УДК 004.056

Д.П. КУРНІЦЬКИЙ

МОДЕЛЬ ТА МЕТОД ОЦІНЮВАННЯ НАДІЙНОСТІ ЗРАЗКА КЛАВІАТУРНОГО ПОЧЕРКУ ДЛЯ ПОВЕДІНКОВОЇ АВТЕНТИФІКАЦІЇ КОРИСТУВАЧА

Вінницький Національний Технічний Університет,
вул. Хмельницьке шосе, 95, м. Вінниця, Україна, 21021

Анотація. Розглянуто задачу оцінювання надійності (якості) поточного зразка клавіатурного почерку (keystroke dynamics) у поведінкової автентифікації під час входу користувача. На відміну від робіт, що оптимізують рішення поведінкової автентифікації або злиття контексту й поведінки, у центрі дослідження — питання: «чи можна довіряти саме цьому зразку введення, щоб застосовувати поведінковий канал?» Запропоновано формальне означення надійності зразка як імовірнісної оцінки його придатності/корисності для біометричного зіставлення та наведено модель ознак деградації: повнота подій, ефективна довжина послідовності, варіативність таймінгів, індикатори автозаповнення/вставки, ознаки зміни пристрою й квантування часових міток. Розроблено метод інтегрального оцінювання надійності на основі логістичної моделі якості та показано, як надійність пов'язана зі зростанням помилок поведінкового верифікатора. Додатково розглянуто калібрування ймовірностей (Platt scaling, ізотонічна регресія, Bayesian binning) для перетворення «сирих» скорів у добре інтерпретовані ймовірності. Експериментальну перевірку виконано на відкритому еталонному датасеті DSL-StrongPassword (51 користувач, 400 введень пароля на користувача) із синтетичним моделюванням деградацій (втрата подій, укорочення послідовності, джиттер/квантування). Отримано, що відбір зразків за високими значеннями показника надійності підвищує якість поведінкового зіставлення (AUC зростає з 0.856 до 0.890 для підмножини зразків із $q \geq 0.8$ за покриття ≈53%), а також суттєво змінює профіль помилок, зменшуючи відмови для легітимних користувачів у «якісних» зразках. Наведено практичні рекомендації щодо порогів придатності та політик «використати/не використовувати» поведінковий канал у продакшн-системах.

Ключові слова: динаміка натискань клавіш; поведінкова біометрія; якість біометричного зразка; надійність зразка; калібрування ймовірностей; Platt scaling; ізотонічна регресія; ROC; EER.

Abstract. This paper addresses the problem of estimating the reliability (sample quality) of the current keystroke-dynamics sample in behavioral user authentication at login time. Unlike risk-based fusion or context–behavior score combination, the focus is: "Can this particular behavioral sample be trusted enough to use the behavioral channel?" We provide a formal definition of sample reliability as a probabilistic estimate of sample usability/utility for biometric comparison, and propose a feature model capturing degradation factors: event completeness, effective sequence length, timing variability, autofill/paste indicators, device-change signals, and timestamp quantization/jitter. An integral reliability estimation method based on a logistic quality model is developed, and the relationship between estimated reliability and the behavioral verifier's error is analyzed. We also discuss probability calibration (Platt scaling, isotonic regression, Bayesian binning) to convert raw scores into well-interpretable probabilities. Experimental validation is performed on the public DSL-StrongPassword benchmark dataset (51 users, 400 password typings per user) with controlled synthetic degradations (event loss, truncation, jitter/quantization). Results show that reliability-based filtering improves behavioral matching performance (AUC increases from 0.856 to 0.890 for samples with $q \geq 0.8$ at ≈53% coverage) and changes the error profile, reducing false rejects for legitimate users in high-quality samples. Practical deployment recommendations for reliability thresholds and "use/do-not-use" gating policies are provided.

Keywords: keystroke dynamics; behavioral biometrics; biometric sample quality; sample reliability; probability calibration; Platt scaling; isotonic regression; ROC; EER.

