SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Forensic-by-Design Framework Using Multimodal Behavioral Biometrics for Mobile Pharmacy Security

Bashor Fauzan Muthohirin[1)*], Novan Visia Farmasari[2)], Rizky Pamuji[3)]
[1)] Department of Informatics, Universitas Muhammadiyah Malang, Malang, Indonesia,
[2)]Department of Pharmacy, Universitas Muhammadiyah Malang, Malang, Indonesia,
[3)] Department of Computer Science Education, Universitas Lambung Mangkurat, Banjarmasin, Indonesia
[1)]bashorfauzan@umm.ac.id, [2)]novanvisia@umm.ac.id, [3)]rizky.pamuji@ulm.ac.id

Submitted :Sep 15, 2026 | **Accepted** : Oct 16, 2026 | **Published** : Oct 20, 2026

Abstract: In the mobile pharmacy sector, digital transformation increases the exposure of sensitive data and opens up insider threats that are not easily addressed by traditional security mechanisms such as login authentication and encryption, whose traces are only collected after an incident and often do not meet evidentiary standards. In this study we develop a multimodal behavioral biometric framework based on Forensic-by-Design principles to passively and continuously detect user interaction anomalies in mobile pharmacy applications. We develop three modules with hybrid architecture (Long Short-Term Memory network for keystroke dynamics, eXtreme Gradient Boosting for scrolling behavior and time threshold heuristic for detecting input that comes from paste or auto-injection). The keystroke module was tested on the Carnegie Mellon University keystroke benchmark (51 subjects) session disjoint split over 5 training runs and the framework was piloted in a mobile pharmacy application with 12 users and 19 sessions. For the keystroke module using key release intervals equal error rate of 11.1% was obtained (identification accuracy 74.4%) when key hold times and press latencies were added. For copy-paste heuristic requiring three consecutive intervals below 50 ms false positives on 20,400 genuine typing samples dropped from 16.3% to 0%. These results show that the proposed multimodal approach can support forensic readiness of mobile pharmacy applications by generating forensic-relevant interaction traces but more testing on a larger user population is needed to validate the scroll module, fusion layer and external validity of the findings.

Keywords: Forensic-by-Design; Insider Threat; Keystroke Dynamics; Multimodal Behavioral Biometrics; Mobile Pharmaceutical Security;

INTRODUCTION

Digital transformation of the healthcare sector has been enormous over the past decade with the widespread adoption of e-health services, telemedicine, electronic medical records and mobile health applications that have slowly replaced traditional healthcare interactions (Stoumpos et al., 2023). Such digital pharmaceutical and healthcare platforms are repositories of highly sensitive and high-value data assets, including medical records, medical histories and even patient identities. If not properly managed with good security governance, they pose very high levels of risk. This strategic imperative of health data has made this sector a key target for cybercrime worldwide.

The healthcare industry is rapidly becoming a target for cyberattacks as it means cyber attacks can result in the loss of operations or patient information (He et al., 2021). A literature mapping study of international publications during the COVID-19 pandemic shows that phishing, ransomware, DDoS attacks and malware are the most common cyberattack methods against the healthcare sector, and access management and incident response capacity in the sector is lacking (He et al., 2021).

Indonesia illustrates these risks. In 2021 personal data of about 279 million Indonesians linked to BPJS Kesehatan (national health insurance agency) were leaked and traded on a hacker forum due to IT security weakness and lack of continuous monitoring (Lubis, 2024). Other government-owned health applications such as

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

electronic Health Alert Card (eHAC) and vaccine certificate services have also been hacked (Firmansyah et al., 2022). These incidents show that the national digital health infrastructure is not safe.

To continue with this ever-changing threat landscape, most of the research and development of healthcare-pharmaceutical applications still focuses on reactive prevention, such as data encryption and static credential-based authentication systems. It's not new; a study of dozens of online pharmacy websites found that the majority of these platforms leak sensitive user data, namely, prescription drugs searched for or ordered, to third-party analytics services and that privacy policies rarely disclose these practices transparently (Rauti et al., 2024). This lack of post-incident preparedness has a direct impact on the investigation process for a healthcare application when the healthcare application is breached. Forensic investigations are slow, expensive and could produce digital evidence that is not legal because the system was not originally designed to legally record digital footprints.

This gap was identified and led to the Digital Forensic Readiness (DFR) paradigm which states that organisations need to use digital data to its full potential and reduce the cost of investigating them, and this framework has been evolving through modern DFR maturity models (Bankole et al., 2022). This idea later evolved into Forensic-by-Design (FbD) where forensic requirements are baked into every component in the system design process and the goal is to have a system that is forensically ready through continuous forensic readiness assessment (Akilal & Kechadi, 2022): Passive recording of biometric data and user interactions from very beginning of operation so that when an incident happens traces can be used as evidence for analysis.

In this study, an insider threat refers to the misuse of legitimate access by an authorized user, such as manipulating stock records, misusing prescriptions, or diverting drugs. Because unusual typing, scrolling, or copy-paste behavior can also arise from legitimate variation, this study distinguishes four levels: (1) normal behavioral variation; (2) anomalous behavior, defined as a statistical deviation from a learned profile; (3) suspicious behavior, defined as an anomaly for which the fused session score crosses the decision threshold; and (4) a confirmed insider threat, which can only be established through subsequent human investigation. The proposed framework operates at the second and third levels. It flags suspicious sessions and preserves their traces, but it does not determine intent or attribute an attack.

