

Continuous Behavioral Age Verification: Estimating Underage Users from Free-Text Typing Patterns

Sarker Tanveer Ahmed Rume¹, Ayon Shahrier¹, Md. Abid Raihan¹ and Farkhunda Dorin

¹ Department of Computer Science and Engineering, University of Dhaka, Dhaka, Bangladesh

² Bangladesh Institute of Management, Dhaka, Bangladesh

Abstract: Keystroke dynamics has been studied for decades as a behavioral biometric, but the vast majority of this work has been directed toward user authentication. This study instead investigates an underexplored application: estimating user age from keystroke patterns in order to identify underage users, offering a lightweight, privacy-preserving complement to existing age-verification methods. Using free-text keystroke data collected from 80 volunteers aged 16 to 60, we formulate age estimation as a binary classification problem distinguishing users under 21 from adults, and evaluate three classifiers, K-Nearest Neighbor, Support Vector Machine, and Random Forest, on a comprehensive set of timing-derived features. Our best-performing model achieves a recall of 0.86, with the two strongest classifiers averaging an F1-score of 0.71, indicating that the incidental behavioral signal present in ordinary typing carries meaningful information about a user's age. These findings point to keystroke dynamics as a promising tool for addressing the growing challenge of child safety online.

Keywords: Keystroke Dynamics; Age Verification; Machine Learning; Behavioral Biometrics; Child Online Safety.

1. Introduction

The number of internet users under the age of 18 is rising quickly, driven by the need for technology to complete schoolwork and maintain social connections. This growing presence comes with real risk: a McAfee poll found that 71% of teenagers hide their online activity from parents and 62% disclose personal information without their parents' knowledge [11], while a separate study in India reported that 82% of children had received inappropriate messages on Facebook, with 67% holding accounts before the age of 10 [14]. Government agencies continue working to safeguard children from such unidentified online hazards [8], yet the effectiveness of existing measures, such as platform age limits, remains questionable.

This research proposes keystroke dynamics as an alternative strategy for identifying underage users. Prior work suggests that children can be distinguished from adults through differences in physical, cognitive, and behavioral traits, including keyboard experience and typing posture, all of which shape typing patterns for a given age group [16]. Like a voice-print or signature, a person's typing rhythm is a behavioral biometric that has been studied since World War II, when telegraph operators could be identified by their keying rhythm; similar techniques were later extended to computer keyboards. Keystroke dynamics research has since focused primarily on authentication, though several studies have noted that age, gender, and experience all influence identification outcomes, raising the possibility that age-discriminative traits could be isolated once more personal, identity-specific characteristics are removed.

While most prior work targets authentication or other static user characteristics, this study classifies user age from free-flowing text, enabling continuous, unobtrusive age assessment that can be integrated into existing systems. We see this as a step toward better protecting younger users from scams, cyber predators, and phishing attempts.

Unlike more stable physiological biometrics such as fingerprints or facial features, typing patterns vary considerably over both short and long timescales: fatigue, distraction, movement, and posture all affect typing behavior within a single session, and patterns vary further with keyboard type, layout, and hardware, introducing additional irregularity into the collected data.

The key contributions of this paper are summarized as follows:

- **Novel application domain:** we repurpose keystroke dynamics, a biometric traditionally used for authentication, to address the underexplored problem of estimating user age for child-safety purposes.
- **Free-text, continuous formulation:** we frame age estimation as a binary classification task (under 21 vs. adult) on free-flowing typed text, enabling continuous, unobtrusive verification rather than a one-time, fixed-text check.
- **Comprehensive, interpretable feature set:** we design and extract six categories of timing-derived features, hold-time, seek-time, move-time, bigraph latency, error rate, and pause-between-words, directly from naturalistic, free-form typing.
- **Empirical evaluation on original data:** we collect a new free-text dataset from 80 volunteers aged 16-60 and benchmark three classifiers, KNN, SVM, and Random Forest, achieving a best recall of 0.86 and an average F1-score of 0.71 across the two strongest models.
- **Ethical and deployment framing:** we identify the privacy and consent considerations raised by analyzing incidental typing behavior and outline safeguards, such as on-device processing, data minimization, and a transparent appeals process, needed for a responsible real-world deployment (revisited in Section 6).

The remainder of this paper is organized as follows. Section 2 reviews related literature, including recent advances published between 2023 and 2026. Section 3 details our proposed feature set and classification models. Section 4 describes our data collection and feature extraction methodology. Section 5 presents and analyzes our experimental results. Section 6 concludes with a summary and directions for future work.

