due to variations in behavior or physical conditions (e.g., changes in finger pressure due to injury or stress). This can degrade the system's performance over time, necessitating periodic updates or retraining of the model to maintain accuracy [11]. Touch gesture-based authentication systems are inherently tied to the device on which they are implemented. This means that the system's effectiveness can vary significantly between different devices, depending on factors like screen size, touch sensitivity, and device orientation. This limitation restricts the generalizability of the system across different platforms and devices [28]. Although touch gestures are unique to each user, they are not immune to mimicry attacks. Skilled attackers who observe a user's interactions could potentially replicate the gestures with sufficient accuracy to bypass the authentication system. While this risk is lower than with traditional password-based systems, it remains a concern that must be addressed through additional security layers or combined modalities [16].

### 2.2 Implicit authentication based on keystroke dynamics

Keystroke dynamics, an established behavioral biometric trait for authenticating users on computers [29], has also found application in smartphones. It involves analyzing the unique rhythm, timing, and pressure applied during a user's keyboard input, creating a personalized and distinguishable authentication signature [12, 30]. Similar to the touch gesture, keystroke dynamics support the continuity and transparency, allowing for implicit acquisition without causing any interruption for users while typing [30].

Keystroke dynamics-based authentication operates continuously and transparently in the background, monitoring typing behavior without requiring active input from the user for authentication purposes. This allows for seamless integration into the user's normal interactions, enhancing convenience and reducing friction in the authentication process [30]. The individual typing patterns captured by keystroke dynamics are highly unique, making them effective for distinguishing between users. Features such as typing speed, key hold time, and key release intervals create a biometric signature that is difficult to replicate, thereby improving the security of the system [12]. Keystroke dynamics can be easily integrated into existing systems without the need for specialized hardware. Since most devices already have built-in keyboards, the implementation primarily requires software algorithms to capture and analyze the keystroke data, making it a cost-effective solution for enhancing security [30].

A key challenge with keystroke dynamics is the variability in typing patterns, which can be influenced by factors such as user mood, physical condition (e.g., fatigue or injury), and environmental conditions (e.g., different keyboards or devices). This variability can lead to false rejections, where the system fails to recognize the legitimate user [30]. Over time, a user's typing behavior may change due to factors such as learning effects, changes in typing habits, or even age-related motor skill alterations. This phenomenon, known as template aging, can degrade the performance of the authentication system, necessitating periodic retraining or updating of the user's keystroke profile to maintain accuracy [12]. While keystroke dynamics are unique, they are not entirely immune to mimicry attacks, where an attacker attempts to replicate the typing patterns of a legitimate user. Additionally, keystroke dynamics can be vulnerable to shoulder surfing, where an attacker observes the user's typing and attempts to replicate it, potentially compromising the system's security [30].

### 2.3 Implicit authentication based on sensors data

Various studies have demonstrated the versatility of using sensor data for implicit authentication, showing its ability to capture user behavior [31–33]. For instance, Lee et al. [31] showed the potential of using smartphone sensors for implicit authentication by achieving a good performance with a low false rejection rate (FRR) of 0.9% and a modest false acceptance rate (FAR) of 2.8%. This study highlights the efficacy of sensor-based biometrics in distinguishing genuine users from impostors. Shen et al. [32] conducted a comprehensive evaluation using sensor data and ten one-class detectors, resulting in an impressive equal error rate (EER) of 2.21%. As the field progresses, further exploration and refinement of sensor-based authentication methods promise to enhance both the security and user experience in the landscape of mobile device technology.

Sensor-based authentication leverages a wide array of data sources, such as accelerometer, gyroscope, and magnetometer, which capture detailed information about user movements, device orientation, and environmental context. This rich data enables the system to build a comprehensive profile of the user's behavior, enhancing its ability to accurately distinguish between legitimate users and impostors [31, 33]. One of the key advantages of sensor-based authentication is its ability to operate continuously in the background without requiring explicit actions from the user. This continuous monitoring allows the system to detect anomalies in real time, providing ongoing authentication that increases security without interrupting the user experience [32].

A significant challenge with sensor-based authentication is its susceptibility to environmental factors that can introduce noise or variability into the data. Changes in the user's environment, such as riding in a vehicle or walking on uneven terrain, can affect the sensor readings, potentially leading to increased false rejection or acceptance rates [31]. The collection and processing of sensor data for authentication purposes raise potential privacy concerns. Users may be wary of the extent to which their movements and behaviors are being monitored, even if the data is used solely for security purposes. Ensuring that data collection practices are transparent and handled securely is essential to mitigate these concerns [32].

In comparison with the previous work, most proposed implicit authentication methods are unimodal based [12, 15, 23], facing different challenges such as the intra-class variations and the noisy data [13]. Our proposed framework M$^2$auth introduces unique features and capabilities that works toward addressing these problems by integrating different modalities together. While acknowledging previous work achievements, we recognize the dynamic nature of the field and the continuous evolution of state-of-the-art methods. Our work support in enhancing multimodal behavioral biometric authentication, and we envision further improvements and refinements based on emerging technologies and methodologies. Additionally, our work contributes to the field by introducing a new fusion strategy that can be applied to other domains requiring secure and continuous authentication. We believe that these innovations provide a significant advancement over existing methods, and we have included a more detailed explanation of this in the revised manuscript.

## 3 Threat model

We presume the attacker has physical access to the victim's device. This unauthorized access could be gained by **insider attacks**, which come from someone who is related to the user such as family member or friend, or **stranger attacks**, which come from someone who does not know the user [34]. We do not consider unauthorized access that happened by malicious apps [35] or the apps that access data without the user intention [36].

The potential of an attacker is to obtain access to applications and private data or steal private information. We focus on detecting and preventing these types of attacks as described next by our developed framework M$^2$auth.

## 4 The proposed framework

M$^2$auth is a framework that encompasses three main modules as depicted in Fig. 1: the data collection module, the data analysis and feature extraction module for each authentication modality, and the decision fusion module. This framework aims to authenticate smartphone users by calculating authentication scores based on multiple modalities that extract relevant features from user interactions with the device. These interactions vary depending on the context, such as typing or network connectivity. M$^2$auth employs an ensemble method of classifiers to calculate authentication scores for each modality, and a decision fusion technique is utilized to integrate the scores from all modalities and formulate the final authentication decision.

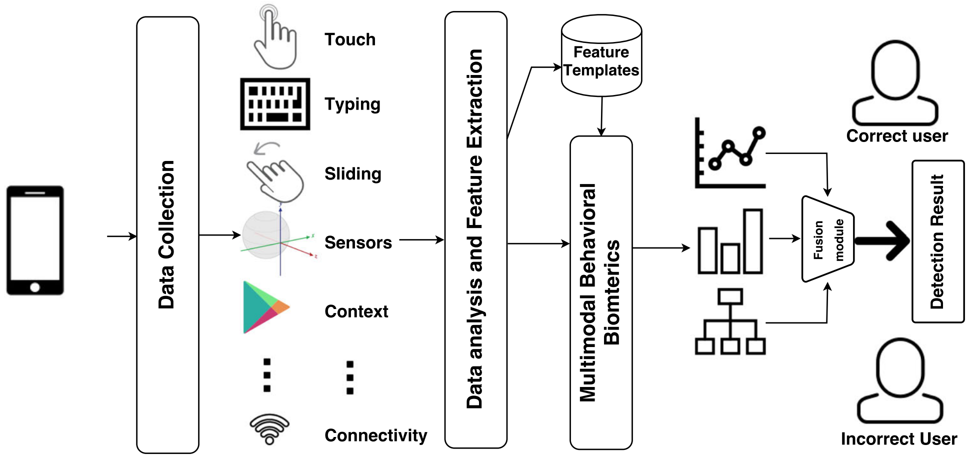


**Fig. 1** M$^2$auth behavioral biometric authenticate framework: It consists of three modules: data collection module that collects actual users' behavior; data analysis and feature extraction module that constructs a discriminative feature vector for each modality; and decision fusion module that combines all decisions from modalities to predict the final decision

## 4.1 Instrumentation

We instrumented the Android Open Source Project with a monitoring tool that records events relevant to touchscreen, keystrokes and profile sensors in a real-life settings without intervention. We used Phonelab test bed [37] to deploy our instrumentation and facilitate the data collection process. The test bed consists more than 200 LG Nexus 5 Android smartphone users who agreed to use custom build devices (i.e., Android 5.1.1). The instrumentation distributed transparently over the air to the smartphones. Participants have the ability to accept or deny the instrumentation at different times during the experiment period which was for two months. We collected three types of events: touch screen events, keystroke events, and sensory events. These events are going to reflect the user behavior, and consequently, it has the ability to differentiate between the owner and the imposter [14].

## 4.2 Data collection

For our study, we used the Phonelab test bed for data collection [37]. This test bed provided a robust platform for conducting experiments with smartphone users. To collect behavioral biometric data, we conducted a field study involving 133 participants, primarily students and staff from the University of Buffalo. The data collection spanned two months, ensuring a substantial period for capturing varied behavioral patterns. Before running the study, we obtained consent from all participants, emphasizing privacy and following the ethical guidelines approved by our university's research ethics board.