At the technical level, behavioral biometric research as a basis for such monitoring has a number of limitations. A comprehensive review of the behavioral biometrics literature on mobile devices shows that most active authentication systems are still built on a single modality such as touch dynamics, keystroke dynamics, or gait recognition, rather than adding another modality (Abuhamad et al., 2021; Stylios et al., 2021). This unimodal tendency is exacerbated by the bias of testing in relatively stable environments; recent work on Hand Movement, Orientation, and Grasp (HMOG), a postural noise rarely accounted for in systems tested only in desktop environments. Training on time series approaches such as Long Short-Term Memory (LSTM) only recently started on large scale text-free keystroke data for biometric authentication (Acien et al., 2021). If there is a memory on mobile devices than it can contribute to accuracy. Multimodal fusion has been shown to improve passive authentication. For example, score-level fusion of typing, scrolling, and other touch and sensor signals reduced error rates compared with individual modalities (Stragapede et al., 2022). However, such work targets identity verification on benchmark data rather than anomaly detection with forensic logging in an operational domain. Smartphones also impose computational constraints, and real target-domain data are often scarce (Wagata & Teoh, 2024). Transfer learning offers a way to address data scarcity, since pretraining on auxiliary data and adapting to a target application has improved accuracy when only limited target data are available (Salunke et al., 2022; Wagata & Teoh, 2024)

These gaps are addressed in this study: we develop and evaluate a Forensic-by-Design framework that uses multimodal behavioral biometrics to detect anomalous interactions in a mobile pharmacy application and record them as forensic relevant traces. The novelty of our approach is not in the algorithms (each of which has been studied before) but in their integration for this domain: first keystroke, scroll and copy-paste signals are combined into a session-level decision whose module outputs and scores are logged according to FbD principles so the purpose of behavioral biometrics changes from login-time authentication to continuous anomaly detection oriented towards forensics; second the framework is adapted to the target domain by pretraining on public benchmark datasets then fine tuning on real interaction data collected from a mobile pharmacy application; third model complexity is matched to each signal (a recurrent network only for keystroke sequences where temporal dependencies matter while tree based and rule based modules do scroll and copy paste behaviour). As previous work has shown (Salunke et al., 2022) transfer learning improves accuracy when target data are scarce this paper asks whether fine-tuning does so too in a mobile pharmacy setting with a small user population. The rest of the paper is structured as follows: Section II discusses related work, Section III describes our method, Section IV presents results, Section V discusses them and Section VI concludes the paper.

LITERATURE REVIEW

Behavioral biometrics authenticate users by what they do with a device (as opposed to what they know or have). Surveys of mobile behavioral biometrics list touch dynamics, keystroke dynamics and motion sensors as the most

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

studied modalities but most systems still use only one modality (Abuhamad et al., 2021; Stylios et al., 2021). For keystroke dynamics recurrent and attention-based models are now dominant: TypeNet (LSTM trained on keystroke sequences from more than 100,000 subjects) achieved EERs of 2.2% on physical keyboards and 9.2% on touchscreen keyboards (Acien et al., 2022); TypeFormer (transformer with LSTM layers) reached an EER of 3.25% on Aalto mobile keystroke database after five enrolment sessions (Stragapede, Delgado-Santos, et al., 2022). Several studies show that combining modalities improves performance. Score-level weighted fusion of typing, scrolling, drawing and tapping with background sensor data yielded EERs of 4–9% on HuMIdb database (Stragapede, Vera-Rodriguez et al., 2022). BehaveFormer fused keystroke and inertial sensor data through spatio-temporal dual attention transformers to get an EER of 2.95% on HuMIdb (Senarath et al., 2023). All these studies focus on authentication on public benchmarks; they do not address anomaly detection in an operational application or use behavioral traces as forensic records.

Behavioral biometric models require many samples per user which are rarely available when a system is first deployed. Transfer learning reuse knowledge learned from other users or domains: A feed-forward network pretrained on other users increased accuracy by 9.76 percentage points over training from scratch for new users in a five-user keystroke and mouse study (Salunke et al., 2022). Cross domain transfer between smartphone applications plus meta learning and feature covariance augmentation achieved state of the art touch stroke authentication with reasonable computational burden on smartphones (Wagata et al., 2024). Synthetic data generation is another strategy: hybrid generative model reduced enrolment requirement by up to 75% across nine behavioral modalities (Abdullah et al., 2025). The success of these strategies depends on how closely the source data resemble the target domain (which has not been tested for pharmacy applications yet).

Insider threats occur when authorized users abuse their legitimate access. In enterprise settings, a SIEM-integrated prototype combined system logs from the CERT insider threat dataset with mouse dynamics from Balabit dataset using per-modality one-dimensional convolutional networks (ROC-AUC values of 0.936–0.949), about 70,500 windows per second on CPU for inference throughput (Mohamed et al., 2026). The most studied insider threat in healthcare is drug diversion: A supervised machine learning model trained on 27.9 million medication movement transactions from ten hospitals detected 22 known blinded diversion cases a median of 74 days earlier than existing methods (Knight et al., 2022). This method only looks at transaction records after the fact — it doesn't capture how a user interacts with the application or whether the person performing a transaction is the account owner.

Digital Forensic Readiness (DFR) tries to maximize evidential value of data while minimizing investigation cost; maturity models have been proposed to measure how ready an organization is (Bankole et al., 2022). Forensic-by-Design (FbD) goes further by embedding forensic requirements into system design: An improved FbD framework for cloud computing aligned these requirements with systems engineering standards (Akilal & Kechadi, 2022). Readiness also depends on how evidence is stored. A secure storage model combining integrity checks, sandboxing, strong encryption and two-factor authentication showed that protecting forensic artifacts collected before and after an incident is possible at low cost (Singh et al., 2022). These works specify how evidence should be collected and protected but do not say which behavioral signals need to be recorded or how they can be analyzed to flag suspicious sessions.