2. Related Work

Password-based authentication, while simple to adopt, suffers from well-known weaknesses such as shoulder surfing, keyloggers, phishing, and brute-force attacks. Keystroke dynamics has emerged as one of several proposed solutions to these weaknesses, and is now a mature technology, with decades of research producing authentication systems with very low equal error rates (EER). Lin et al. [7] used features such as keystroke duration (KD) and up-down digraph latency (UDDL) to detect unauthorized users even when the correct password is known, while Venugopal and Viji [15] combined keystroke durations with four digraph-latency measures (down-down, down-up, up-up, and up-down) to achieve an EER of 0.5% on password-entry data collected in MATLAB. Sun et al. [12] focused on more granular features, including Space, Backspace, and arrow-key timings alongside select digraph latencies, achieving an EER of 2.94% and an AUC of 0.994 using only 13 features from 34 users.

Beyond authentication, keystroke dynamics has also been used to infer other user characteristics: Kołakowska [6] classified emotional state from typing behavior, while Buker et al. [2] and Bandeira et al. [4] separately explored gender prediction, motivated in part by the growing scale of live-chat systems. Bandeira et al. [4] predicted gender from free-text keystrokes across 100 volunteers, with an

SVM achieving 64% accuracy using 29 features, while Brizan et al. [1] predicted handedness from 329 participants, obtaining an f-score of 0.223 for the left-hand class.

Most directly relevant to this work are studies on age classification from keystroke dynamics. Buriro et al. [3] estimated user age from short PIN/password inputs on a mobile device across 150 volunteers and three age groups, comparing Naive Bayes, SVM, Random Forest, MLP, and Deep Neural Network classifiers and finding Random Forest the strongest performer. Pentel [10] similarly favored Random Forest, achieving 61.6% accuracy across six age groups using 134 features from over 7,000 users, while Tsimperidis et al. [13] used 120 down-down digraph latencies from 239 log files across four age groups (18-25, 26-35, 36-45, 46+) to reach 66.1% accuracy with a boosted MLP. Pentel [9] later achieved a maximum F1-score of 0.73 predicting age and gender jointly from both physical and virtual keyboards.

Interest in keystroke dynamics has grown well beyond authentication in recent years. Shadman et al. [16] survey the field's state of the art, covering novel datasets, algorithms, and mobile deployments, and highlight demographic and cognitive-state inference as an expanding research direction. Closer to this study's aim, Kundu et al. [17] used a modified TypeNet architecture to distinguish genuine student writing from generative-AI-assisted writing with up to 85.72% accuracy, reinforcing that fine-grained keystroke timing carries rich, exploitable signal about the user, while Kuruvilla et al. [21] applied keystroke dynamics to detect fake social-media profiles, a safety application aligned with our own motivation.

On the authentication side, Yang et al. [18] proposed a federated, cross-device keystroke authentication scheme to address keyboard and device variability, a challenge also raised in Section 1. Kuric et al. [19] showed that mouse hardware can similarly bias extracted behavioral features, a caution equally relevant to our own heterogeneous, self-selected keyboard data; because our volunteers used their own devices rather than a single controlled one, some of the variance in individual feature vectors, and the misclassifications discussed in Section 5, may stem from this device heterogeneity, a factor future iterations of this work should explicitly model.

Most directly relevant to age estimation, Moon et al. [20] found that smartphone hold-time and press-to-press latency vary systematically with age and time of day, enabling AI-based age estimation with a mean absolute error under four years; their finding that early-morning and late-evening sessions are most age-discriminative offers a promising direction for future session-aware extensions of this work, revisited in Section 6.

3. Proposed Methodologies

Age-Discriminative Keystroke Feature Extraction

Identifying a feature set capable of discriminating user age is central to solving the underlying prediction problem. Because typing behavior manifests primarily through timing variation rather than through the content typed, the great majority of the extracted features are time-based and require precise measurement to be reliably attributed to a given user. These features are derived from intervals between keyboard events, each of which corresponds to either a key press or a key release; for clarity, the press action of a key is denoted "Down" (D) and the release action "Up" (U) throughout the remainder of this paper. On this basis, the following feature categories are defined:

- I. **Hold-time:** Mean time between key press and release for a particular key-X, identified as (D-X, U-X) measured for X.
- II. **Seek-time:** Mean time between pressing of a previous key-X and current key press for particular key-Y, identified as (D-X, D-Y) measured for Y.