Data was gathered continuously as participants interacted with their smartphones naturally, without any specific instructions or interventions. This approach ensured the collection of authentic usage data. Throughout the study, data was logged from three primary sources: touch screen interactions, virtual keyboard inputs, and inertial measurement units (IMUs) encompassing accelerometer, gyroscope, and magnetometer sensors.

In order to ensure the quality and reliability of the data, we conducted data cleaning procedures and excluded participants who were not actively using their smartphones throughout the entire experiment period. Our final analysis focused on 52 participants who consistently utilized their phones for at least 30 days or more. It is worth noting that each participant had the freedom to use their smartphone without any intervention, allowing the data to be logged and subsequently uploaded to a backend server.

As shown in Table 1, we collected approximately 3 million interaction points from touch gestures, 1 million from keystrokes, and 1.5 million from sensors, totaling around 5.5 million action points. Each participant contributed over 3500 action points on average per day.

The large volume of high-quality data collected without any intervention further enhances the value of the formulated dataset. The substantial amount of data enables a more comprehensive and detailed analysis, capturing a wide range of user behaviors and interactions. With a large dataset, we can uncover intricate patterns, detect rare events, and gain deeper insights into user behavior. The extensive data also facilitates the training of machine learning models and algorithms, improving their accuracy and performance.

**Table 1** Number of user interactions in the dataset

| Modality | # of interaction points |
|---|---|
| Touch gesture | 2969975 |
| Keystrokes | 1004735 |
| Sensor | 1560096 |

## 4.3 Data preprocessing

Data preprocessing is a fundamental step in the development of a reliable and accurate multimodal biometric system. It involves cleaning and organizing raw data to ensure that it is in the best possible condition for feature extraction and analysis. Given the diversity and complexity of data sources such as touch gestures, keystrokes, and sensor readings, effective preprocessing is crucial for mitigating noise, handling variability, and ensuring consistency across different data modalities.

Touch gesture data is prone to noise from unintended screen interactions or device sensitivities. To address noise and variability in touch gesture data, we implemented a set of filtering criteria aimed at cleaning the raw data. These criteria include removing duplicated touch events to eliminate redundancy and ensure that the data accurately represents the user's intentional interactions. By focusing on these preprocessing strategies, we enhance the robustness of the system, enabling more accurate feature extraction and reliable biometric authentication. This approach ensures that the processed data is a true reflection of user behavior, reducing the potential for errors during the authentication process.

Keystroke data often contains irregularities such as accidental key presses or variations in typing speed. These anomalies can distort the analysis if not properly managed. To address this, we use outlier detection methods to identify and remove such irregularities, resulting in a dataset that more accurately represents the user's typical typing behavior. This step is essential for ensuring that the extracted features are both relevant and reliable.

Processing raw sensor data involves reducing noise and enhancing the accuracy of the data to improve the authentication model's performance. We employ a forward–backward digital filtering technique, which effectively removes noise and eliminates unwanted gravitational forces that might interfere with the analysis. After filtering, the sensor signals are segmented into smaller, predefined time windows. This segmentation step is critical for organizing the data and ensuring that the subsequent feature extraction phase operates on consistent and meaningful chunks of sensor readings. By dividing the data into time windows, we capture the temporal dynamics and patterns within specific intervals, enabling more accurate and contextually relevant feature extraction.

Given the inherent variability in data collected from different devices and users, normalization is essential. We standardize data across all modalities, touch gestures, keystrokes, and sensors, to ensure consistency. This involves adjusting the data to a common scale, making it comparable across modalities, and enhancing the effectiveness of subsequent feature extraction and fusion. Furthermore, to ensure that all features contribute equally to the analysis, we apply feature scaling techniques such as min–max scaling or standardization. This alignment is particularly important when combining data from different modalities, where the value ranges can vary significantly, facilitating more effective integration and classification.

To capture broader behavioral patterns and reduce the impact of transient noise or anomalies, we aggregate data over specified time windows or user sessions. This approach helps smooth out short-term fluctuations and provides a more comprehensive view of the user's behavior. Aggregated data is inherently more robust, making it better suited for accurate feature extraction and improving the overall reliability of the biometric system.

Effective data preprocessing is vital for enhancing the performance of a multimodal biometric system. By cleaning, normalizing, and organizing the data from various sources, we reduce the likelihood of errors during feature extraction and classification. This process ensures that the data accurately reflects genuine user behavior, leading to more reliable and consistent biometric authentication. Preprocessed data forms the foundation for building robust models that perform well in real-world scenarios, ultimately contributing to the system's overall effectiveness and user trust.

## 4.4 Feature extraction

### 4.4.1 Gesture features

Touch gesture in M2auth refers to a hand-drawn shape created by the user's fingers on a touch screen. It comprises a sequence of touch points, where each point is defined by its $x$ and $y$ coordinates. Figure 2 illustrates examples of touch gestures. During data collection, the following raw data is recorded for each touchpoint: timestamp (indicating the time of the touch), coordinates (the $x$ and $y$ positions on the touch screen), pressure (the amount of pressure applied), size (the size of the touch area), touch_down (a flag indicating the beginning of a touch), touch_up (a flag indicating the end of a touch), touch_move (a flag indicating movement during the touch), and action_code (a code representing the touch state). These raw data elements provide detailed information about the user's touch interactions, enabling subsequent analysis and authentication within the M2auth framework.

In M2auth, the action_codes are utilized to identify and delineate a stroke, denoted as **S**. A stroke is characterized by a sequence of touch points, represented by their corresponding $x_i$ and $y_i$ coordinates. It begins with the touch_down action and concludes with the touch_up action. Each touchpoint in the stroke is associated with additional

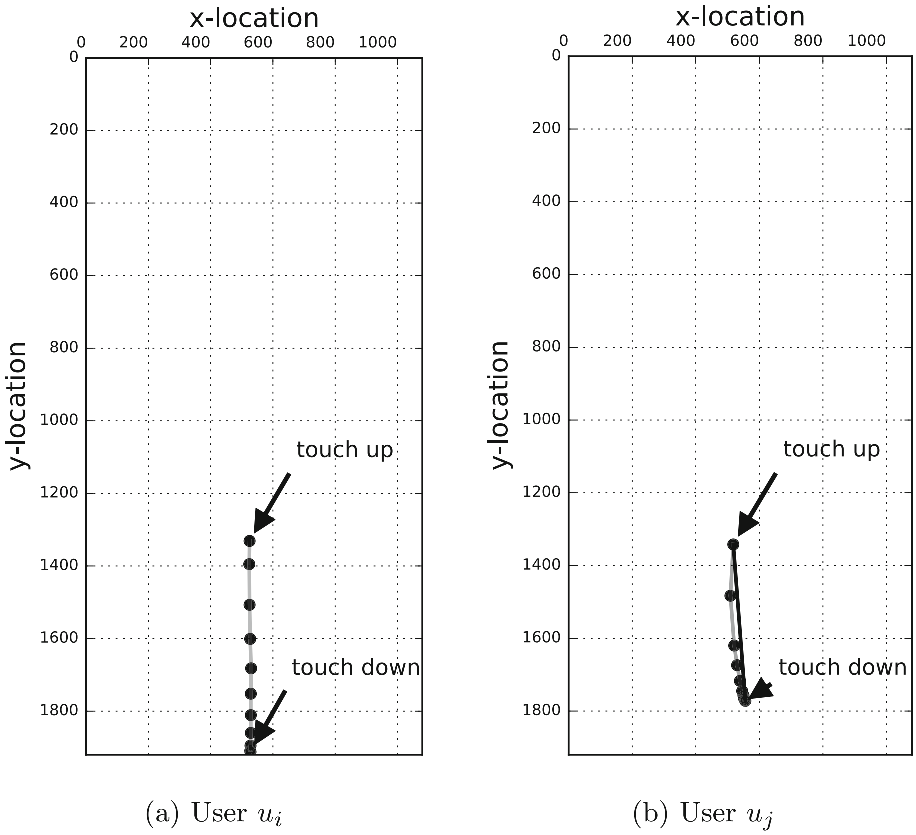


**Fig. 2** Different stroke samples from two different users. Both strokes were from down to up

information: $p_i$ denotes the touch pressure, $a_i$ represents the size area of the touch, and $t_i$ indicates the timestamp. The index $i = 1, \ldots, n$, where $n$ signifies the total number of touch points within the stroke $S$. These attributes provide essential details about the touch interactions, enabling subsequent analysis and authentication processes within the M2auth framework.

In order to extract features from strokes, we employed both geometric and motion dynamics analysis techniques.

For geometric analysis, we extracted six features that focus on stroke geometry. Four features were derived from touch_down and touch_up locations, namely $x_{\text{down}}, y_{\text{down}}, x_{\text{up}}, y_{\text{up}}$. These represent the coordinates of the initial touch and the final release points. Additionally, two features, stroke length $S_{\text{length}}$ and the curvature $S_{\text{curvature}}$, were calculated. $S_{\text{length}}$ measures the length of the stroke and is calculated by summing the Euclidean distances between consecutive touch points, following the formula:

$$S_{\text{length}} = \sum_{i=2}^{n} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{1}$$

where $n$ represents the number of points in the stroke $S$. Stroke curvature $S_{\text{curvature}}$ quantifies the deviation of the stroke from a straight line. It is determined by calculating the curvature for each touchpoint in the stroke, as shown in Fig. 2b. The curvature ($K$) is computed using the formula [38]:

$$K = \frac{|X'Y'' - Y'X''|}{(X'^2 + Y'^2)^{\frac{3}{2}}} \tag{2}$$

where $X$ and $Y$ represent the vectors of the touchpoints in the stroke. Their primes refer to the first and second order of derivatives. $K$ is the curvature row vector for each touchpoint in the stroke. Then we calculate the $S_{\text{curvature}}$ by taking the mean of the $K$ as follows:

$$S_{\text{curvature}} = \frac{1}{N} \sum_{i=1}^{N} K_i \tag{3}$$

These geometric features provide insights into the shape, length, and curvature of the stroke, contributing to the multimodal behavioral biometric authentication process in M2auth.