Comparison and research gap Table 1 compares representative studies along dimensions relevant to this work. Three gaps are apparent: Behavioral biometric studies report authentication accuracy on public benchmarks but rarely real world validation or computational cost; insider threat and drug diversion studies analyze logs or transactions not interaction behavior; forensic readiness studies collect evidence but do not behavioral analysis. No study reviewed here combines interaction-level behavioral biometrics with session-level anomaly flagging and forensic oriented logging in a mobile pharmacy application. This study does. Its contribution is the integration (and evaluation) not any single algorithm.

METHOD

This section describes the research design, data acquisition, preprocessing, model architecture, training scheme, fusion layer, and evaluation protocol.

A.Research Design and Flow

The keystroke and scroll modules are formulated as user verification: for each session the module estimates how likely it belongs to the user whose account is active. The copy-paste module flags input patterns that cannot be from manual typing (e.g., no shift+t). Module outputs are fused in a fusion layer into a session level score which classifies each session as AUTHENTIC or SUSPICIOUS (third level of behavioral hierarchy introduced in Introduction). Everything from data acquisition through log storage as forensic artifact can be seen in Fig 1.

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

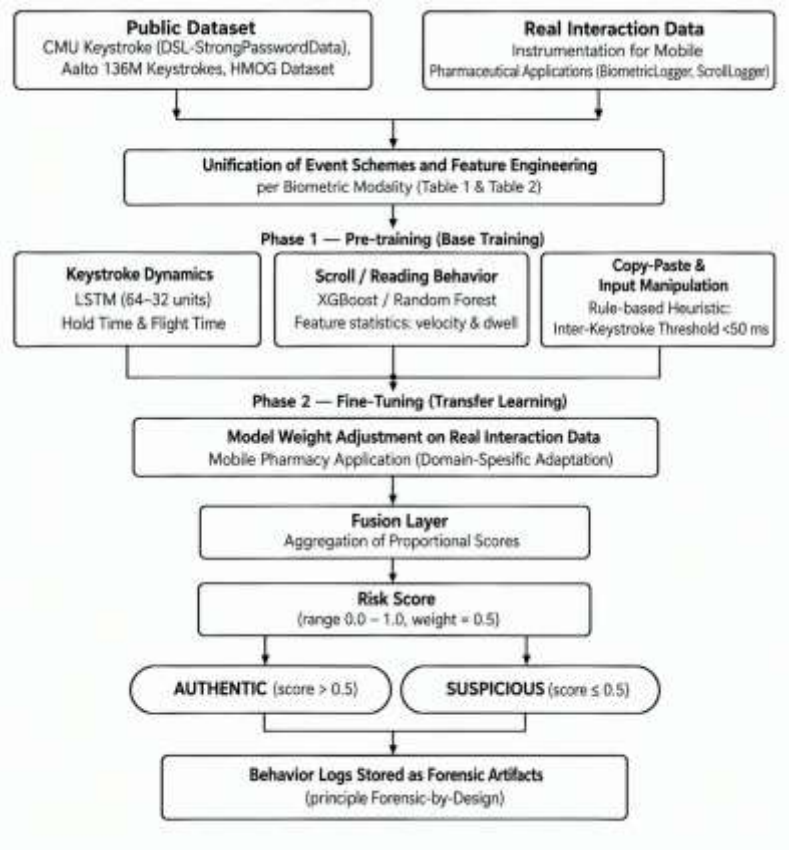


Fig 1. So that it matches this flow

B. Data Acquisition

1. Public Dataset for Basic Training (Phase 1)

Public datasets were selected based on their modality suitability for each trained module (Table 1). The keystroke datasets used were the CMU Keystroke Dynamics Benchmark (also known as DSL-StrongPasswordData) collected from 51 subjects typing a fixed password (".tie5Roanl") 400 times over 8 sessions, yielding hold time and down-down latency per letter (Killourhy & Maxion, 2009), and the Aalto 136M Keystrokes, a very large free-text typing corpus involving hundreds of thousands of participants (Dhakal et al., 2018). For the scrolling modality, the HMOG (Hand Movement, Orientation, and Grasp) Dataset was used, which records touch and motion features on a smartphone in various user postures (sitting, walking, and leaning) (Sitová et al., 2016). All three datasets were converted into a unified event schema (session_id, user_id, event_type, event_timestamp, raw_json) which matches the application's production log format (so public and real-world data go through the same preprocessing pipeline). These are benchmark datasets that are cited in papers and used only for base training.