- III. **Move-time:** Mean time between release of a previous key-X and current key press for particular key-Y, identified as (U-X, D-Y) measured for Y.
- IV. **Bigraph Latency:** Mean time between 2 consecutive different key presses beginning with the first key press of key-X and ending with the second key press of key-Y making a X-Y bigraph delay. This is identified as (D-X, D-Y) measured for XY.
- V. **Error rate:** Ratio of total number of BACKSPACE key presses and total number of keys.
- VI. **Pause between words:** Mean time between a SPACE key press followed by any other different key press.

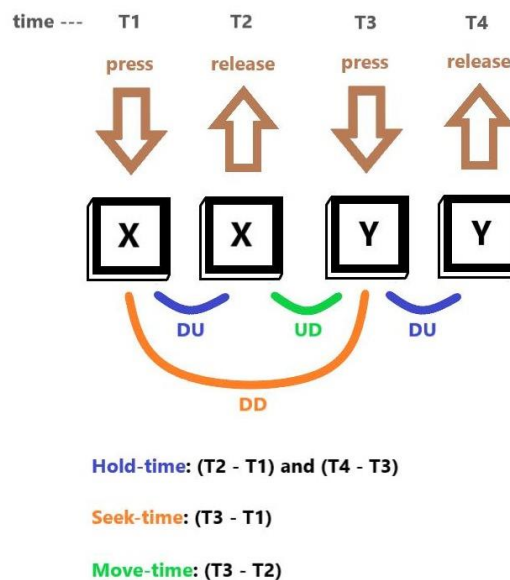


Figure 1: Inter-key timing intervals (hold-, seek-, and move-time) computed from consecutive key-press (D) and key-release (U) events, forming the basis of the age-discriminative feature set

Figure 1 illustrates these inter-key latencies graphically for a representative sequence of key events; each delay is computed by subtracting the timestamp of the earlier event from that of the later event, as indicated.

Table 1: Time-Based Keystroke Feature Set, Defining Key-Event Pairs, Observed Instance Counts, and Mean Measured Time Across All Participants

No.	Feature Name	Typecode	Key	Total	Mean Time (ms)
1	Hold-time	D-X, U-X	X	34	101.79
2	Seek-time	D-X, D-Y	Y	30	311.54
3	Move-time	U-X, D-Y	Y	31	261.89
4	Bigraph Latency	D-X, D-Y	XY	288	272.27
5	Pause between words	U-SPACE, D-Y	SPACE	1	472.71

Once extracted, only features present in at least one-third of the dataset are retained in the final feature set, so as to limit the influence of sparsely populated columns on downstream model training. Table 1 summarizes the resulting time-based features, together with the total number of observed instances and

the mean measured time for each, across all participants. The average error rate, computed as a ratio over all users, was 0.0767.

Classification Models for Underage-User Detection

The core prediction problem addressed in this work is framed as a binary classification task: determining whether a user's age falls below a given threshold. After retaining the features that satisfy the coverage criterion described above, the dataset is partitioned into training and testing subsets. Missing values are then imputed using the corresponding column mean computed from the training set, after which all features are scaled accordingly.

The K-Nearest Neighbor (KNN) classifier is adopted as the primary model, motivated by its distance-based formulation, which parallels the template-matching approach long used in keystroke-based authentication. Its performance is benchmarked against two other widely used classifiers, a Support Vector Machine (SVM) and a Random Forest Classifier.

The Random Forest Classifier is an ensemble of decision trees trained on bootstrapped subsets of the data and of the feature set, whose majority vote is used as the final prediction; it is included as a benchmark given its strong prior performance in related age-classification studies such as Buriro et al. [3] and Pentel et al. [10].

Principal component analysis (PCA) is applied to reduce feature dimensionality prior to training, and grid-search optimization is used to select the best-performing estimator for each model via cross-validation on the training set. Model performance is then evaluated across multiple scoring metrics, as reported in Section 5.

4. Implementation

Data Acquisition

Researchers require input data collected from computer users in various real-world scenarios in order to do research in the field of keystroke dynamics biometrics. The keys touched on the keyboard and the moments at which they are pressed indicate the critical information. Only a constrained or regulated setting will allow for the collection of this information. As such, the data collection carried out for this research followed acquisition methods similar to those used to gather the publicly shared dataset of Iapa and Cretu [5].