DOI: 10.31649/1681-7893-2026-51-1-24-32

ВСТУП

Поведінкова автентифікація за клавіатурним почерком (динамікою натискань клавіш) використовує часові характеристики натискань і відпускань клавіш як поведінкову біометричну ознаку для верифікації користувача під час введення пароля або тексту. Типові ознаки включають тривалість утримання клавіші (hold/dwell time) та інтерклавішні інтервали (flight time), а також похідні характеристики швидкості й помилок введення. [1]

Однак у реальних умовах один і той самий користувач може генерувати зразки дуже різної якості: частина спроб містить неповні події (втрата keyup/keydown), дуже короткі послідовності (короткий пароль або дострокове завершення введення), аномально «рівні» або аномально «шумні» таймінги, квантування часових міток, вплив іншої клавіатури/пристрою, віддалений доступ з джиттером та втратами пакетів тощо. Для таких випадків поведінковий канал може бути не стільки «поганим», скільки «непридатним для застосування саме в цій спробі». Вплив мережевого джиттера/packet loss на придатність клавіатурної біометрії для безперервної автентифікації у віддалених середовищах демонструвався окремими роботами. [2]

Саме тому у біометрії сформована окрема лінія стандартів і практик щодо якості (sample quality) біометричних зразків. Зокрема, стандарт ISO/IEC 29794-1 визначає призначення та інтерпретацію якісних скорів, мотивацію до нормалізації, агрегацію та оцінювання алгоритмів якості. [3]

Це дослідження формалізує й емпірично перевіряє ідею: перед тим, як використовувати клавіатурний почерк як канал поведінкової автентифікації, слід оцінити надійність поточного зразка. Така постановка забезпечує іншу наукову новизну порівняно з задачами злиття або оптимізації risk-based рішень.

Метою роботи є розроблення моделі та методу оцінювання надійності зразка клавіатурного почерку, а також експериментальна перевірка зв'язку між рівнем надійності зразка та ймовірністю помилки поведінкової автентифікації користувача.

АНАЛІЗ ПОПЕРЕДНІХ ДОСЛІДЖЕНЬ

Відомою основою для порівняння методів у верифікації за паролем є набір DSL-StrongPassword, де 51 користувач набирає один і той самий «сильний» пароль 400 разів; дані збирались у багатьох сесіях з точною часовою прив'язкою. Це дає репродуковану базу для дослідження залежності помилок від умов збирання та деградацій.

Інший клас даних — багатоклавіатурні та багатосесійні протоколи (наприклад, GREYC Keystroke), де 133 користувачі вводили фразу «greyc laboratory» у кількох сесіях і на різних клавіатурах; опубліковано розміри та протокол збирання (7555 зразків). Це важливо для моделювання «зміни пристрою» як джерела деградації.

Для безперервної/вільнотекстової поведінкової автентифікації існують неконтрольовані датасети: Clarkson II (101 файл на користувача, подієві логи натиск/відпуск із високою роздільністю часових міток) та великі набори на кшталт Aalto 136M, де дані й код публічно випущені для наукового використання.

Порівняльні огляди/метаопис датасетів також узгоджують розміри й умови збору Buffalo (SUNY Buffalo), де підкреслюються 157 учасників, 2.14M натискань і наявність підмножини із різними розкладками клавіатур.

У загальній біометрії питання якості зразка формалізується стандартами та тестовими підходами, де метрики помилок, throughput і правила звітності описані в ISO/IEC 19795-1. [4] У контексті клавіатурної динаміки, окрім «індивідуальності» сигналу, дослідження підкреслюють складність вільного тексту та варіативність як джерело високих помилок.

Ключова відмінність нашої роботи: ми не намагаємося безпосередньо «покращити класифікатор» у кожному випадку, а пропонуємо модуль надійності зразка, який визначає, чи взагалі зразок варто використовувати для поведінкового зіставлення у поточній спробі. [5, 15,16]

ПОСТАНОВКА ЗАДАЧІ ОЦІНЮВАННЯ НАДІЙНОСТІ ЗРАЗКА

Нехай зразок клавіатурного почерку, отриманий під час однієї спроби входу, задається послідовністю подій

$$S = \{e_j\}_{j=1}^{m}, \quad e_j = (k_j, t_j, \delta_j),$$

де k_j — код клавіші, t_j — часова мітка, $\delta_j \in \{down, up\}$ - вказує тип події, тобто натискання або відпускання клавіші.

У задачах поведінкової верифікації така послідовність S перетворюється у вектор ознак x ∈ R^d, який містить часові характеристики введення. До цього вектора можуть входити тривалості утримання клавіш, інтервали між натисканнями, диграфи, показники швидкості введення, а також ознаки, пов'язані з помилками під час набору. В оглядових і прикладних роботах такі характеристики зазвичай описують через dwell або hold time, flight time, швидкість введення та кількість помилок [6].

Метою є побудова функції надійності, або якості, зразка

$$q = Q(S) \in [0,1],$$

яка може тлумачитися як ймовірнісна оцінка того, наскільки цей зразок придатний для використання у поведінковій автентифікації саме в поточній спробі входу.