In the dynamic analysis of strokes, we focused on the motion dynamics and extracted four features. As the finger moves on the touchscreen, the stroke is formed by a sequence of touchpoints. We detected these touchpoints based on the finger's motion, which is often curvilinear rather than linear. Consequently, we calculated the displacement of the stroke, represented by the straight line length between the touch_down and touch_up points, as shown in Fig. 2b. Additionally, we computed the velocity of the stroke at each touchpoint using the following equation:

$$V = \sqrt{((X')^2 + (Y')^2)} \tag{4}$$

where $X$ and $Y$ are sets of touchpoints in the stroke and $V$ represents the velocities vector at each touchpoint. From these velocity values, we extracted two features: the mean velocity $S_{\text{mean(V)}}$ and the maximum velocity $S_{\max(\text{V})}$. Furthermore, we analyzed the acceleration of the stroke using the following equation [38]:

$$A = \frac{d^2 s}{dt^2} T + k\left(\frac{ds}{dt}\right)^2 N \tag{5}$$

Here, $s$ represents the stroke, $d$ denotes the derivatives with respect to the time, and $T$ and $N$ represent the tangent and normal vectors, respectively.

By examining the dynamic aspects of the stroke, including displacement, velocity, and acceleration, we extracted features that capture the motion characteristics. These features provide valuable information about the dynamic behavior of the stroke, contributing to the multimodal behavioral biometric authentication process in M2auth.

In addition to the geometric and dynamic features, M2auth also includes temporal features and pressure/size features for stroke analysis.

The temporal features capture the timing aspects of the stroke. The duration of a stroke, denoted as $S_{\text{duration}}$, represents the time spent by the user's finger to perform the stroke. This feature provides insights into the speed or deliberate nature of the gesture. Additionally, the interstroke duration, $S_{\text{interduration}}$, measures the time duration between the previous stroke and the current one. This temporal information can be useful in understanding the rhythm or pattern of user interactions.

The pressure and size features focus on the touch pressure and size associated with each touchpoint in the stroke. For every touchpoint, we extract ten features, including touch_down and touch_up values. These features encompass the average, maximum, and minimum values of touch pressure and size. By analyzing the variations in pressure and size throughout the stroke, we can gain insights into the user's touch behavior and potentially identify unique patterns or characteristics.

Together, the temporal features and pressure/size features provide additional dimensions for characterizing strokes in M2auth. By considering these aspects, the framework can capture a more comprehensive representation of user behavior and enhance the accuracy of authentication.

#### 4.4.2 Keystroke features

In the context of keystroke analysis, different applications require users to interact daily by entering inputs via a keyboard, such as messaging apps. Each user's input is characterized by a specific typing rhythm, which describes their unique way of typing. A keystroke session, illustrated in Fig. 3, represents the sequence of keystrokes starting from the key press $p_1$ and ending with the last key release event $r_n$ before the soft keyboard disappears. Each press and release event represents a keystroke $k$, and multiple keystroke sessions were collected while users interacted with various apps.

The feature extraction process for keystrokes involves analyzing their geometry, dynamics, spatial characteristics, temporal properties, size, and pressure. These features are divided into two groups: one for individual keystrokes ($k_i$) and the other for the keystroke session ($S_{\text{keystrokes}}$), where the session features rely on the individual keystroke features.

From spatial analysis, we extract four features based on the coordinates of the press ($p_i$) and release ($r_i$) events: $x_{\text{press}}$, $y_{\text{press}}$, $x_{\text{release}}$, and $y_{\text{release}}$. Although there is no dynamic movement within a single keystroke, we extract the displacement ($k_{\text{displacement}}$) as a feature, representing the distance between the previous and current keystrokes. Temporal analysis provides two features: the duration ($d_i$), calculated as the time difference between the press and release of the current key ($r_i - p_i$), and the latency ($l_i$), calculated as the time difference between the press of the current key and the release of the previous key ($p_i - r_{(i-1)}$) as shown in Fig. 3. Additionally, we extract the average size and average pressure of the keystroke as features.

The keystroke session, which consists of a sequence of keystrokes, enables the extraction of temporal features. The duration ($D$) represents the total time between the first key press ($p_1$) and the last key release ($r_n$). The average duration per keystroke ($d_{\text{average}}$) provides insights into the typical duration of individual keystrokes within the session. The average latency ($l_{\text{average}}$) represents the average time interval between keystrokes in the session. These temporal features capture the timing and rhythm of the user's typing behavior during a keystroke session.

#### 4.4.3 Sensor features

In sensor modality, we authenticate smartphone users based on a discriminative set of features that execrated from off-the-shelf motion sensors on the smartphone. These features helped us to build a user behavior profile that reflects the user interactions with the mobile sensors and services [39]. We use three different sensors to make

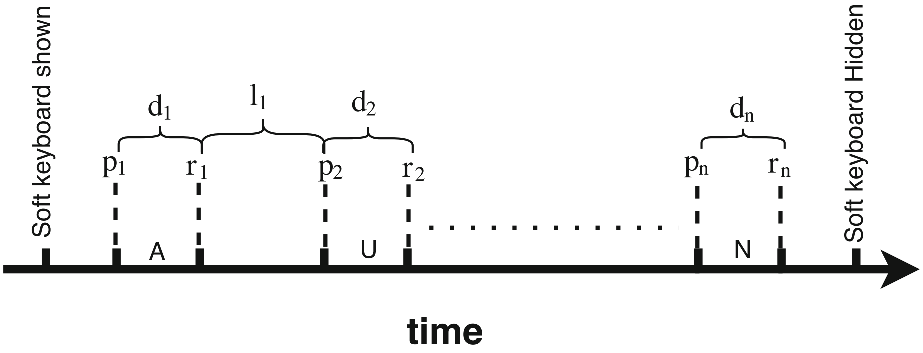


**Fig. 3** Keystroke session events: during the Soft keyboard shown event, the virtual keyboard pop up to enable users enter characters. The $p_1$ represents the first press, $r_1$ represents the first release, $d_1$ represents the consumed time between press and release, $l_1$ represents the latency, the time between the release and the press of the previous and the next key, respectively

authentication: accelerometer, gyroscope, and magnetometer [26]. During the filed study, we collected sensor events of the instantaneous reading of sensors' axes, x, y, and z for every subtle action that user did on the smartphone.

For feature extraction, two types of analysis are conducted: time-domain analysis and frequency-domain analysis. In the time-domain analysis, statistical features are extracted from the raw sensor data over time. These features provide valuable information about the distribution and characteristics of the sensor readings. As illustrated in Table 2, the feature vector includes statistical measures such as mean, median, maximum, minimum, standard deviation, variance, kurtosis, skewness for each of the three axes ($x$, $y$, $z$), as well as the magnitude. These statistical features effectively capture the identity of the users based on their sensor interactions [26].

In the frequency-domain analysis, the goal is to capture the intensity of user activities on the smartphone. This is achieved by applying the fast Fourier transform (FFT) to the sensory data, transforming it from the time domain to the frequency domain. From the frequency domain, energy and entropy features are extracted. The energy feature reflects the distribution of signal energy across different frequency components, indicating the intensity or strength of user actions. The entropy feature provides insights into the complexity or randomness of the sensor data in the frequency domain. These frequency-domain features are particularly useful in capturing the fine-grained movements and actions performed by the user, such as tapping or swiping on the phone [40].

By combining both time-domain and frequency-domain analysis, a comprehensive feature set is obtained, capturing various aspects of the user's sensor interactions. These features significantly contribute to identifying and authenticating users based on their unique behavioral patterns and movements recorded by the sensors.

### 4.5 Fusion and decision module

In the fusion and decision module, using only a single source of information to authenticate the user presents several limitations and problems [13]. These include noisy data affected by impreciseness in measurements, non-universality, incorrect interaction with sensors, intraclass variations, changes in behavioral characteristics at different time instances, lack of uniqueness to differentiate between two users, and vulnerabilities such as spoofing and robot attacks.

The goal is to differentiate between legitimate user and impostors based on information acquired from multiple behavioral biometric traits. Fusing these different traits to perform authentication offers several advantages in overcoming the aforementioned limitations. This includes improving the performance and reliability of the authentication process, increasing protection against attacks, as it becomes harder for impostors to spoof multimodal biometric traits compared to unimodal ones, enhancing data availability, and reducing data ambiguity (Fig. 4).