The authors created their own environment to collect data from participants for the study's objective. Participants performed the experiment from their own computers at their convenience, unlike earlier research when the observers had total control over the data collection. Each user's text entry from the keyboard was free-form as well and did not require them to duplicate a certain prepared text. This marks a significant difference from previous work on short, fixed texts, since this should be more applicable to real-world scenarios.

In spite of the fact that practically all participants were familiar with computers, typing competence was not a criterion for this study. A set of general questions was created for each text box to direct the user to a specific section of the text being filled out. There were inquiries regarding the climate, the perfect day, and the educational system. Participants were given permission to answer freely as they please within a given time frame. The goal of this is to record users' natural, free-flowing text-writing process. After due time, volunteers gave their consent and submitted their written paragraphs capturing the keystroke timings.

Dataset Description

A total of 80 volunteers participated in the data collection process including 36 males and 44 females. The average age of the users is 28 years with the youngest user being 16 years old and the oldest being 60. The total user count for each age group can be found in Table 2. Users were allowed to use any keyboard-attached device to perform the task, so participants could work on laptops or desktops as they chose.

Table 2: Age group of volunteers

Age-group	User-count
16-22	24
23-29	23
30-34	20
35-44	9
45+	4

A sum of 410,633 key events occurred in total over all participants that are being used for the feature extraction. That amounts to an average of 5133 key events per user. There were more than 100 different keys monitored throughout the procedure that helped produce the keystroke timings with the SPACE key understandably being the most popular event. Multiple keys can be pressed at the same time, cases in which the hold-time for those keys will overlap, such as using a SHIFT key to capitalize a letter.

Feature Extraction

Timings are all recorded in milliseconds, marking the moment a key is pressed or released by a user. The recording for each user is stored in separate text files.

To measure the interval times, only keys pressed within 3 seconds of each other are accounted for. Keys pressed after this threshold are considered to be a pause from the user for whatever environmental reason and are thus discarded from our system. Multiple consecutive keys pressed within 50 milliseconds are similarly left out, since they are likely to represent a mistaken key hold. Double-presses of a key were also ignored, so bigraphs with the same key on both sides do not appear in our models, improving measurement quality.

For move-time, the SHIFT key is excluded from measurement, since it is always released after a different key is pressed, which would otherwise yield a negative move-time value; the number of such exclusions therefore varies depending on how often SHIFT is used within a given paragraph.

Any feature not recorded for a user is left blank, and only features that occurred in at least one-third of the dataset were finally passed to our model. The missing blank values were then replaced by the column mean during model training, since it represents the average time taken by users for that feature.

5. Experiment Results

We train and test each of our three classifier models multiple times and average the results to obtain a reliable measurement of the prediction scores. Grid-search hyperparameter tuning selects the best estimator for each model from its given parameter set using 5-fold cross-validation, with the F1-score used to evaluate the performance of each combination.

Table 3: Model Prediction Scores on Evaluation Metrics

Model	Recall	Accuracy	F1-Score
K-Nearest Neighbor	0.86	0.77	0.74
Support Vector Machine	0.79	0.71	0.67
Random Forest Classifier	0.60	0.57	0.58

Table 3 displays the results of our current findings with different prediction models. Upon further observation, the K-Nearest-Neighbor (KNN) algorithm seems to perform significantly better than the two other classifiers while the Support Vector Machine (SVM) model showcases worse outcomes albeit with a decent recall score. The Random Forest Classifier has produced far weaker results in comparison.

To place these scores in context, Table 4 summarizes the reported performance of the most closely related prior studies on age classification from keystroke dynamics discussed in Section 2, alongside the KNN result obtained in the present study. Direct comparison must be treated with caution given the differences in dataset size, age-group boundaries, text-entry mode (fixed versus free text), and evaluation protocol across studies, but the comparison nonetheless situates our free-text, single age-threshold approach within the broader body of work.