Щоб пов'язати поняття надійності з практичними задачами безпеки та перевірки користувача, введемо подію правильності поведінкового рішення:

$$C = 1\{\text{поведінковий верифікатор прийняв правильне рішення на зразку } S\},$$

Тоді бажаною є інтерпретація

$$q \approx P\big(C = 1 \mid \phi(S)\big)$$

де $\varphi(S)$ — вектор ознак, що описує деградацію або якість зразка. Такий підхід узгоджується з логікою стандарту ISO/IEC 29794-1, згідно з яким якісний скор має бути інтерпретованим, нормалізованим і придатним для оцінювання ефективності алгоритмів, що працюють із показниками якості [7].

Для практичної побудови вектора φ(S) у цій роботі пропонується прикладна модель ознак деградації, яка дає змогу відокремити інформативний зразок клавіатурного почерку від такого, що є малопридатним або непридатним для надійної поведінкової автентифікації. Ця модель задається не як жорстко фіксований перелік окремих чисел, а як система класів ознак, кожен із яких може реалізовуватися через один або кілька конкретних числових показників. Такий підхід є зручним для практичного впровадження, оскільки дозволяє адаптувати склад ознак до середовища збору даних, типу клієнтського застосунку та технічних обмежень платформи.

Першим важливим класом є повнота подій. У багатьох середовищах реєстрації, зокрема при роботі з браузерними подіями, у віддалених сеансах або при використанні фонової телеметрії, можливі втрати подій натискання чи відпускання клавіш, а також поява розривів у часовій послідовності. Такі втрати безпосередньо впливають на відтворюваність таймінгових ознак, оскільки частина інтервалів або взагалі не може бути обчислена, або обчислюється неточно. Особливо це стосується віддалених середовищ, де додатково діють джиттер, нестабільність каналу зв'язку та втрата пакетів, через що клавіатурна динаміка в окремих випадках може стати непридатною для достовірного використання [8].

Другим класом є ефективна довжина послідовності. Короткі зразки містять меншу кількість диграфів і триграфів, а отже, дають значно менше статистично стійкої інформації про індивідуальний стиль введення. У такій ситуації навіть незначні випадкові коливання сильніше впливають на результат верифікації. Крім того, важливе значення має правильне вирівнювання послідовностей, оскільки зсуви, пропуски або частково некоректне зіставлення окремих елементів можуть призводити до помітного зростання помилок. Експериментальні результати, наведені в літературі, показують, що саме погано вирівняні зразки здатні істотно погіршувати якість розпізнавання [9].

Ще одним важливим класом ознак є варіативність таймінгів. Надмірно велика розсіюваність часових характеристик може свідчити про шум, нестабільність введення або зміну умов збору даних. Водночас аномально мала варіативність також не є ознакою доброякісного зразка, оскільки вона може вказувати на неприродно рівномірну структуру таймінгів, автоматизацію введення або вставлення готового тексту. У практиці дослідження клавіатурного почерку до цього класу також відносять показники нерівномірності ритму введення та характеристики, пов'язані з помилками користувача під час набору [10].

Окрему групу формують ознаки автозаповнення або вставлення. У веб-формах поява тексту в полі може бути наслідком не реального введення з клавіатури, а автоматичного заповнення браузером, вставлення із буфера обміну або програмного встановлення значення елемента форми. У такому разі

зовнішньо система отримує рядок символів, однак часовий профіль такого рядка не відображає реальної моторної поведінки користувача. Специфікації UI Events визначають події keydown і keyup як події, пов'язані саме з натисканням клавіш, проте для сценаріїв автозаповнення реалізація подій залишається неоднорідною, а в обговореннях WHATWG прямо підкреслюється необхідність чітко визначати, які клавіатурні події мають або не мають генеруватися в режимі автодоповнення. Через це виявлення таких ситуацій є принципово важливою складовою оцінювання надійності зразка.

Наступним класом є ознаки, пов'язані зі зміною пристрою, клавіатури або розкладки. Наявні набори даних і результати попередніх досліджень показують, що навіть за незмінного користувача перехід на іншу клавіатуру або в інше середовище введення може помітно змінити часові характеристики набору. Саме тому в окремих роботах і наборах даних, зокрема в багатоклавіатурних протоколах на кшталт GREYC, спеціально досліджується вплив зміни апаратного середовища на значення метрик і рівень EER. Для задачі оцінювання надійності це означає, що нетиповість пристрою або конфігурації введення доцільно розглядати як окреме джерело деградації.