**Table 2** Extracted time- and frequency-domain features

| Domain | Features |
|---|---|
| Time domain | Max, Min, Mean, Median, Standard deviation, Variance, Mean crossing rate Jitter, Skewness, Kurtosis |
| Frequency domain | Energy, Entropy |

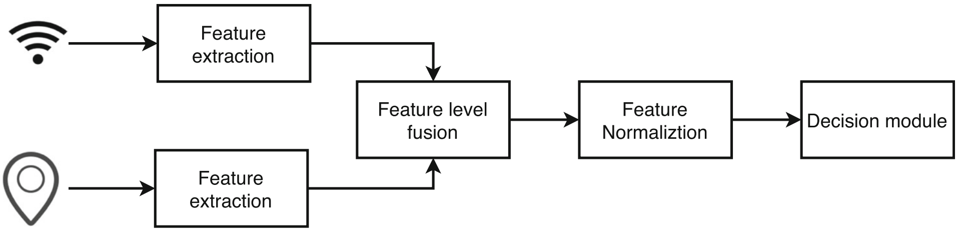


**Fig. 4** Feature-level fusion on behavioral profiling traits

### 4.5.1 Fusion levels

Different fusion levels can be applied to the behavioral biometric framework modules to enhance the robustness and accuracy of authentication systems. These fusion levels include the followings [41, 42]:

- Sensor-level fusion: involves combining the raw data from different sensors for the same biometric traits. This approach integrates the initial data streams, allowing for a more comprehensive capture of the biometric signals before any feature extraction occurs.
- Feature-level fusion: merges different feature vectors extracted from multiple biometric modalities into a single new feature vector. By integrating features at this level, the system can utilize a richer set of characteristics derived from the raw sensor data, enhancing the discriminatory power of the authentication process.
- Score-level fusion: applies the combination based on the matching scores generated by each authentication modality. Each biometric modality produces a score indicating the likelihood of a match, and these scores are then combined to improve the overall authentication accuracy.
- Decision-level fusion: involves combining the decisions made by multiple classifiers to make the final authentication decision. This method aggregates the outputs of various classifiers, leveraging the strengths of each to achieve a more reliable and robust final decision.

The first and second levels of fusion classified as fusion before matching, but the third and fourth levels classified as fusion after matching. The main distinction between them is that the fusion after matching has abstract summary about the input pattern and easier to access and combine than the fusion before matching. Consequently, after matching fusion is the most common used fusion in multimodal biometric systems [41]. By employing these different fusion levels, a behavioral biometric framework can significantly enhance its performance, leveraging the strengths of various data integration strategies to improve both security and user experience.

### 4.5.2 Feature-level fusion

In our multimodal biometric framework, we implemented feature-level fusion to enhance the accuracy and robustness of the system. Feature-level fusion involves combining feature vectors extracted from different biometric modalities into a single, comprehensive feature vector. This approach is more effective than matching score or decision-level fusion because it retains richer information from the input biometric data [43].

In our study, we extracted three distinct feature vectors from each behavioral profile, specifically from gesture features, keystroke features, and sensor features. To leverage the complementary information provided by these different modalities, we employed multiple fusion strategies. We conducted fusion in three binary combinations, where we combined pairs of feature vectors (e.g., gesture with keystroke, gesture with sensor, and keystroke with sensor). Additionally, we performed a trinary fusion, which involved combining all three feature vectors (gesture, keystroke, and sensor) into a single, unified feature vector.

Let $\mathbf{F_1}, \mathbf{F_2}$ and $\mathbf{F_3}$ represent the feature vectors extracted from different biometric modalities, such as gesture features, keystroke features, and sensor features, respectively. These feature vectors can be defined as:

$$\mathbf{F}_1 = [f_{11} \quad f_{12} \quad \ldots \quad f_{1p}]^T, \quad \mathbf{F}_2 = [f_{21} \quad f_{22} \quad \ldots \quad f_{2q}]^T, \\ \mathbf{F}_3 = [f_{31} \quad f_{32} \quad \ldots \quad f_{3r}]^T \tag{6}$$

where $p$, $q$, and $r$ are the dimensions of the respective feature vectors. One of the simplest and most common methods of feature-level fusion is linear vector concatenation [44]. In this approach, the feature vectors from different modalities are concatenated to form a single, unified feature vector:

1. Binary Combination:
   - Gesture and keystroke feature fusion:

$$\mathbf{F}_{12} = \left[\mathbf{F}_1^T \quad \mathbf{F}_2^T\right]^T = [f_{11} \quad f_{12} \quad \ldots \quad f_{1p} \quad f_{21} \quad f_{22} \quad \ldots \quad f_{2q}]^T \tag{7}$$

   - Gesture and sensor feature fusion:

$$\mathbf{F}_{13} = \left[\mathbf{F}_1^T \quad \mathbf{F}_3^T\right]^T = [f_{11} \quad f_{12} \quad \ldots \quad f_{1p} \quad f_{31} \quad f_{32} \quad \ldots \quad f_{3r}]^T \tag{8}$$

- Keystroke and sensor feature fusion:

$$\mathbf{F}_{23} = \left[\mathbf{F}_2^T \quad \mathbf{F}_3^T\right]^T = [f_{21} \quad f_{22} \quad \ldots \quad f_{2q} \quad f_{31} \quad f_{32} \quad \ldots \quad f_{3r}]^T \tag{9}$$

2. Trinary Combination:

$$\mathbf{F}_{123} = \left[\mathbf{F}_1^T \quad \mathbf{F}_2^T \quad \mathbf{F}_3^T\right]^T = \begin{bmatrix} f_{11} \\ f_{12} \\ \ldots \\ f_{1p} \\ f_{21} \\ f_{22} \\ \ldots \\ f_{2q} \\ f_{31} \\ f_{32} \\ \ldots \\ f_{3r} \end{bmatrix} \tag{10}$$

In this formulation, $\mathbf{F}_{12}, \mathbf{F}_{13}, \mathbf{F}_{23}$ and $\mathbf{F}_{123}$ are the fused feature vectors that combine the information from the selected modalities. The fused vector retains the comprehensive information from all contributing modalities, providing a more discriminative representation of the biometric traits.

This method of feature fusion aims to create a more discriminative representation of the user's behavior by integrating diverse data sources. By addressing challenges such as data diversity and feature incompatibility, and by applying normalization techniques, we ensured that the fused feature vectors were consistent and ready for further processing. The resulting fused vectors were then used to improve the detection accuracy of our multimodal biometric system, supporting real-time authentication and robust user recognition.

### 4.5.3 Decision-level fusion

In our multimodal biometric system, decision-level fusion plays a critical role in enhancing the accuracy and reliability of the authentication process. Unlike feature-level fusion, where features are combined before classification, decision-level fusion involves aggregating the decisions made by classifiers for each biometric modality. This approach leverages the strengths of each modality while mitigating the weaknesses that may arise from relying on a single source of biometric data.

As illustrated in Fig. 5, in this scenario, we combine the local decisions obtained from each modality, such as gesture, keystroke, and sensor features. The combination is performed using the majority voting method. We use this method because (i) it is the simplest combination method to implement, and (ii) it is effective as more complicated schemes perform to improving the results.

The majority voting method operates by evaluating the local decisions $d_{ij}$ generated by each modality's classifier. Each classifier outputs a decision that either accepts or rejects the authentication attempt. The final decision $D$ is determined by taking the plurality vote of these local decisions [45]:

$$D = *\max_j^C \sum_{i=1}^{L} d_{ij} \tag{11}$$

where $j$ represents the class dimension, typically between 0 and 1 (i.e., accept or reject) and $i$ represents the number of modalities, with $L$ being the total number of modalities involved.

Our goal of combining different modalities is to improve the detection accuracy, given that every modality has a unique mechanism to make a decision. One modality could fail to classify a sample but others not. By observing the complementarity of decisions across modalities, we are able to achieve a more reliable and accurate overall decision [45, 46].

Instead of relying on a single modality for authentication, decision-level fusion ensures that the system benefits from the combined expertise of multiple classifiers. Each classifier is cross-validated and trained independently, and the final authentication decision is made by applying the majority voting method to the local decisions. This approach significantly boosts the performance of the authentication process, providing a robust mechanism for user verification.

## 5 M$^2$auth evaluation results

Evaluating the multimodal behavioral biometric framework is a challenging task, due to (i) the diversity of performance measures under various conditions [41], and (ii) the lack of performance evaluations standards in the literature, where the majority of the existing standards are established to evaluate single traditional biometric modality [47, 48]. We perform the evaluation of M$^2$auth based on a matching performance framework.

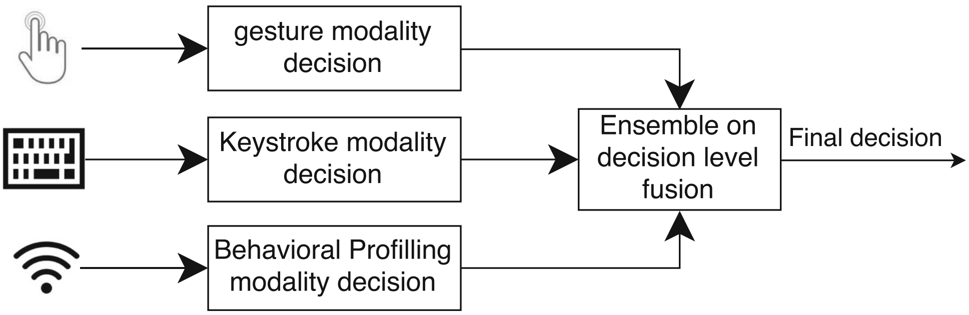

gesture modality decision
Keystroke modality decision
Behavioral Profilling modality decision
Ensemble on decision level fusion
Final decision

**Fig. 5** The fusion was applied on the decision level on multiple biometrics traits (i.e., we have three biometric traits, gesture, keystrokes and behavioral profile)

## 5.1 Dataset

We formulated a dataset from the collected data of M$^2$auth as described in Sect. 4.2. The dataset contains three different feature vectors extracted from gesture, keystroke, and accelerometer as described in Sects. 4.4.1, 4.4.2 and 4.4.3. The dataset contains more than 5.5M events collected in the wild from 52 participants who used the phones for 30 days or more in unconstrained environment as shown in Table 1. Participants used the smartphones without any intervention; consequently, all collected events reflect the behavior of using the smartphone in a real-world manner.