Table 1. Datasets Used

Dataset	Modality	Key Characteristics	Usage Phase
CMU Keystroke Benchmark (DSL-StrongPasswordData) (Killourhy & Maxion, 2009)	Keystroke (fixed-text)	51 subjects, 8 sessions × 50 password repetitions	Phase 1 – Base Training
Aalto 136M Keystrokes (Dhakal et al., 2018)	Keystroke (free-text)	Large-scale population dataset, natural key press–release latency	Phase 1 – Base Training

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

HMOG Dataset (Sitova et al., 2016)	Touch & Motion	Smartphone touch and motion features under sitting/walking conditions	Phase 1 – Base Training
Real Interaction Data from the Mobile Pharmacy Application	Keystroke, Scroll, Copy-Paste	[N] participants, [S] sessions, [E] events; keydown/keyup, scroll, and input-source logs from routine and scripted scenarios	Phase 2 – Fine-Tuning

2. Real Interaction Data Acquisition in Mobile Pharmaceutical Applications

The real-time data are obtained through direct instrumentation on the mobile pharmacy application interface by a dedicated logger module embedded on the client side. The BiometricLogger module records every keydown and keyup event along with its timestamp to calculate Hold Time (the time difference between pressing and releasing the same key) and Flight Time (the time difference between releasing the previous key and pressing the next key). The ScrollLogger module records the scroll position (scroll_y), scroll speed (pixels/second), viewport progress percentage, and dwell time per page section with a sampling frequency of 20 Hz (throttle interval 50 ms). Scroll speeds below the 50 pixels/second threshold are considered as reading behavior, including specifically the drug side effect information section, which is relevant to patient compliance and safety in the pharmacy application. All events are exported in CSV format with event-level resolution and stored as a real-time data corpus for the fine-tuning stage.

Text insertions are also labeled by source (typed, pasted, autofill or keyboard suggestion) using browser input-type information. Data were collected from 12 participants between 9 and 12 july 2026 on 12 devices running smartphone. Each participant completed at least [k] sessions of each of two routine scenarios: (i) searching for and entering drug information (keystroke data) and (ii) reading drug information pages (scroll data). To provide ground truth for the copy-paste module participants also completed [m] scripted sessions in which drug quantities or prescription details were entered by pasting or by an automated typing script; each scripted session was labelled at collection time. Altogether there are [S] sessions and [E] interaction events.

All participants gave informed consent before data collection. The logger only records event types, timestamps and positional values (not characters typed) and user and session identifiers were pseudonymized before analysis

C. Preprocessing and Feature Engineering

Events were grouped by session and features were extracted for each modality (Table 2). For keystroke dynamics inter-key intervals between consecutive keyup events were computed, padded or truncated to a fixed length of 20 intervals and standardized with statistics computed on the training partition only. Scrolling: A 10 dimensional session level vector was extracted: average, maximum, minimum and standard deviation of velocity; total scroll distance; number of scroll events; number of direction changes; average interval between scroll events; maximum viewport progress; scroll range. Each feature was standardized in the same way. For copy-paste module inter-key intervals and input source of each text insertion were retained.

Table 2. Features by Biometric Modality

Modality	**Extracted Features**	**Granularity**
Keystroke Dynamics	Inter-key interval (time difference between consecutive keyup events)	Per-session sequence (fixed dimension, padded)
Scroll/Reading Behavior	avg/max/min/std velocity; total scroll distance; n_scroll_events; n_direction_changes; avg_time_between_scrolls; viewport_progress_max; scroll_range	Per-session aggregate vector (10 dimensions)
Copy-Paste & Input Manipulation	Inter-key interval (ms) and input source	Per-sequential keystroke pair

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

D.Data partitioning and Augmentation

Since the number of real sessions is small we partitioned at session level. Performance was estimated with stratified group k-fold cross validation (k = [5]) using each original session as a group so that no part of a session appears in both training and validation folds; another [15]% of original sessions per participant were held out as an independent test set and used only once for final evaluation. Within each training fold only, the original sessions were augmented by generating [50] synthetic copies per session with Gaussian noise added to event timestamps (σ = 5 ms), scroll velocity (σ = 0.05) and scroll position (σ = 10 px). Validation and test partitions contain only original non-augmented sessions.

E. Model Architecture and Configuration

We choose the architecture of each module according to the signal characteristics of each modality, considering the balance between accuracy and computational load on mobile devices, in line with the novelty of the hybrid architecture proposed in the Introduction (Table 3).

Table 3. Model Architecture and Configuration per Module

Modale	Architecture	Training Configuration
Keystroke Dynamics	Two-layer LSTM (64 → 32 units) + Batch Normalization + Dropout + Dense(64, ReLU) + Dense(Softmax) (Soni & Prabakar, 2022)	Optimizer Adam (lr = 0.001), categorical cross-entropy, 20 epochs, batch size 64, EarlyStopping (patience = 5), ReduceLROnPlateau
Scroll/Reading Behavior	XGBoost Classifier(Estrela et al., 2021)	n_estimators = 100, max_depth = 4, learning_rate = 0.1
Copy-Paste & Input Manipulation	Rule-based threshold heuristics	Flagged when ≥ 5 consecutive inter-key intervals fall below 50 ms, or when pasted content enters a sensitive field; autofill and keyboard suggestions excluded

1. Keystroke Dynamics Module

   This module is implemented as a two-layer LSTM network that processes sequences of keystroke temporal features on each session using a time series approach that has been found to be effective for large-scale keystroke biometrics (Soni & Prabakar, 2022). The first LSTM layer (64 units) returns the entire sequence to the second LSTM layer (32 units) which then applies Batch Normalization and Dropout to avoid overfitting the sequence and then classifies it in a Dense layer with softmax activation function.

2. Modul Scroll/Reading Behavior

   This module classifies scroll statistical feature vectors using the gradient-boosted trees (XGBoost) algorithm. A tree ensemble was used because the scroll features are tabular session-level aggregates so they fit well with models like this.

3. Module Copy-Paste and Input Manipulation module

   This module is a fixed rule rather than a trained model. Insertions from autofill, password managers and keyboard word suggestions are not considered. Pasted content is flagged only when it enters sensitive fields (dosage, quantity, prescription or stock). A typing sequence is auto-injected if at least five consecutive inter-key intervals are below 50 ms long. A run of intervals is required not just one interval since occasionally sub-50ms intervals occur in fast human typing due to key rollover. The module was evaluated on the labeled scripted sessions and sensitivity analysis was done over thresholds 30–100 ms and run lengths 3–7 intervals.