Нарешті, важливим класом залишаються ознаки квантування та джиттера часових міток. Якщо часові значення фіксуються надто грубо, мають великий крок дискретизації або надходять з непостійною затримкою, структура таймінгових ознак спотворюється. Подібний вплив мають також джиттер, втрати пакетів та нерівномірність передачі подій, особливо в сценаріях віддаленої роботи або у віртуалізованих середовищах. У таких умовах самі часові характеристики можуть уже не відображати справжню поведінку користувача, а радше особливості каналу передачі або програмної інфраструктури, що безпосередньо знижує придатність клавіатурного почерку для задач автентифікації [11].

У сукупності наведені класи ознак формують основу для побудови вектора $\varphi(S)$, який використовується надалі для оцінювання надійності поточного зразка. Саме через такий опис стає можливим перейти від загального ймовірнісного означення показника q до його практичного обчислення на основі характеристик, що безпосередньо відображають повноту, стабільність і достовірність зафіксованого поведінкового сигналу.

МОДЕЛЬ ТА МЕТОД ОЦІНЮВАННЯ НАДІЙНОСТІ

Нехай S_Rp — вектор ознак деградації поточного зразка клавіатурного почерку, який у базовому варіанті охоплює повноту c, тобто частку наявних часових подій і таймінгових ознак відносно очікуваних; ефективну довжину ℓ, що характеризує кількість придатних для аналізу значень утримання клавіш і диграфів; показники варіативності v, зокрема стандартне відхилення та коефіцієнт варіації для часових інтервалів; ознаку a, що відображає ймовірність автозаповнення або вставлення фрагмента тексту й визначається за евристиками на кшталт надто малої сумарної тривалості введення, неприродно малих інтервалів утримання клавіш або нехарактерно рівномірної структури таймінгів; а також ознаку k, що вказує на грубу квантизацію часових міток, тобто наявність такого великого кроку дискретизації, який може пояснити значну частину зафіксованих значень.

Тоді інтегральний показник надійності зразка доцільно визначити у вигляді

$$q = \sigma(w^T S\_Rp + b),$$

де $\sigma(\cdot)$ — логістична функція, а w і b — параметри, оцінені за навчальними даними. Таке подання є зручним не лише з обчислювального погляду, а й з погляду інтерпретації, оскільки дає змогу відобразити внесок окремих ознак деградації в єдину шкалу від 0 до 1. Обраний спосіб узгоджується із загальною логікою ISO/IEC 29794-1, де наголошується не на фіксованому алгоритмі оцінювання якості, а на необхідності нормалізації, інтерпретації та експериментальної перевірки ефективності показника якості.

Для практичного використання одного порога недостатньо, тому доцільно вводити три зони надійності. Якщо $q < \tau\_low$, поточний зразок слід вважати низьконадійним і непридатним для використання в поведінковому каналі. Якщо $\tau\_low \le q < \tau\_high$, зразок є обмежено придатним: його можна враховувати лише за обережної політики, наприклад із пониженим коефіцієнтом довіри або в поєднанні з іншими чинниками. Якщо ж $q \ge \tau\_high$, зразок слід вважати придатним для подальшого використання, а рішення поведінкового класифікатора може залучатися з мінімальним додатковим контролем. Отже, кінцеве призначення модуля полягає не просто в обчисленні числового показника, а у формуванні практичної рекомендації щодо того, чи варто взагалі використовувати поточний поведінковий зразок у процедурі автентифікації.

Окрему увагу слід приділити калібруванню ймовірностей. У реальних системах і модулі надійності, і поведінкові верифікатори часто повертають не безпосередньо ймовірності, а сирі оцінки у

вигляді відстаней, логітів, виходів дерев рішень або інших величин, які не мають властивості коректної ймовірнісної інтерпретації. Саме тому після обчислення первинного значення доцільно застосовувати процедури калібрування. Одним із класичних підходів є сигмоїдне калібрування Платта, що переводить початкові оцінки в ймовірнісну форму. Більш гнучким варіантом є ізотонічна регресія, яка не накладає жорсткої параметричної структури, але за малих калібрувальних вибірок може давати перенавчання. Ще одним підходом є байєсівське бінування за квантилями, яке враховує невизначеність у розбитті значень на інтервали та в оцінюванні частот. Якість такого калібрування доцільно перевіряти за допомогою критерію Brier score, а також через показники очікуваної й максимальної помилки калібрування та діаграми надійності, які відображають, наскільки близько передбачені ймовірності узгоджуються з емпіричними частотами.