## 5.2 Matching model

Any biometric systems can use two types of matching: verification and identification [41, 48]. Both are different and depend on the context of application that the biometric system will operate on. Consequently, the evaluation is different based on the chosen one.

In the verification process, which is a one-to-one matching scenario, the model validates the claimed biometric identity by comparing it with the stored template, where the matching algorithm either accepts or rejects the claimed identity using a matching score and a predefined threshold.

In the identification process, which involves one-to-many matching, the model recognizes the presented biometric identity by comparing it with all stored templates for each user. The matching algorithm estimates the identity of the sample by selecting the highest match score from the multiple matching scores generated, and this decision is made based on a predefined threshold.

In this study, we assessed the M$^2$auth framework in a single-user context, where each mobile device is exclusively used by one user. Our primary objective is to thwart unauthorized access by distinguishing between the legitimate owner and potential impostors, effectively framing this as a binary-class classification problem. As such, our evaluation centers on the matching system within the verification mode.

## 5.3 Classification model

Verification is essentially a binary classification task. For every user, denoted as $u_i$, the classifier computes an authentication score, denoted as $p(u_i)$, which falls within the range of (0, 1). This score reflects the likelihood of $u_i$ being a valid user, and it is determined in relation to a predefined threshold, $\alpha$, also within the range (0, 1). To execute this classification, we employed a one-vs-all scheme, where data from other users was utilized to represent impostors in the evaluation.

A random forest (RF) classifier [49] was used, and it is one of the best performed classifiers in the literature of biometric authentication [25]. RF operates through an ensemble classification approach, wherein it fits decision tree classifiers on various subsamples of the dataset. Subsequently, it aggregates the predictions from these subclassifiers through averaging, thereby enhancing accuracy and mitigating overfitting. The selection of RF was based on its superior performance relative to other classifiers in the existing literature.

## 5.4 Validation model

For classifier evaluation, we partitioned the dataset into training set and testing set (i.e., unseen data). Then we performed tenfold cross-validation where the data is divided into 10 subsets. One subset is used as a test set, and the other 9 subsets are used as a training set. This procedure is repeated 10 times. Then the mean error across all folds is calculated. This was used with the grid search to determine the best hyper-parameters. Table 3 shows the search space for random forest hyper-parameters in addition to the

**Table 3** Search Space for hyper-parameters

| Parameter | Search space | Optimal value |
|---|---|---|
| # of estimators | 10, 50, 100, 200 | 200 |
| Tree depth | 2, 4, 5, 6, 7, 8 | 8 |
| # of features | sqrt, log2 | sqrt |

determined optimal value of each parameter. These values were used to set up the model for testing.

## 5.5 Performance metrics

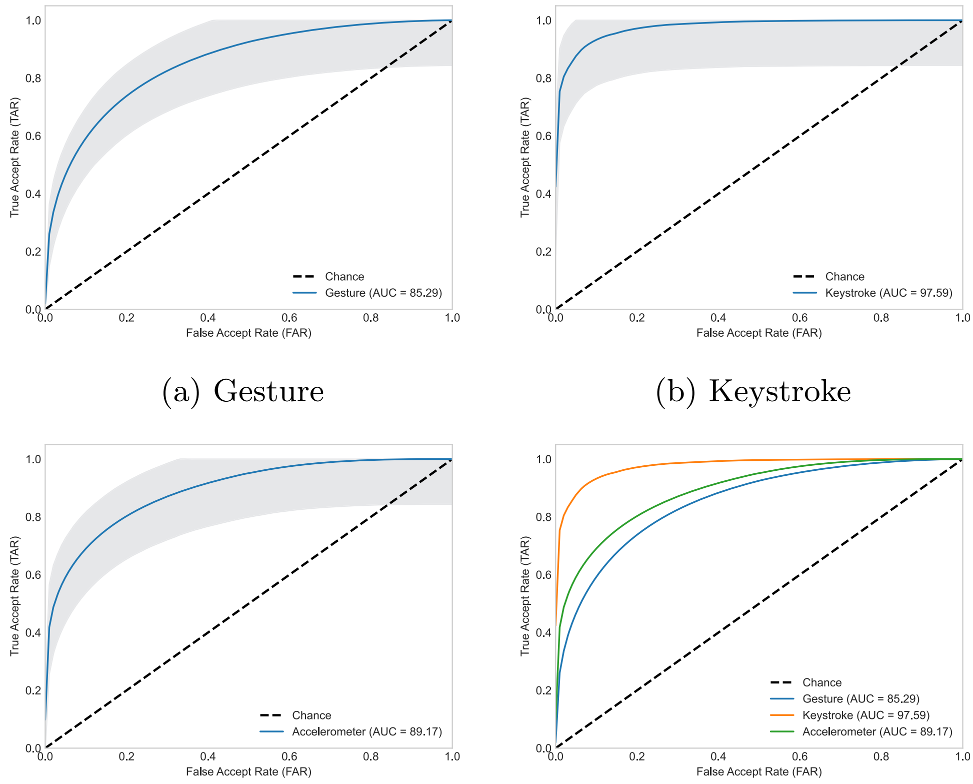
To evaluate the behavioral biometric performance, we used receiver operating characteristic (ROC) curve. It shows the trade-off between the true accept rate (TAR) and the false accept rate (FAR) at various threshold. Also, we used the area under the curve (AUC) to measure the quality of the model as an alternative to the accuracy, which is useful even when there is an imbalanced dataset (i.e., one of the classes dominates). When the TAR equal to one and FAR equal to zero, the performance of the model that measured by AUC will equal to 100% and the plot will hit the top left corner.

## 5.6 Experimental results

We present the performance of each biometric modality individually followed by the results of the fusion scenarios on feature level and decision level.

### 5.6.1 Biometric modalities

In our experiment, we measured the performance of each authentication modality using the AUC and EER metrics. We cross-validated and trained a RF classifier for each participant. The experimental results were conducted on a test set and averaged over all participants. Figure 6 shows ROC curves for each authentication modality. The best performed one is keystroke modality as shown in Fig. 6b, and its AUC reached 97.59% with an EER of 8.21%. This indicates that the typing rhythm scenario could be considered as a strong behavioral biometric traits for authenticating smartphone users. The extracted feature vector of this modality contains highly discriminative features that can differentiate between two identities. The flooding of accelerometer sensory data that were used to train a RF classifier achieved a reasonable accuracy with an AUC of 89.17% and EER of 19.83% as shown in Fig. 6d. Given that these data were collected in the background without any explicit interaction with the user, the smallest performed one was the gesture modality, where the result of the developed classifier that was trained on feature set extracted from the touch gesture was 85.29% and 23.07% for AUC and EER, respectively, as shown in Fig. 6a. Table 4 presents a summary of all results.



(a) Gesture

(b) Keystroke

(c) Accelerometer

(d) All modalities

**Fig. 6** Performance of authentication modalities. ROC Analysis plots TAR against FAR over different thresholds using RF classifier for gesture, keystroke, and accelerometer

**Table 4** Results for all modalities using RF classifier

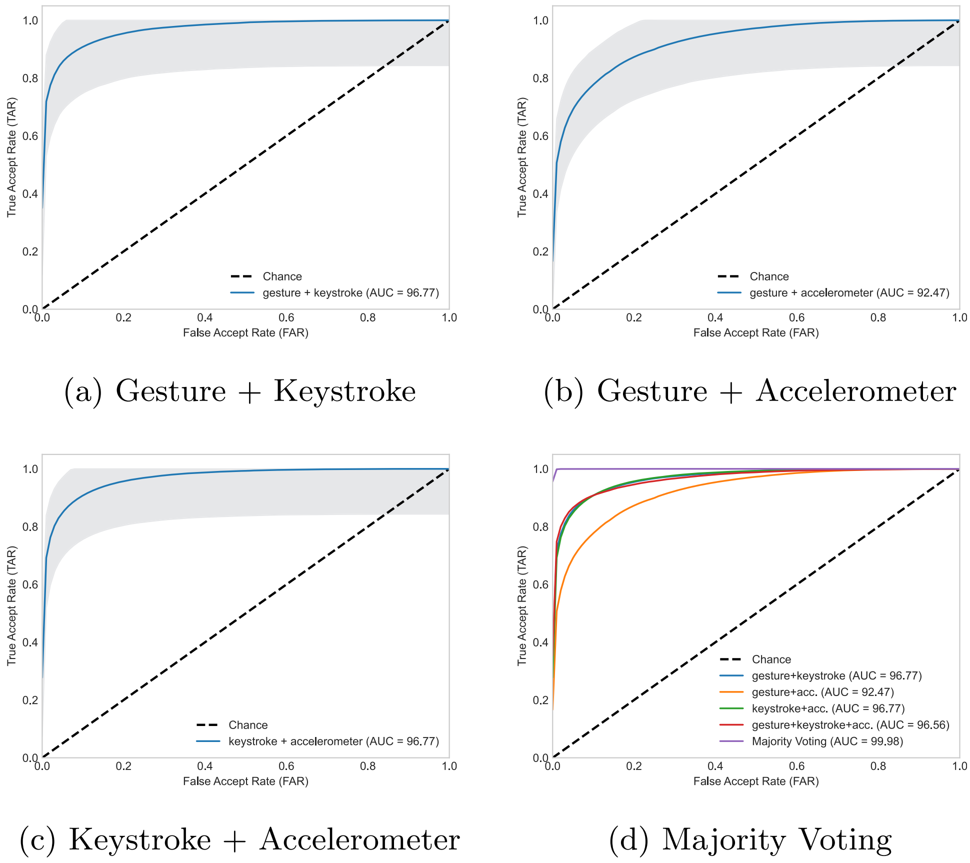
| | Gesture | Keystrokes | Accelerometer |
|---|---|---|---|
| AUC (%) | 85.29 | 97.59 | 89.17 |
| EER (%) | 23.07 | 8.21 | 19.83 |