Table 4: Comparison with Related Age-Classification Studies

Study	Data / Text Mode	Best Model	Reported Score
Buriro et al. [3]	PIN/password, 150 users	Random Forest	Best of 5 classifiers
Pentel [10]	Free text, 6 age groups	Random Forest	61.6% accuracy
Tsimperidis et al. [13]	Free text, 4 age groups	MLP + boosting	66.1% accuracy
Pentel [9]	Mixed keyboards, 6 groups	5 ML models (Weka)	F1 up to 0.73
Present study	Free text, binary (<21)	K-Nearest Neighbor	0.86 recall / 0.74 F1

The KNN model's recall of 0.86 compares favorably with the recall-oriented goal of this study, which prioritizes correctly flagging underage users over overall accuracy, since a missed detection (false negative) is more costly in a child-safety context than an occasional false positive. While the multiclass studies of Pentel [9] and Tsimperidis et al. [13] address a more fine-grained age-grouping problem, our binary formulation around a single, policy-relevant threshold (age 21) is more directly comparable to the deployment scenario described in Section 1, namely flagging accounts for further review rather than estimating exact age.

Discussion

Our primary focus in this study was to extend previous work on age-based classification with keystroke dynamics from fixed-text to free-text input that can contribute to continuous verification. Since we wish to identify underage users, correctly classifying each positive case is of utmost importance. Recall is the most relevant metric for this goal, and two of our models produce promising results in this regard.

The assumption that KNN would perform significantly better is also supported by our results. The model uses a distance-based proximity algorithm, which is analogous to the authentication techniques historically applied with keystroke dynamics. It can therefore be assumed that the KNN model can, to some extent, accurately predict a user's correct age group.

It is worth noting that the Random Forest Classifier, despite being the strongest performer in several of the related studies discussed in Section 5, produced the weakest results of the three models evaluated here. We attribute this discrepancy primarily to the comparatively small size of our volunteer pool (80 users) relative to the datasets used by Pentel [10] (over 7,000 logs) and Buriro et al. [3] (150 users), since ensemble tree-based methods typically require a larger sample to avoid overfitting to a high-dimensional feature space of 385 candidate features. This observation is consistent with the future work outlined in Section 6, where expanding the dataset is identified as a priority for improving the robustness of all three classifiers, and particularly the ensemble-based approaches. Some of this misclassification may also be attributable to device heterogeneity, since volunteers were free to use their own laptop or desktop keyboards rather than a single controlled input device, as discussed in Section 2.

6. Conclusion

This study investigated keystroke dynamics as a lightweight, privacy-preserving signal for identifying underage users from free-text typing. Using a dataset of 80 volunteers aged 16–60, we extracted a comprehensive set of timing-derived features and evaluated three classifiers on a binary age-threshold task. The K-Nearest Neighbor model achieved the strongest performance, with a recall of 0.86, confirming that the incidental behavioral signal present in ordinary typing carries meaningful information about a user's age. Positioned within a broader trust-and-safety pipeline rather than as a stand-alone mechanism, this approach offers a promising complement to existing age-verification methods and could help flag accounts for review, including cases of child grooming where an adult poses as a younger user.

Several limitations remain: the boundaries that separate age groups are not fully understood, it is unclear whether users could deliberately alter their typing to evade detection, and our limited sample size constrained the performance of ensemble methods such as Random Forest. Future work should therefore prioritize expanding the dataset in size, age granularity, and language coverage; incorporating additional features such as trigrams and session-aware, time-of-day signals as suggested by Moon et al. [20]; and exploring federated, on-device deployment along the lines of Yang et al. [18] to preserve privacy across heterogeneous devices. Realizing these directions would move continuous, privacy-preserving age verification from typing behavior closer to practical deployment.

REFERENCES

- [1] D. G. Brizan, A. Goodkind, P. Koch, K. Balagani, V. V. Phoha, and A. Rosenberg, "Utilizing linguistically-enhanced keystroke dynamics to predict typist cognition and demographics," *International Journal of Human-Computer Studies*, vol. 82, May 2015.
- [2] A. A. N. Buker, G. Roffo, and A. Vinciarelli, "Type like a man! Inferring gender from keystroke dynamics in live-chats," *IEEE Intelligent Systems*, vol. 34, no. 6, pp. 53–59, 2019.
- [3] A. Buriro, Z. Akhtar, B. Crispo, and F. Del Frari, "Age, gender and operating-hand estimation on smart mobile devices," in *Proc. 2016 Int. Conf. of the Biometrics Special Interest Group (BIOSIG)*, 2016, pp. 1–5.
- [4] D. R. Cavalcante Bandeira, A. M. de Paula Canuto, M. Da Costa-Abreu, M. Fairhurst, C. Li, and D. S. Costa do Nascimento, "Investigating the impact of combining handwritten signature and keyboard keystroke dynamics for gender prediction," in *Proc. 2019 8th Brazilian Conf. on Intelligent Systems (BRACIS)*, 2019, pp. 126–131.
- [5] C. Iapa and V. Cretu, "Shared data set for free-text keystroke dynamics authentication algorithms," May 2021. doi:10.20944/preprints202105.0255.v1.