Підсумкова процедура роботи модуля має такий вигляд: спочатку з множини подій однієї спроби входу виділяються звичайні ознаки клавіатурного почерку — тривалості утримання клавіш, інтервали між натисканнями, швидкість введення, кількість виправлень тощо; далі з них і з первинної часової послідовності обчислюються ознаки надійності, що характеризують повноту, довжину, варіативність, наявність ознак вставлення або автозаповнення, а також можливу зміну пристрою чи грубу дискретизацію часових даних; після цього формується сире значення показника надійності, яке за потреби калібрується на валідаційній вибірці; на завершальному етапі відкаліброване значення порівнюється з двома порогами τ_low і τ_high, після чого зразок відноситься до одного з трьох рівнів — низького, середнього або високого.

На рисунку 1 наведено алгоритм обчислення показника надійності та віднесення зразка до одного з трьох рівнів надійності.

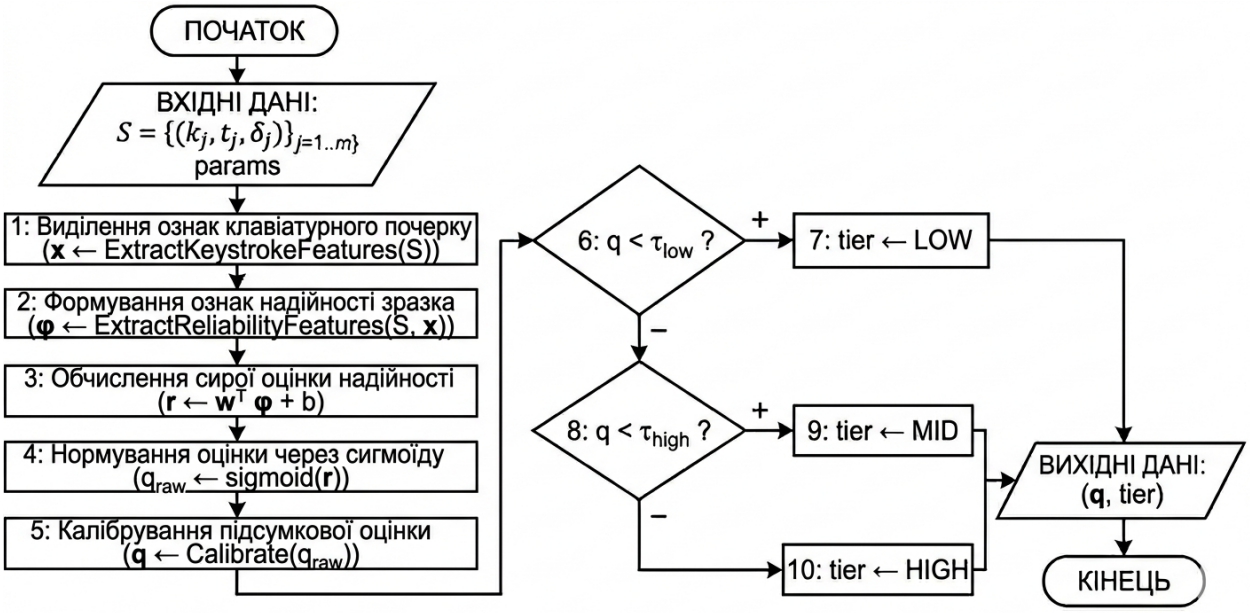

Рисунок 1 – Алгоритм обчислення показника надійності та віднесення зразка до одного з трьох рівнів надійності

У логіці практичного використання це означає, що за низького рівня надійності поведінковий канал не повинен використовуватися й система має перейти до альтернативних способів перевірки, за середнього рівня дозволяється лише обмежене використання з консервативною політикою прийняття рішень, а за високого рівня можна повноцінно враховувати поведінковий канал у схемі автентифікації.

У життєвому циклі такого модуля спочатку визначаються набір ознак φ(S) та порогові значення τ_low і τ_high, далі збираються валідаційні та калібрувальні зразки разом із позначеннями типових деградацій, таких як автозаповнення, втрата частини подій або зміна пристрою. На етапі навчання оцінюються параметри w і b, а також перевіряється якість калібрування за відповідними показниками. Під час експлуатації для кожної нової спроби входу в режимі онлайн обчислюється значення q, а агреговані характеристики роботи системи зберігаються для подальшого аналізу. Надалі модуль має періодично проходити переоцінювання, перекалібрування та перевірку на дрейф даних, оскільки надійність поведінкового зразка залежить не лише від самого користувача, а й від умов введення, змін інтерфейсу, типу пристрою та якості зчитування подій.