### 5.6.2 Feature-level fusion results

We extracted three feature vectors from each behavioral profile, as shown in Sects. 4.4.1, 4.4.2, and 4.4.3. We fused these vectors together in three binary combination, in addition to a trinary one. The first combination was done by adding the gesture features with the keystroke features together, let us call it (G + K) and it contained 41 features. Cross-validation and training were conducted on an RF classifier. The evaluation was done on unseen data, the test set, and the performance of the classifier was 96.77% and 9.51% for AUC and EER, respectively, as shown in Fig. 7a. This combination has improved the performance compared to the gesture unimodal one, whose AUC was 85.29%, and EER was 23.07%. On the other hand, if we compare this combination with the gesture and keystroke, its performance was slightly smaller than the performance of the keystroke unimodal alone, as shown in Table 4. This means that adding the gesture features penalized the stroke features and downgraded its performance from 97.59 to 96.77% as shown in Figs. 6b and 7a, respectively.

The second combination was conducted by adding the features of gesture and accelerometer modalities (G + A). The combined feature vector contained 70 features, and an RF classifier was cross-validated and trained. The evaluation result of this combination was 92.47% and 15.86% for AUC and EER, respectively. Even though the performance of (G + A) is worse than the performance of (G + K), its performance is better than the performance of each individual modality alone, where the AUC of gesture was 85.29% and AUC of accelerometer was 89.17%. This means the fusion of (G + A) has improved the performance by 7.18% compared to the unimodal gesture modality and by 3.3% compared to the unimodal accelerometer modality.

The last binary combination was between the keystroke features and the accelerometer features (K + A). This fusion formulated a feature vector of 67 features and the performance result of this fusion is similar to the performance of (G + K), which is 96.77% and 9.58% for AUC and EER, respectively. Hence, we can see that adding the keystroke features has improved the accelerometer



(a) Gesture + Keystroke

(b) Gesture + Accelerometer

(c) Keystroke + Accelerometer

(d) Majority Voting

**Fig. 7** Performance of authentication modalities. ROC Analysis plots TAR against FAR over different thresholds using RF classifier for gesture, keystroke, accelerometer, and majority voting

**Table 5** Results for fusion modalities using RF classifier

| | Feature fusion | | | | Decision fusion |
|---|---|---|---|---|---|
| | G + K | G + A | K + A | G + K + A | Majority |
| AUC (%) | 96.77 | 92.47 | 96.77 | 96.56 | 99.98 |
| EER (%) | 9.51 | 15.86 | 9.58 | 9.48 | 0.84 |

modality performance by 7.6%. Fusing all features of all modalities together helped us to formulate a feature vector of 89 features. Similarly, we cross-validated and trained an RF classifier; then, we evaluated this classifier on a combined test set of unseen data. The performance of this combination (G + K + A) was 96.56% and 9.48% for AUC and EER, which is similar to all other combination that contained the keystroke features. Tables 4 and 5 show a summary of individual and fused results.

#### 5.6.3 Decision-level fusion results

Instead of relying on only one modality to perform the authentication process, we use the decision fusion scenario as described in Sect. 4.5.3 to make the authentication. So, the classifier of each modality is cross-validated and trained. A decision fusion scenario is applied to boost the performance of the authentication process. The decision of the authentication was conducted by applying the majority voting method on the local decision of each modality. This fusion scenario increased the performance of the authentication to reach 99.98% for AUC with 0.84% EER. Figure 7 shows the fusion results for decision level and feature level, as you can see the magenta line that represents the performance of the decision fusion which outperform all other feature fusion scenarios. In comparison with the unimodal ones, the decision fusion performance has increased the results by 14.69%, 10.81%, and 2.39% for gesture, accelerometer and keystroke modalities, respectively. This perfect result was achieved based on the hypothesis, *if one modality failed to classify a sample, others may not*. We leveraged the notion of *complementarity* to lead us to this perfect decision.

## 6 Discussion

M$^2$auth provides a high accurate detection rate to differentiate between legitimate user and imposters as shown in the previous section. In this section, we discuss different aspects of the framework, in terms of datasets, extracted features, and comparison with other systems.

### 6.1 Real-world datasets

The majority of the state-of-the-art (e.g., 89% of experiments reported by Teh et al. [50]) behavioral biometric systems that achieve very low error rates are tested on samples acquired under controlled conditions with cooperative users, which tends to be far from the real-world scenarios [50, 51]. Some of these datasets were conducted in a laboratory study [52, 53], and others used some specific applications [23, 54–56]. Our dataset was collected in unconstrained environment without any form of intervention. Remarkably, we did not confine the data collection to specific applications; instead, we instrumented the Android OS, granting users complete freedom to install and customize their devices as they saw fit. This approach ensured that our dataset encapsulates authentic user behavior across a diverse range of application contexts, rendering it highly realistic and representative of real-world usage scenarios.

### 6.2 The influence of feature vector

Feature engineering is a challenging process and takes a lot of time. It requires a comprehensive analysis of the raw data and a solid background in the problem domain. The most common proposed behavioral authentication methods were built using simple features that are not strongly correlated with the users [51]. On the other hand, some methods used large feature sets such as in [23], but unfortunately, the extraction process takes a lot of time because of the high correlation in the generated feature set. In this paper, we conducted a comprehensive analysis to extract and select an informative and independent feature vector as described in Sect. 4.4.

### 6.3 User experience and usability

The implementation of M2auth brings significant advancements in security through multimodal behavioral biometric authentication, but it also presents certain user experience challenges that need to be addressed to ensure widespread adoption. One key usability challenge is the potential learning curve for users unfamiliar with multimodal systems. Additionally, the continuous nature of data collection may be perceived as intrusive, raising concerns about privacy. Response time is another critical factor; any delays in the authentication process could frustrate users and hinder adoption. Moreover, compatibility across a wide range of devices and operating systems is essential to ensure a seamless user experience.

To mitigate these challenges, several strategies can be employed. First, an effective onboarding process, including tutorials and user education materials, can help users quickly become comfortable with the system. Offering customization options allows users to tailor the frequency and types of biometric checks to their preferences, reducing the perceived intrusiveness of the system. Ensuring that M2auth is optimized for speed and minimal impact on device performance is also crucial for maintaining a smooth user experience. Broad device compatibility can be achieved through extensive testing and adaptation, ensuring that all users have a consistent experience.

To further enhance user acceptance and adoption, it is important to emphasize strong privacy protections, such as data encryption and minimal data retention. Implementing feedback mechanisms will allow users to report their experiences and suggest improvements, fostering a sense of involvement and trust.

M2auth offers robust security benefits, and addressing usability challenges and actively improving the user experience are critical for ensuring its acceptance and adoption. By focusing on user education, customization, performance optimization, and privacy, M2auth can achieve a balance between security and usability, leading to broader adoption across diverse user groups.

The main goal of M$^2$auth is to accept the benign and reject the imposter. The problem lies when the framework reject a valid user (i.e., false rejection) raising a usability issue or incorrectly accept an imposter (i.e., false acceptance) raising a security issue. The decision-level fusion did a great job in decreases these errors as shown in Fig. 7 and Table 5. Moreover, to address these issues, we avoid using all-or-nothing scheme that simply accepts or rejects a user, as this approach tends to reduce usability for the users [57]. We can use a continuous trust score to authenticate users based on the context-aware applications [58]. For instance, in case of security, we can allow access to a banking app with a very high trust score (i.e., decrease FAR and increases the security). On the other hand, for usability, we can allow access, might be a game app, with a low trust score (i.e., decreases FRR and increases the usability).

### 6.4 Comparative analysis

Table 6 presents a performance comparison of M2auth with several state-of-the-art benchmark models in the field of behavioral biometric authentication. This comparison highlights the effectiveness of different authentication methods across various modalities, including gestures, keystrokes, and accelerometer data, as well as their fusion results.