F. Two-Phase Training Scheme (Transfer Learning)

In Phase 1 (base training), the keystroke and scroll modules were trained on public datasets (Table 1). In Phase 2 (fine-tuning) we adapted the Phase 1 models to real-world data: for the keystroke module the output layer was replaced by that of enrolled users; LSTM layers were frozen while new output layer was trained, then unfrozen and trained jointly with a reduced learning rate [0.0001], and for the scroll module boosting was continued from the Phase 1 model with [n] additional trees on real-world training data. Transfer learning has been shown to improve accuracy when target-domain data are limited (Salunke et al., 2022; Wagata et al., 2024). To quantify its contribution in this study each module was also trained from scratch on the same real-world training folds with identical architecture and hyperparameters and each module was

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

evaluated in three conditions: Phase 1 model without adaptation, model trained from scratch and fine-tuned model.

G.Fusion Layer and Decision-Making Scheme

Sessions were evaluated as verification trials: A genuine trial pairs a session with its true user and an impostor trial pairs the same session with every other enrolled user as the claimed identity. The score of each verification module is the predicted probability of the claimed user; copy-paste module contributes 1 if no anomaly was found and 0 otherwise. Module scores are weighted averages (weights proportional to mean validation accuracy for each module across folds). This scheme is transparent and reproducible so logged decisions can be audited. The decision threshold was set at equal error rate on validation folds and applied unchanged to test set. Sessions scoring above threshold are AUTHENTIC; all others are SUSPICIOUS. We compared our proposed fusion with equal weight averaging and tested threshold sensitivity over [0.1,0.9]. Each decision is stored as a log entry with pseudonymized session and user identifiers, timestamp, per-module scores, fusion weights, fused score and verdict. These logs are for later forensic analysis (integrity mechanisms like hashing and tamper evident storage, formal chain-of-custody procedures etc are not addressed here).

H.Evaluation Protocol

All metrics were computed directly from model predictions (no post-hoc adjustment): accuracy, precision, recall, F1-score, false acceptance rate (FAR), false rejection rate (FRR), equal error rate (EER), ROC-AUC, PR-AUC and confusion matrices. Results are reported as mean ± standard deviation across folds and final performance is reported on the independent test set. The three training conditions were compared using Wilcoxon signed rank test on fold level scores. Model size, inference latency per session (median over [N] runs) CPU and RAM usage were measured on [server specification]; overhead of client-side logger was measured on [phone model] with browser developer tools.

RESULT

A. Pilot development Data

The framework was deployed in a web-based mobile pharmacy application between 9 and 12 july 2026. During this pilot we recorded 19 original sessions of 12 users (425 interaction events; median session duration 66 s range 10–359 s). Scroll events were recorded in 15 sessions (175 events) and keystroke events in one session (22 keyup events). Most users only contributed one session and no known malicious activity sessions were recorded. Pilot data therefore do not support an estimate of detection accuracy: they were used to check the logging pipeline and examine copy-paste rule on genuine typing (Section IV-C).

Table 4. Pilot Deployment Data

Item	Value
Collection period	9-12 July 2026
users/original sessions	12/19
Interactions events	425
Sessions with scroll events (events)	15 (175)
Sessions with keystroke events(keyup events)	1(22)
Median session duration (range)	66 s(10-359s)
Sessions with know malicious activity	0

B. Keystroke Module on the CMU Benchmark

The keystroke module was tested on the CMU benchmark with sessions 1–5 for training, session 6 for validation and threshold selection and sessions 7–8 for testing (51 subjects; 12,750 / 2,550 / 5,100 samples). Training was repeated five times using different random seeds. Results are given in Table 5. Using inter-key intervals between consecutive keyup events as in the deployed module we obtained 51.1 ± 9.3% correct subject identification rate (95% CI 39.6–62.6%) on test samples; verification EER of 11.1 ± 2.5% (95% CI 8.0–14.2%); ROC-AUC 0.956 ± 0.016. At the threshold selected on the validation session false acceptance rate was 8.6 ± 2.0% and false rejection rate 14.3 ± 3.4%. Performance varied a lot across seeds (test accuracy 41.7–63.0). If we instead used hold time, keydown–keydown latency and keyup–keydown latency (which

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the logger already records), then accuracy was 74.4 ± 0.8% (95% CI 73.4–75.4%), EER was 5.5 ± 0.2% (95% CI 5.3–5.8%), ROC-AUC was 0.987 ± 0.001, FAR was 3.6 ± 0.2% and FRR was 7.9 ± 0.3%. Variability across seeds was also lower. Fig. 2 show the For comparison Scaled Manhattan detector under the original benchmark protocol (Killourhy & Maxion, 2009) had an EER of 9.5% with all 31 features and 14.7% with keyup intervals only. But because that protocol tests each subject as a separate anomaly detector these values are meant to be used in context rather than as direct comparisons.