ОБГОВОРЕННЯ РЕЗУЛЬТАТІВ

Таблиця 1 подає придатні для дослідження «надійності зразка» публічні набори даних (для статичної «login»-верифікації найрелевантніші фіксовані паролі/фрази; для факторів зміни середовища — багатоклавіатурні та неконтрольовані набори).

Таблиця 1 – Порівняння публічних наборів даних для дослідження «надійності зразка».

Датасет	Тип даних	К-сть користувачів	Обсяг	Особливості/шум	Доступність	Першоджерело
CMU DSL-StrongPassword	Фіксований пароль; таймінги hold/digraph	51	400/користувач	8 сесій; точні часові мітки; контроль помилок	Відкрито	Carnegie Mellon University
GREYC Keystroke	Фіксована фраза; multi-keyboard	133 (100 ≥5 сесій)	7555	2 клавіатури; протокол чергування	Відкрито (архів)	GREYC-Keystroke benchmark
Clarkson II	Вільний ввід; подієві логи	101 файлів	подієво	неконтрольоване середовище; timestamp 100 нс; очищено PII	CITeR	Clarkson University
Buffalo (WIFS)	Фіксований + вільний текст	157	2.14M натискань	multi-session; subset multi-keyboard	На запит/обмежено	University at Buffalo
Aalto 136M	Тест друку речень; великі логи	≈168K	136M натискань	неконтрольовано; desktop/laptop	Відкрито	Aalto University

Емпіричну частину виконано на CMU DSL-StrongPassword (51×400 введень).

Для експериментальної перевірки для кожного користувача перші 200 введень використовувалися для побудови еталонного шаблону у вигляді середніх значень і стандартних відхилень за ознаками клавіатурного почерку. Наступні 200 введень того самого користувача розглядалися як легітимні приклади під час тестування. Приклади зловмисника формувалися як чужі введення, подані від імені вибраного користувача; для кожного заявленого облікового запису використовувалося по 200 таких прикладів. Як базовий верифікатор застосовувалася стандартизована Manhattan-відстань між поточним вектором ознак та еталонним шаблоном, а підсумковий скор визначався як від’ємне значення цієї відстані.

$$d(x,\mu) = \frac{1}{d}\sum_{i=1}^{d}\left|\frac{x_i - \mu_i}{\sigma_i}\right|, \quad score = -d.$$

Для дослідження стійкості методу додатково моделювалися синтетичні сценарії деградації з імовірністю 0,3 при фіксованому значенні генератора випадкових чисел seed = 7. До таких сценаріїв належали втрата частини ознак, укорочення послідовності, додавання часових збурень, імітація автозаповнення, квантування часових міток та зсув характеристик, пов’язаний зі зміною пристрою. Модуль оцінювання надійності реалізовувався у вигляді логістичної моделі, побудованої на ознаках якості зразка, серед яких враховувалися повнота, довжина, варіативність, величина кроку грубого квантування та індикатор автозаповнення. Цю модель навчали на окремій вибірці користувачів набору CMU із застосуванням такого самого генератора деградацій при seed = 111 та seed = 222. Для порівняння ROC-кривих, значень AUC і перевірки статистичної значущості відмінностей у загальному випадку доцільно застосовувати тест DeLong для корельованих ROC-кривих, а також непараметричні критерії для зіставлення методів на множині задач або наборів даних. [12]

На чистому наборі даних, тобто без штучно внесених деградацій, базовий верифікатор показує значення AUC = 0,911 та EER ≈ 0,166. Після переходу до режиму з деградаціями при p = 0,3 якість

розпізнавання помітно погіршується: значення AUC зменшується до 0,856, а EER зростає приблизно до 0,215. Таке погіршення є очікуваним, оскільки деградації порушують повноту подій і спотворюють часові характеристики введення, на яких ґрунтується клавіатурний почерк. Наведені значення отримано в межах відтворюваного експерименту за описаним протоколом, а метрики AUC та EER використано як стандартні показники якості для задач оцінювання клавіатурного почерку. Після цього доцільно навести ROC-криві для базового верифікатора на наборі CMU, а також окремо для підмножини зразків із високою надійністю, щоб наочно показати вплив якості даних на результат верифікації. [13]

Окремо було проведено експеримент із поділом зразків на низько-, середньо- та високонадійні. Для цього всі зразки впорядковувалися за значенням показника q, після чого розбивалися на три рівні частини. Такий поділ дав змогу простежити, як змінюються метрики верифікації залежно від рівня надійності зразка, а також як відрізняється сам профіль даних у кожній із груп. Результати цього порівняння доцільно подати у таблиці. Поріг t* при цьому фіксувався за значенням EER на чистому наборі даних, що забезпечувало коректне зіставлення профілів помилок між різними групами зразків. На таблиці 2 наведено результати експерименту.