BehavePassDB [22], which uses an LSTM RNN model, shows moderate performance with an AUC of 87.20% in the fusion of gesture, keystroke, and accelerometer data. However, its individual modality accuracies reached 73.22% for gestures, 57.48% for keystrokes, and 66.23% for accelerometer data, indicate a limitation in handling the complexity of multimodal biometric data.

MMauth [16] employs a deep learning-based support vector data description (DeSVDD) algorithm, focusing on one-class learning, but reports an EER of 14.9% without providing detailed modality-specific accuracy metrics. This suggests challenges in achieving high accuracy and low error rates, possibly due to the limitations inherent in one-class learning models.

IncreAuth [11] uses a gradient boosting decision tree supported by a neural network (GBDTNN) model, achieving a high overall accuracy (ACC) of 95.96% in its fusion results. However, the lack of specific performance data for individual modalities limits a detailed comparison.

SBAS (swipe-based authentication system) [59] relies on random forest (RF) and support vector machine (SVM) classifiers, achieving an AUC of 96.9% through the fusion of swipe data. Although this shows strong performance, it does not incorporate other modalities such as keystrokes or accelerometer data, which could enhance overall accuracy.

BioGamesAuth [60] utilizes LSTM and MLP models to combine touch gesture and keystroke dynamics, achieving an accuracy of 98.3% in its best fusion scenario with MLP. This high accuracy reflects the system’s effectiveness, particularly in keystroke dynamics, where it records an accuracy of 97.17%. However, the absence of accelerometer data limits its ability to fully capture diverse user behaviors.

**Table 6** Performance comparison of M2auth with other benchmark models

| Model | Method | Gesture | Keystroke | Accelerometer | Fusion |
|---|---|---|---|---|---|
| BehavePassDB [22] | LSTM RNN | 73.22 | 57.48 | 66.23 | AUC: 87.20 |
| MMauth [16] | DeSVDD | – | – | – | EER: 14.9 |
| IncreAuth [11] | GBDTNN | – | – | – | ACC: 95.96 |
| SBAS [59] | RF, SVM | – | – | – | AUC: 96.9 |
| BioGamesAuth [60] | LSTM, MLP | 77.5 | 97.17 | – | ACC: 98.3 |
| **M2auth (AUC)** | **RF** | **85.29** | **97.59** | **89.17** | AUC: **99.98** |

In comparison, M2auth significantly outperforms these models. It achieves an AUC of 85.29% for gesture-based authentication, 97.59% for keystroke dynamics, and 89.17% for accelerometer data. Most notably, M2auth excels in the fusion of all modalities, reaching an impressive AUC of 99.98%. This outstanding performance is largely attributed to the high-quality data that M2auth leverages, which accurately reflects real-world user behavior. M2auth's dataset is comprehensive, capturing a wide range of user interactions in diverse scenarios, ensuring that the behavioral biometric traits it extracts are not only accurate but also representative of natural user engagement with their devices. In Table 6, the bold values indicate M2auth's performance, emphasizing its results across different modalities and fusion scenarios to showcase the method's superior accuracy and effectiveness.

M2auth's advanced multimodal integration and its superior fusion performance position it as a leading solution in behavioral biometric authentication. Compared to the other benchmark models, M2auth offers enhanced security and accuracy, making it an ideal choice for secure and user-friendly authentication in real-world applications. This analysis highlights the importance of a comprehensive multimodal approach, particularly in achieving high levels of performance in behavioral biometric systems.

## 7 Conclusions and future work

We studied how the combination of multiple authentication modalities optimize the authentication accuracy and addressed some problems that attached with the unimodal systems such as noisy data. We conducted a large-scale field study on Android phone users where we collected more than 100 GB of actual user behavioral data. We developed a multimodal behavioral biometric authentication framework M$^2$auth to authenticate smartphone users over different contexts. We developed three modalities, gesture modality, keystroke modality, and behavioral profiling modality. Our evaluation results show that M$^{2-}$auth framework outperforms the single-modal systems in terms of the error rate and the accuracy.

In the future, we are looking forward to develop and evaluate other modalities by leveraging more sensors, in addition to evaluating the framework in a real-time manner and measuring its usability.

**Acknowledgements** The authors would like to thank participants for their volunteering in this experiment. A special thanks to our colleagues for their feedback on the earlier version of this paper.

**Availability of data and material** Data will be available upon request.

**Code availability** Code will be available upon request.

## Declarations

**Conflict of interest** The authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethics approval** We obtained approval from our university's research ethics board before initiating the data collection process.

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

**Consent for publication** The article was submitted with the consent of all the authors and institutions for publication.

## References

1. Marques D, Muslukhov I, Guerreiro T, Carriço L, Beznosov K (2016) Snooping on mobile phones: Prevalence and trends. In: Twelfth Symposium on Usable Privacy and Security (SOUPS 2016), pp. 159–174. USENIX Association, Denver, CO. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/marques
2. Song Y, Cai Z, Zhang Z-L (2017) Multi-touch authentication using hand geometry and behavioral information. In: 2017 IEEE Symposium on Security and Privacy (SP), pp. 357–372. IEEE
3. Walia KS, Shenoy S, Cheng Y (2020) An empirical analysis on the usability and security of passwords. In: 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), pp. 1–8. https://doi.org/10.1109/IRI49571.2020.00009
4. Tolosana R, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J (2020) Biotouchpass2: Touchscreen password biometrics using time-aligned recurrent neural networks. IEEE Trans Inf Forensics Secur 15:2616–2628
5. Shi D, Tao D, Wang J, Yao M, Wang Z, Chen H, Helal S (2021) Fine-grained and context-aware behavioral biometrics for pattern lock on smartphones. Proc ACM Interact Mobile Wearable Ubiquitous Technol. https://doi.org/10.1145/3448080
6. Agrawal M, Mehrotra P, Kumar R, Shah RR (2022) Gantouch: An attack-resilient framework for touch-based continuous authentication system. IEEE Trans Biomet Behav Ident Sci 4(4):533–543. https://doi.org/10.1109/TBIOM.2022.3206321
7. Egelman S, Jain S, Portnoff RS, Liao K, Consolvo S, Wagner D (2014) Are you ready to lock? In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. CCS '14, pp. 750–761. ACM, New York, NY, USA. https://doi.org/10.1145/2660267.2660273
8. Harbach M, Zezschwitz E, Fichtner A, Luca AD, Smith M (2014) It's a hard lock life: A field study of smartphone (Un)Locking behavior and risk perception. In: 10th Symposium On Usable Privacy and Security (SOUPS 2014), pp. 213–230. USENIX Association, Menlo Park, CA. https://www.usenix.org/conference/soups2014/proceedings/presentation/harbach
9. Raja KB, Raghavendra R, Vemuri VK, Busch C (2015) Smartphone based visible iris recognition using deep sparse filtering. Pattern Recogn Lett 57:33–42
10. Fathy ME, Patel VM, Chellappa R (2015) Face-based active authentication on mobile devices. In: Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference On, pp. 1687–1691. https://doi.org/10.1109/ICASSP.2015.7178258
11. Shen Z, Li S, Zhao X, Zou J (2023) Increauth: Incremental learning based behavioral biometric authentication on

smartphones. IEEE Internet of Things J. https://doi.org/10.1109/JIOT.2023.3289935