Table 5. Pilot Deployment Data

Features	Accuracy (%)	Precision(%)	Recall (%)	F1 (%)	EER (%)	ROC-AUC	FAR (%)	FRR (%)
Keyup intervals (deployed)	51,1± 9,3	51,1±9,9	51,1±9,3	49,3±9,8	11,1±2,5	0,956±0, 016	8,6±2,0	14,3±3,4
Hold + DD+UD	74,4±0,8	76,3±0,6	74,4±0,8	73,5±0,9	5,5±0,2	0,987±0, 001	3,6±0,2	7,9±0,3
Scaled Manhattan, 31 features(reference)*	0	0	0	0	0	9,5	0	0
Scaled Manhattan, keyup intervals(reference)*	0	0	0	0	0	14,7	0	0

*Original benchmark protocol (Killourhy & Maxion, 2009); not directly comparable. DD = keydown–keydown latency; UD = keyup–keydown latency.

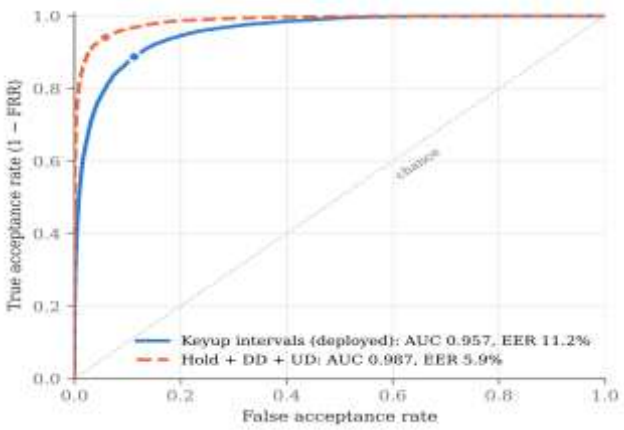


Fig. 2 ROC Curves of the Two Keystroke Feature Configurations (test sessions, run with seed 0; dots mark the EER)

C. Copy-paste Rule on Genuine Human Typing

To estimate false positives we applied the rule to all 20,400 genuine passwords in CMU dataset (204000 inter-key intervals). Only 1.9% of keyup–keyup were <50 ms but because each entry has ten intervals a rule that flags any single sub-50 ms interval would have flagged 16.3% of genuine entries (8.9% when using keydown–keydown intervals). Two consecutive sub-50 ms had to be required to reduce this to 0.15% and three to 0% (Table 6). Negative keyup–keyup was also present down to −145 ms since often a key was released after the next one pressed so key rollover is not producing very short intervals in normal typing. The

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

same pattern was seen in the pilot data too: 5 of 21 intervals in the single typing session were below 50 ms (minimum 33.7 ms), at most two consecutive so the single-interval rule would have flagged this genuine session but a run-length rule of three would not. The detection rate on real paste or injection events cannot be determined because no labeled sessions of that kind were recorded.

Table 6. Percentage of Genuine CMU Entries Flagged by the Copy-Paste Rule (keyup–keyup intervals)

Threshold /minimum run length	1	2	3	5
30 ms	**5,92**	**0,01**	**0,00**	**0,00**
50 ms	**16,26**	**0,15**	**0,00**	**0,00**
70 ms	**35,87**	**1,42**	**0,02**	**0,00**
100 ms	**68,16**	**15,02**	**2,11**	**0,00**

D. Computational Cost

Table 7 shows the computational cost of the deployed models (measured over 100 inference runs on a single real session on a MacBook Pro with an Apple M1 chip (8-core CPU), 8 GB RAM running macOS 26.6.2, Python 3.12.4 and TensorFlow 2.21.0): keystroke model (about 35,000 parameters) required median 35.3 ms per session (95th percentile 38.6 ms); scroll model 0.18 ms (0.22 ms). Logger overhead on mobile devices and energy consumption were not measured. The scroll module could not be evaluated because pilot data have no participant more than four scroll sessions and HMOG dataset was not used in this study. Fusion layer could also not be evaluated because it requires sessions where all modules produce scores and sessions with known suspicious activity. Both are reported as design components whose validation is left to future work.

Table 7. Computational cost of the deployed models

Model	SIze(KB)	Median latency(ms)	P95 latency(ms)
Keystroke (LSTM)	**448,8**	**35,3**	**38,6**
scroll	**71,5**	**0,18**	**0,22**

E. Modules not evaluated

The scroll module could not be evaluated: the pilot data contain no participant with more than four scroll sessions, and the HMOG dataset was not used in this study. The fusion layer could likewise not be evaluated, because it requires sessions in which all modules produce scores and sessions with known suspicious activity. Both are reported as design components whose validation is left to future work.

DISCUSSIONS

The most consistent finding of this study is about the feature representation of the keystroke module. With keyup–keyup intervals only (as used in the deployed module) the LSTM had EER 11.1% and correctly classified about half of test samples; with hold time, keydown–keydown latency and keyup–keydown latency it was 5.5% and accuracy was 74.4%. A possible explanation for this is that each keyup–keyup interval is the sum of a latency and the difference between two consecutive hold times. Different typing patterns can produce the same interval but ten intervals are less informative than 31 separate timing features. The keyup-only configuration was also less stable: test accuracy varied from 41.7% to 63.0% across random seeds and training took longer (60 vs 39 on average); richer configuration varied by less than two percentage points. The same pattern occurred with the Scaled Manhattan detector; EER increased from 9.5% to 14.7% when restricted to keyup intervals. Because the logger already records keydown and keyup events this has a direct practical consequence: the deployed module can be improved by changing its feature extraction without additional instrumentation.

Work The EER of 5.5% obtained with the richer feature set is in line with what has been reported for mobile behavioral biometrics (score-level fusion of several touch and sensor modalities gave EERs of 4–9%; Stragapede et al., 2022). It is higher than 2.2% reported by TypeNet for free-text typing on physical keyboards (Acien et al., 2022). That result was however with much longer sequences and a population scale training set whereas here we used a ten character password and only 51 subjects. Also beware when comparing to Scaled Manhattan detector.