- [6] A. Kołakowska, "Recognizing emotions on the basis of keystroke dynamics," in *Proc. 2015 8th Int. Conf. on Human System Interaction (HSI)*, 2015, pp. 291–297.
- [7] C.-H. Lin, J.-C. Liu, and K.-Y. Lee, "On neural networks for biometric authentication based on keystroke dynamics," *Sensors and Materials*, vol. 30, pp. 385–396, Jan. 2018.
- [8] S. Newman, "Children and the dangers of the virtual world," *Cybercrime and Security Awareness*. [Online]. Available: <https://informationsecuritybuzz.com/children-and-the-dangers-of-the-virtual-world/>
- [9] A. Pentel, "Predicting age and gender by keystroke dynamics and mouse patterns," in *Proc. 25th Conf. on User Modeling, Adaptation and Personalization*, 2017, pp. 381–385.
- [10] A. Pentel, "Predicting user age by keystroke dynamics," in *Artificial Intelligence and Algorithms in Intelligent Systems*, Springer International Publishing, Jan. 2019, pp. 336–343.
- [11] R. Siciliano, "Do you know what your kids are hiding?," in *Digital Deception: Exploring the Online Disconnect between Parents and Kids*, June 2013. [Online]. Available: <https://www.mcafee.com/blogs/family-safety/digital-divide>
- [12] Y. Sun, H. Çeker, and S. J. Upadhyaya, "Anatomy of secondary features in keystroke dynamics – achieving more with less," in *Proc. Int. Conf. on Identity, Security and Behavior Analysis (ISBA)*, New Delhi, India, 2017, pp. 1–6.
- [13] I. Tsimperidis, S. Rostami, and V. Katos, "Age detection through keystroke dynamics from user authentication failures," *International Journal of Digital Crime and Forensics*, vol. 9, pp. 1–16, 2017.
- [14] M. Variyar, "82% children on Facebook get vulgar messages," *Hindustan Times*, Mumbai. [Online]. Available: <https://www.hindustantimes.com/mumbai/82-children-on-facebook-get-vulgar-messages/story-0d532SUH4E2kYDN4o1ja8H.html>
- [15] P. C. Venugopal and K. S. A. Viji, "Applying empirical thresholding algorithm for a keystroke dynamics based authentication system," *Journal of Information and Communication Technology*, vol. 18, no. 4, pp. 383–413, Sep. 2019.
- [16] R. Shadman, A. A. Wahab, M. Manno, M. Lukaszewski, D. Hou, and F. Hussain, "Keystroke dynamics: concepts, techniques, and applications," *arXiv preprint arXiv:2303.04605*, 2023.
- [17] D. Kundu, A. Mehta, R. Kumar, N. Lal, A. Anand, A. Singh, and R. R. Shah, "Keystroke dynamics against academic dishonesty in the age of LLMs," in *Proc. 2024 IEEE Int. Joint Conf. on Biometrics (IJCB)*, Buffalo, NY, USA, 2024, pp. 1–10.
- [18] Y. Yang, B. Guo, Y. Liang, K. Zhao, and Z. Yu, "Cross-device free-text keystroke dynamics authentication using federated learning," *Personal and Ubiquitous Computing*, vol. 28, no. 3–4, pp. 491–505, 2024.
- [19] E. Kuric, P. Demčák, M. Krajčovič, and P. Nemček, "Is mouse dynamics information credible for user behavior research? An empirical investigation," *Computer Standards & Interfaces*, vol. 90, art. 103849, 2024.
- [20] J. Moon, Y. L. Huh, H. Y. Cho, C. Kim, H. Lee, D. P. Jang, and B. H. Cho, "Exploring age-related patterns in smartphone keystroke dynamics considering temporal variability: cross-sectional study with AI-based analysis," *JMIR mHealth and uHealth*, vol. 13, art. e80094, 2025.
- [21] A. Kuruvilla, R. Daley, and R. Kumar, "Spotting fake profiles in social networks via keystroke dynamics," in *Proc. 2024 IEEE 21st Consumer Communications & Networking Conf. (CCNC)*, 2024.