12. Krishnamoorthy S, Rueda L, Saad S, Elmiligi H (2018) Identification of user behavioral biometrics for authentication using keystroke dynamics and machine learning. In: Proceedings of the 2018 2Nd International Conference on Biometric Engineering and Applications. ICBEA '18, pp. 50–57. ACM, New York, NY, USA. https://doi.org/10.1145/3230820.3230829 . http://doi.acm.org/10.1145/3230820.3230829
13. Khaleghi B, Khamis A, Karray FO, Razavi SN (2013) Multisensor data fusion: a review of the state-of-the-art. Inf Fusion 14(1):28–44. https://doi.org/10.1016/j.inffus.2011.08.001
14. Mahfouz A, Mahmoud TM, Eldin AS (2017) A survey on behavioral biometric authentication on smartphones. J Inf Secur Appl 37:28–37
15. Meng W, Wang Y, Wong DS, Wen S, Xiang Y (2018) Touchwb: Touch behavioral user authentication based on web browsing on smartphones. J Netw Comput Appl 117:1–9. https://doi.org/10.1016/j.jnca.2018.05.010
16. Shen Z, Li S, Zhao X, Zou J (2022) Mmauth: A continuous authentication framework on smartphones using multiple modalities. IEEE Trans Inf Forensics Secur 17:1450–1465. https://doi.org/10.1109/TIFS.2022.3160361
17. Al-Saraireh J, AlJa'afreh MR (2023) Keystroke and swipe biometrics fusion to enhance smartphones authentication. Comput Secur 125:103022
18. Zaidi AZ, Chong CY, Jin Z, Parthiban R, Sadiq AS (2021) Touch-based continuous mobile device authentication: state-of-the-art, challenges and opportunities. J Netw Comput Appl 191:103162
19. Mahfouz A, Hamdy A, Eldin MA, Mahmoud TM (2024) B2auth: A contextual fine-grained behavioral biometric authentication framework for real-world deployment. Pervas Mobile Comput. https://doi.org/10.1016/j.pmcj.2024.101888
20. Xu X, Yu J, Chen Y, Hua Q, Zhu Y, Chen Y-C, Li M (2020) Touchpass: Towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations. In: Proceedings of the 26th Annual International Conference on Mobile Computing and Networking, pp. 1–13
21. Stylios I, Kokolakis S, Thanou O, Chatzis S (2021) Behavioral biometrics & continuous user authentication on mobile devices: a survey. Inf Fusion 66:76–99
22. Stragapede G, Vera-Rodriguez R, Tolosana R, Morales A (2023) Behavepassdb: Public database for mobile behavioral biometrics and benchmark evaluation. Pattern Recogn 134:109089. https://doi.org/10.1016/j.patcog.2022.109089
23. Frank M, Biedert R, Ma E, Martinovic I, Song D (2013) Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. IEEE Trans Inf Forensics Secur 8(1):136–148. https://doi.org/10.1109/TIFS.2012.2225048
24. Peng G, Zhou G, Nguyen DT, Qi X, Yang Q, Wang S (2017) Continuous authentication with touch behavioral biometrics and voice on wearable glasses. IEEE Trans Human Mach Syst 47(3):404–416
25. Buriro A, Crispo B, Conti M (2019) Answerauth: A bimodal behavioral biometric-based user authentication scheme for smartphones. J Inf Secur Appl 44:89–103. https://doi.org/10.1016/j.jisa.2018.11.008
26. Shen C, Li Y, Chen Y, Guan X, Maxion RA (2018) Performance analysis of multi-motion sensor behavior for active smartphone authentication. IEEE Trans Inf Forensics Secur 13(1):48–62. https://doi.org/10.1109/TIFS.2017.2737969
27. Syed Z, Helmick J, Banerjee S, Cukic B (2019) Touch gesture-based authentication on mobile devices: the effects of user posture, device size, configuration, and inter-session variability. J Syst Softw 149:158–173. https://doi.org/10.1016/j.jss.2018.11.017
28. Yang Y, Guo B, Wang Z, Li M, Yu Z, Zhou X (2019) Behavesense: Continuous authentication for security-sensitive mobile apps using behavioral biometrics. Ad Hoc Netw 84:9–18
29. Gunetti D, Picardi C (2005) Keystroke analysis of free text. ACM Trans Inf Syst Secur 8(3):312–347. https://doi.org/10.1145/1085126.1085129
30. Mondal S, Bours P (2017) Person identification by keystroke dynamics using pairwise user coupling. IEEE Trans Inf Forensics Secur 12(6):1319–1329. https://doi.org/10.1109/TIFS.2017.2658539
31. Lee W-H, Lee RB (2017) Implicit smartphone user authentication with sensors and contextual machine learning. In: 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 297–308. IEEE
32. Shen C, Chen Y, Guan X (2018) Performance evaluation of implicit smartphones authentication via sensor-behavior analysis. Inf Sci 430–431:538–553. https://doi.org/10.1016/j.ins.2017.11.058
33. Jorquera Valero JM, Sánchez Sánchez PM, Fernández Maimó L, Huertas Celdrán A, Arjona Fernández M, De Los Santos Vélchez S, Marténez Pérez G (2018) Improving the security and qoe in mobile devices through an intelligent and adaptive continuous authentication system. Sensors **18**(11) https://doi.org/10.3390/s18113769
34. Muslukhov I, Boshmaf Y, Kuo C, Lester J, Beznosov K (2012) Understanding users' requirements for data protection in smartphones. In: Proceedings of the 2012 IEEE 28th International Conference on Data Engineering Workshops. ICDEW '12, pp. 228–235. IEEE Computer Society, Washington, DC, USA. https://doi.org/10.1109/ICDEW.2012.83 . http://dx.doi.org/10.1109/ICDEW.2012.83
35. Zhou Y, Jiang X (2012) Dissecting android malware: Characterization and evolution. In: Proceedings of the 2012 IEEE Symposium on Security and Privacy. SP '12, pp. 95–109. IEEE Computer Society, Washington, DC, USA. https://doi.org/10.1109/SP.2012.16
36. Yang Z, Yang M, Zhang Y, Gu G, Ning P, Wang XS (2013) Appintent: analyzing sensitive data transmission in android for privacy leakage detection. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. CCS '13, pp. 1043–1054. ACM, New York, NY, USA. https://doi.org/10.1145/2508859.2516676
37. Nandugudi A, Maiti A, Ki T, Bulut F, Demirbas M, Kosar T, Qiao C, Ko SY, Challen G (2013) Phonelab: A large programmable smartphone testbed. In: Proceedings of First International Workshop on Sensing and Big Data Mining. SENSEMINE'13, pp. 4–146. ACM, New York, NY, USA. https://doi.org/10.1145/2536714.2536718
38. Weisstein EW (2022) Curvature, A Wolfram Web Resource. January 2022
39. Li F, Clarke N, Papadaki M, Dowland P (2014) Active authentication for mobile devices utilising behaviour profiling. Int J Inf Secur 13(3):229–244. https://doi.org/10.1007/s10207-013-0209-6
40. Sitová Z, Šeděnka J, Yang Q, Peng G, Zhou G, Gasti P, Balagani KS (2016) Hmog: New behavioral biometric features for continuous authentication of smartphone users. IEEE Trans Inf Forensics Secur 11(5):877–892. https://doi.org/10.1109/TIFS.2015.2506542
41. Ross AA, Jain AK, Nandakumar K (2006) Decision level fusion. Handbook of Multibiometrics, 91–142
42. Chen CH, Chen CY (2013) Optimal fusion of multimodal biometric authentication using wavelet probabilistic neural network. In: 2013 IEEE International Symposium on Consumer

Electronics (ISCE), pp. 55–56. https://doi.org/10.1109/ISCE.2013.6570127

43. Haghighat M, Abdel-Mottaleb M, Alhalabi W (2016) Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans Inf Forensics Secur 11(9):1984–1996. https://doi.org/10.1109/TIFS.2016.2569061
44. Cheng G, Han J (2016) A survey on object detection in optical remote sensing images. ISPRS J Photogram Remote Sens 117:11–28. https://doi.org/10.1016/j.isprsjprs.2016.03.014
45. Kuncheva LI (2004) Combining Pattern Classifiers: Methods and Algorithms. John Wiley & Sons, ???
46. Ho TK (2002) Multiple classifier combination: Lessons and next steps. In: Hybrid Methods in Pattern Recognition, pp. 171–198. World Scientific, ???
47. Patel VM, Chellappa R, Chandra D, Barbello B (2016) Continuous user authentication on mobile devices: recent progress and remaining challenges. IEEE Signal Process Magaz 33(4):49–61. https://doi.org/10.1109/MSP.2016.2555335
48. Monroe D (2012) Biometrics Metrics Report v3. 0. December
49. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
50. Teh PS, Zhang N, Teoh ABJ, Chen K (2016) A survey on touch dynamics authentication in mobile devices. Comput Secur 59(C):210–235. https://doi.org/10.1016/j.cose.2016.03.003
51. Jain AK, Nandakumar K, Ross A (2016) 50 years of biometric research: accomplishments, challenges, and opportunities. Pattern Recogn Lett 79:80–105
52. Buschek D, De Luca A, Alt F (2015) Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. CHI '15, pp. 1393–1402. ACM, New York, NY, USA. https://doi.org/10.1145/2702123.2702252
53. Khan H, Hengartner U, Vogel D (2015) Usability and security perceptions of implicit authentication: Convenient, secure, sometimes annoying. In: Eleventh Symposium On Usable Privacy and Security (SOUPS 2015), pp. 225–239. USENIX Association, Ottawa. https://www.usenix.org/conference/soups2015/proceedings/presentation/khan
54. Draffin B, Zhu J, Zhang J (2013) Keysens: Passive user authentication through micro-behavior modeling of soft keyboard interaction. In: International Conference on Mobile Computing, Applications, and Services, pp. 184–201. Springer
55. Khan H, Hengartner U (2014) Towards application-centric implicit authentication on smartphones. In: Proceedings of the 15th Workshop on Mobile Computing Systems and Applications. HotMobile '14, pp. 10–1106. ACM, New York, NY, USA. https://doi.org/10.1145/2565585.2565590
56. Xu H, Zhou Y, Lyu MR (2014) Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. In: Symposium On Usable Privacy and Security (SOUPS 2014), pp. 187–198. USENIX Association, Menlo Park, CA. https://www.usenix.org/conference/soups2014/proceedings/presentation/xu
57. Hayashi E, Riva O, Strauss K, Brush AJB, Schechter S (2012) Goldilocks and the two mobile devices: Going beyond all-or-nothing access to a device's applications. In: Proceedings of the Eighth Symposium on Usable Privacy and Security. SOUPS '12. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/2335356.2335359
58. Elmalaki S, Wanner L, Srivastava M (2015) Caredroid: Adaptation framework for android context-aware applications. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 386–399. ACM
59. Chao J, Hossain MS, Lancor L (2023) Swipe gestures for user authentication in smartphones. J Inf Secur Appl 74:103450. https://doi.org/10.1016/j.jisa.2023.103450
60. Stylios I, Chatzis S, Thanou O, Kokolakis S (2023) Continuous authentication with feature-level fusion of touch gestures and keystroke dynamics to solve security and usability issues. Comput Secur 132:103363. https://doi.org/10.1016/j.cose.2023.103363

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.