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

That detector models each subject separately as an anomaly detector whereas LSTM is a closed set classifier that can verify users enrolled during training. So its lower EER is not due to better model but different task setting.

The analysis of genuine typing shows that a fixed 50 ms threshold applied to single intervals (as originally planned) is not an effective signal for non-human input: it would have flagged 16.3% of genuine CMU entries and the only session recorded in the pilot. Very short (even negative) intervals are normal typing since users often press next key before releasing previous one. Three consecutive sub-50 ms intervals were needed on the benchmark data to eliminate these false positives so we can treat run of fast intervals (not a single interval) as the anomaly signal. This also illustrates the distinction drawn in the Introduction between normal variation of behavior and anomalous behavior: a statistically unusual interval by itself is not evidence of manipulation. Two caveats apply. First, the rule targets automated typing; paste operation produces no per-key events and has to be detected from the input source recorded by the logger. Second, the detection rate on genuine paste or injection events remains untested because no labeled sessions of this kind were collected.

The framework was designed according to the Forensic-by-Design principle that interaction traces should be recorded from the beginning (Akilal & Kechadi, 2022). In the pilot deployment we found that the logger captures and structures keystroke, scroll, signature and image-interaction events with timestamps and session identifiers; each module output can be stored along with its score. Traces like these can help investigators reconstruct what happened in a session and prioritize sessions for review. But storing behavioral logs does not by itself make forensic readiness. The current implementation does not provide integrity protection (hashing or tamper-evident storage), timestamp reliability beyond the client clock, chain-of-custody procedures or access control for stored logs (none of these properties were evaluated). The framework should therefore be understood as supporting forensic readiness by generating forensic relevant interaction traces but not as a complete Forensic-by-Design implementation. Also, SUSPICIOUS verdict means that session deviates from learned profile of claimed user. It does not tell you who the actor was, whether the actor acted maliciously or if there was an insider attack; those questions must be answered by human investigation afterwards.

This study has several limitations: The pilot deployment had 12 users and 19 sessions in a single application so accuracy on real pharmacy interactions could not be evaluated; the keystroke module was evaluated on CMU benchmark collected on desktop keyboard with one fixed password (performance on mobile touchscreen keyboards and free-text input may vary); because it is a closed-set classifier it was not tested on users unseen during training nor externally validated on second dataset; scroll module and fusion layer could not be evaluated; contribution to detection of scroll module and fusion layer is unknown; copy-paste rule was only evaluated for false positives; its detection rate on real manipulation is unknown; computational cost was measured on laptop not on mobile devices; energy consumption and logger overhead were not measured; results previously obtained with synthetic augmentation of pilot data were withdrawn and no augmentation was used in reported experiments; integrity and admissibility of stored logs as digital evidence were not validated.

CONCLUSION

Misuse of legitimate access (stock manipulation or prescription misuse) in mobile pharmacy applications is hard to prevent with login authentication and encryption alone (and digital traces are usually only collected after an incident). We proposed a framework that records behavioral interaction traces from the start of operation and analyses them with three modules (LSTM for keystroke dynamics, gradient boosted tree classifier for scrolling behavior and rule for automated input detection). Three main conclusions were drawn from evaluation. First on CMU benchmark the keystroke module had EER of 11.1% when keyup intervals only were used (as in deployed module), and 5.5% when hold time and key latencies were included. Since logger already records these events then the deployed module can be improved by changing its feature extraction only. Second a single sub-50 ms interval is not a good indication of non-human input: this rule would have flagged 16.3% of genuine entries but required three short intervals consecutively to produce no false positives on benchmark data. Third, a pilot deployment in a web-based mobile pharmacy application showed that the logging pipeline captures and stores interaction events as expected but only 19 sessions were collected to evaluate detection accuracy on real pharmacy interactions. The framework therefore provides forensic readiness by detecting behavioral anomalies (so investigators can reconstruct sessions and prioritize them), generates forensic-relevant interaction traces (so they have something to work with) but does not establish who did what (or if there was an attack), whether malicious intent was present or not, etc., nor has stored logs been validated against digital evidence requirements.

Future work should: collect a larger real world dataset from pharmacy staff (multiple sessions per user and labeled misuse scenarios) so that scroll module, copy-paste rule and fusion layer can be evaluated; evaluate keystroke module on mobile touchscreen typing and on users not seen during training; measure latency, logger overhead and energy consumption on a range of mobile devices; add integrity protection, reliable timestamping and access control for the stored logs and validate them against forensic standards (ISO/IEC 27043).

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (Kementerian Pendidikan Tinggi, Sains, dan Teknologi/Kemdiktisaintek) for the financial support provided through the BIMA Research Funding Program Year 2026, which has significantly contributed to the successful completion of this research and the preparation of this article. The authors also appreciate all parties who have provided valuable assistance, support, and constructive feedback during the research process, as their contributions have been instrumental in the completion of this work.

REFERENCES

Abdullah, A. B., et al. (2025).Advancing Multi-Modal Behavioral Biometric Authentication: A Deep Learning Approach with Synthetic Data Generation. IEEE Access

Abuhamad, M., Abusnaina, A., Nyang, D., & Mohaisen, D. (2021). Sensor-Based Continuous Authentication of Smartphones' Users Using Behavioral Biometrics: A Contemporary Survey. *IEEE Internet of Things Journal*, *8*(1), 65–84. https://doi.org/10.1109/JIOT.2020.3020076

Acien, A., Morales, A., Monaco, J. V., Vera-Rodriguez, R., & Fierrez, J. (2022) TypeNet: Deep learning keystroke biometrics. IEEE Transactions on Biometrics, Behavior, and Identity Science, 4(1), 57–70. https://doi.org/10.1109/TBIOM.2021.3112540.