Таблиця 2 – Результати верифікації та вплив надійності

Рівень надійності (тертилі q)	Покриття	AUC (всередині рівня)	EER (всередині рівня)	FPR@t*	FNR@t*	Сер. повнота	Сер. к-сть клавіш
низька	0.333	0.802	0.261	0.104	0.419	0.880	9.73
середня	0.333	0.900	0.178	0.131	0.220	0.9999	11.00
висока	0.333	0.882	0.191	0.209	0.179	1.0000	11.00

Головний висновок із таблиці 2 полягає в тому, що низька надійність зразка насамперед призводить до істотного зростання FNR, тобто до збільшення кількості відмов для легітимних користувачів. Це безпосередньо підтверджує основну тезу роботи: поведінковий канал недоцільно використовувати без попередньої перевірки якості поточного зразка. Водночас показник FPR не обов'язково має змінюватися монотонно зі зростанням q, оскільки сам показник надійності відображає не легітимність спроби входу, а якість виміряного поведінкового сигналу. Іншими словами, якісний зразок може належати як справжньому користувачеві, так і сторонній особі. Для кращої ілюстрації цього ефекту нижче доцільно навести графік, який показує залежність FNR і FPR від бінів надійності q за фіксованого порогу t*. [14]

ВИСНОВКИ

У роботі запропоновано модель ознак, яка дає змогу оцінювати надійність зразка клавіатурного почерку за сукупністю вимірюваних характеристик, зокрема повнотою подій, ефективною довжиною послідовності, варіативністю часових параметрів, ознаками автозаповнення або вставлення, зміни пристрою, квантування та часових збурень. На цій основі розроблено метод інтегрального оцінювання у вигляді ймовірнісного показника $q \in [0,1]$, який може бути використаний як у дослідницьких, так і в прикладних сценаріях. Важливо, що такий показник придатний до калібрування стандартними підходами й може оцінюватися за загальноприйнятими мірами, що робить його не лише формально визначеним, а й придатним для практичного використання.

Експериментально встановлено, що рівень надійності зразка безпосередньо пов'язаний з імовірністю помилки поведінкового верифікатора. Показано, що низьконадійні зразки суттєво частіше спричиняють відмови для легітимних користувачів, тобто мають вищий FNR за фіксованого порога. Це підтверджує, що не всі зібрані поведінкові дані можна безпосередньо використовувати в автентифікації, а модуль оцінювання надійності слід розглядати як обов'язковий попередній етап перед застосуванням поведінкового каналу. Для практичної інтерпретації показника q запропоновано двопорогову політику з рівнями low і high, яка дозволяє відносити зразки до трьох категорій: непридатні до використання, обмежено придатні та придатні до повноцінного використання.

З практичного погляду доцільно застосовувати таку політику: якщо $q < low$, клавіатурний почерк не слід використовувати як джерело доказу й систему варто переключати на альтернативний механізм перевірки або повторний збір подій; якщо $low \le q < high$, поведінковий канал доцільно використовувати лише як допоміжний чинник; якщо $q \ge high$, його можна використовувати повноцінно, але без ототожнення доброї якості зразка з легітимністю спроби. Верхній і нижній пороги мають підбиратися з

урахуванням придатності зразків, покриття та виграшу за метриками AUC, EER і FNR. У виробничому сценарії такий підхід дозволяє уникати ситуацій, коли низькоякісний зразок спричиняє несправедливу відмову або некоректну оцінку, і за потреби переводити перевірку на інший фактор автентифікації.

Разом із тим дослідження має певні обмеження: частина деградацій моделювалася синтетично, а набір CMU DSL-StrongPassword не повністю відображає реальні помилки введення, специфіку автозаповнення та збої збору подій. Тому подальша перевірка методу потребує використання інших наборів даних і умов, ближчих до реальної експлуатації. Окремо слід враховувати питання приватності й правового регулювання, оскільки клавіатурний почерк належить до поведінкової біометрії. Це означає необхідність мінімізації збережених даних, обмеження зберігання сирих подій, використання прозорих механізмів згоди та надання користувачеві альтернативного, не біометричного способу проходження автентифікації. Регулярний моніторинг калібрування та перекалібрування модуля також є обов'язковими, оскільки навіть незначні зміни в середовищі збору можуть впливати на таймінгові ознаки й на оцінку надійності.