Akilal, A., & Kechadi, M.-T. (2022). An improved forensic-by-design framework for cloud computing with systems engineering standard compliance. *Forensic Science International: Digital Investigation*, *40*, 301315. https://doi.org/10.1016/j.fsidi.2021.301315

Bankole, F., Taiwo, A., & Claims, I. (2022). An extended digital forensic readiness and maturity model. *Forensic Science International: Digital Investigation*, *40*, 301348. https://doi.org/10.1016/j.fsidi.2022.301348

Dhakal, V., Feit, A. M., Kristensson, P. O., & Oulasvirta, A. (2018). Observations on Typing from 136 Million Keystrokes. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3173574.3174220

Estrela, P. M. A. B., Albuquerque, R. D. O., Amaral, D. M., Giozza, W. F., & Júnior, R. T. D. S. (2021). A Framework for Continuous Authentication Based on Touch Dynamics Biometrics for Mobile Banking Applications. *Sensors*, *21*(12), 4212. https://doi.org/10.3390/s21124212

He, Y., Aliyu, A., Evans, M., & Luo, C. (2021). Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review. *Journal of Medical Internet Research*, *23*(4), e21747. https://doi.org/10.2196/21747

Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. *2009 IEEE/IFIP International Conference on Dependable Systems & Networks*, 125–134. https://doi.org/10.1109/DSN.2009.5270346

Knight, T., May, B., Tyson, D., McAuley, S., Letzkus, P., & Enright, S. M. (2022). Detecting drug diversion in health-system data using machine learning and advanced analytics. *American Journal of Health-System Pharmacy*, *79*(16), 1345–1354. https://doi.org/10.1093/ajhp/zxac035

Mohamed, M. S., Et Al. (2026). A SIEM-Integrated Cybersecurity Prototype For Insider Threat Anomaly Detection Using Enterprise Logs And Behavioural Biometrics. Https://Doi.Org/10.3390/Electronics15010248

Rauti, S., Carlsson, R., Mickelsson, S., Mäkilä, T., Heino, T., Pirjatanniemi, E., & Leppänen, V. (2024). Analyzing third-party data leaks on online pharmacy websites. *Health and Technology*, *14*(2), 375–392. https://doi.org/10.1007/s12553-024-00819-w

Singh, A. K., et al. (2022). Secure storage model for digital forensic readiness. doi:10.1109/ACCESS.2022.3151403.

Salunke, S. V., Ouda, A., & Gagne, J. (2022). Transfer learning for behavioral biometrics-based continuous user authentication. In 2022 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1–6). IEEE. https://doi.org/10.1109/ISNCC55209.2022.9851764

Stragapede, G., Delgado-Santos, P., Tolosana, R., Vera-Rodriguez, R., Guest, R., & Morales, A. (2022). TypeFormer: Transformers for mobile keystroke biometrics. Neural Computing and Applications.https://doi.org/10.1007/s00521-024-10140-2

Schmitt, V. (2022). Medical Device Forensics. *IEEE Security & Privacy*, *20*(1), 96–100. https://doi.org/10.1109/MSEC.2021.3127490

Senarath, D., Tharinda, S., Vishvajith, M., Rasnayaka, S., Wickramanayake, S., & Meedeniya, D. (2023). BehaveFormer: A Framework with Spatio-Temporal Dual Attention Transformers for IMU-enhanced

*name of corresponding author

This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 4, October 2026
DOI : https://doi.org/10.33395/sinkron.v10i4.16852
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Keystroke Dynamics. *2023 IEEE International Joint Conference on Biometrics (IJCB)*, 1–9. https://doi.org/10.1109/IJCB57857.2023.10448997

Sitova, Z., Sedenka, J., Yang, Q., Peng, G., Zhou, G., Gasti, P., & Balagani, K. (2015). *HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users*. https://doi.org/10.48550/ARXIV.1501.01199

Soni, J., & Prabakar, N. (2022). KeyNet: Enhancing Cybersecurity with Deep Learning-Based LSTM on Keystroke Dynamics for Authentication. In J.-H. Kim, M. Singh, J. Khan, U. S. Tiwary, M. Sur, & D. Singh (Eds.), *Intelligent Human Computer Interaction* (Vol. 13184, pp. 761–771). Springer International Publishing. https://doi.org/10.1007/978-3-030-98404-5_67

Stoumpos, A. I., Kitsios, F., & Talias, M. A. (2023). Digital Transformation in Healthcare: Technology Acceptance and Its Applications. *International Journal of Environmental Research and Public Health*, *20*(4), 3407. https://doi.org/10.3390/ijerph20043407

Stylios, I., Kokolakis, S., Thanou, O., & Chatzis, S. (2021). Behavioral biometrics & continuous user authentication on mobile devices: A survey. *Information Fusion*, *66*, 76–99. https://doi.org/10.1016/j.inffus.2020.08.021

Wagata, K., & Teoh, A. B. J. (2024). Cross-Domain Cross-Task Transfer Mobile Touch-Stroke Authentication. In ICASSP 2024 – 2024 IEEE International Conference On Acoustics, Speech And Signal Processing (ICASSP) (Pp. 4445–4449). https://doi.org/10.1109/ICASSP48485.2024.10445813

*name of corresponding author
This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.