СПИСОК ЛІТЕРАТУРИ / REFERENCES

1. Killourhy K. S., Maxion R. A. Keystroke Dynamics – Benchmark Data Set [Електронний ресурс]. Carnegie Mellon University, 2009. URL: https://www.cs.cmu.edu/~keystroke/
2. Giot R., El-Abed M., Rosenberger C. GREYC Keystroke: A Benchmark for Keystroke Dynamics Biometric Systems. In: 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems. 2009. P. 1–6. DOI: 10.1109/BTAS.2009.5339051.
3. ISO/IEC 29794-1:2024. Information technology – Biometric sample quality – Part 1: Framework. Geneva : International Organization for Standardization, 2024.
4. ISO/IEC 19795-1:2021. Information technology – Biometric performance testing and reporting – Part 1: Principles and framework. Geneva : International Organization for Standardization, 2021.
5. Kuzminykh I., Ghita B., Silonosov A. On Keystroke Pattern Variability in Virtual Desktop Infrastructure. Computer Modeling and Intelligent Systems. 2021. Vol. 2864. P. 238–248. DOI: 10.32782/cmis/2864-21.
6. Wahab A. A., Hou D., Banavar M., Schuckers S., Eaton K., Baldwin J., Wright R. Shared Multi-Keyboard and Bilingual Datasets to Support Keystroke Dynamics Research. In: CODASPY '22: Proceedings of the Twelfth ACM Conference on Data and Application Security and Privacy. 2022. P. 236–241. DOI: 10.1145/3508398.3511516.
7. Dhakal V., Feit A. M., Kristensson P. O., Oulasvirta A. Observations on Typing from 136 Million Keystrokes. In: CHI '18 Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018. DOI: 10.1145/3173574.3174220.
8. Guo C., Pleiss G., Sun Y., Weinberger K. Q. On Calibration of Modern Neural Networks. In: Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research. 2017. Vol. 70. P. 1321–1330.
9. Platt J. C. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In: Advances in Large Margin Classifiers. Cambridge, MA : MIT Press, 1999. P. 61–74.
10. Zadrozny B., Elkan C. Transforming Classifier Scores into Accurate Multiclass Probability Estimates. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2002. P. 694–699. DOI: 10.1145/775047.775151.
11. Naeini M. P., Cooper G., Hauskrecht M. Obtaining Well Calibrated Probabilities Using Bayesian Binning. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2015. Vol. 29, No. 1. P. 2901–2907. DOI: 10.1609/aaai.v29i1.9602.
12. Brier G. W. Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review. 1950. Vol. 78. P. 1–3. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
13. DeLong E. R., DeLong D. M., Clarke-Pearson D. L. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988. Vol. 44, No. 3. P. 837–845. DOI: 10.2307/2531595.
14. Demšar J. Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research. 2006. Vol. 7. P. 1–30.
15. Bisikalo, O.; Kharchenko, V.; Kovtun, V.; Krak, I.; Pavlov, S. Parameterization of the Stochastic Model for Evaluating Variable Small Data in the Shannon Entropy Basis. *Entropy* 2023, *25*, 184.
16. Intellectual technologies in medical diagnostics, treatment and rehabilitation: monograph / [S.V. Pavlov,

O.G. Avrunin, S.M. Zlepko, E.V. Bodianskyi and others]; edited by S. Pavlov, O. Avrunin. – Vinnytsia: PP “TD “Edelweiss and K”, 2019. – 260 p.

Дата надходження: 15.02.2026

Дата прийняття до друку після рецензування: 25.03.2026

Дата публікації: 18.06.2026

КУРНІЦЬКИЙ ДМИТРО ПЕТРОВИЧ – аспірант групи 126-23а, факультету інтелектуальних інформаційні технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, ***e-mail: dmytro.kurnitskiy@gmail.com***, **https://orcid.org/0009-0000-3190-9514**

Dmytro KURNITSKIY

MODEL AND METHOD FOR ASSESSING THE RELIABILITY OF A KEYBOARD HANDWRITING SAMPLE FOR BEHAVIORAL USER AUTHENTICATION

Vinnytsia National Technical University