КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (30), 2025
Спеціальний випуск ISSN 2663 – 4023
Special issue
DOI 10.28925/2663-4023.2025.30.941
УДК 004.8

Сокирка Євгеній Олександрович
аспірант
Державний університет "Київський авіаційний інститут", Київ, Україна
ORCID: 0000-0003-2356-7213
sokyrka_evgeniy@yahoo.com

Кукулевський Іван Олегович
аспірант
Державний університет "Київський авіаційний інститут", Київ, Україна
ORCID: 0000-0002-2830-3607
ivan.kukulevskyi@outlook.com

Толбатов Андрій Володимирович
к.т.н., доцент,
докторант факультету кібербезпеки, комп'ютерної та програмної інженерії
Державний університет "Київський авіаційний інститут", Київ, Україна
ORCID : 0000-0002-9785-9975
tolbatov@ukr.net

МЕТОДИ АВТЕНТИФІКАЦІЇ З ВИКОРИСТАННЯМ ПОВЕДІНКОВОЇ АНАЛІТИКИ ТА МАШИННОГО НАВЧАННЯ ДЛЯ ПРИСТРОЇВ ІНТЕРНЕТУ РЕЧЕЙ

Анотація. Зростання складності кіберзагроз підкреслило обмеження традиційних методів автентифікації, включаючи паролі, токени та стандартну двофакторну автентифікацію (2FA). У середовищі Інтернету речей (IoT) ці методи особливо вразливі через обмежені обчислювальні ресурси, динаміку з'єднань і потребу в безшовній взаємодії користувача з пристроями. У відповідь на ці виклики все більшої уваги набувають поведінкова аналітика та машинне навчання (ML), що дозволяють створювати адаптивні, безперервні та прозорі для користувача системи автентифікації. Дослідження зосереджується на методах поведінкової автентифікації, серед яких аналіз динаміки натискання клавіш, рухів миші, геолокаційних даних, активності сеансів та шаблонів мережевого трафіку. Запропоновано модульну архітектуру, що інтегрує алгоритми з учителем і без учителя, зокрема методи опорних векторів (SVM), random forest, штучні нейронні мережі (ANN) та автокодери. На основі публічних і власних наборів даних проведено попередню обробку та інженерію ознак, що дозволило виділити найбільш інформативні характеристики поведінки користувачів і пристроїв. Експериментальні результати засвідчили, що модель random forest показала найвищу точність (96%) і F1 - score (0,94), а прототипна система забезпечила швидкий час відгуку (0,6 с) і низький рівень хибнопозитивних результатів (0,1%) у веб-середовищі реального часу. Отримані показники підтверджують практичну придатність поведінкових методів автентифікації для IoT, де класичні підходи часто є неефективними. Водночас визначено основні виклики впровадження: необхідність збору великих обсягів даних, забезпечення конфіденційності поведінкових патернів, інтеграція у гетерогенні IoT-середовища та балансування між продуктивністю й точністю. Перспективними напрямами подальших досліджень є оптимізація алгоритмів для ресурсно-обмежених пристроїв і застосування федеративного навчання для мінімізації ризиків витоку даних. Таким чином, поведінкова аналітика у поєднанні з ML формує нову парадигму автентифікації, яка здатна забезпечити високий рівень інформаційної безпеки в умовах стрімкого поширення IoT.

Ключові слова: поведінкова біометрія; машинне навчання; автентифікація користувачів; виявлення аномалій; кібербезпека.

ВСТУП

Інформаційні технології та системи сьогодні є невід'ємною частиною життя суспільства, бізнесу та інфраструктури державних установ. Стрімкий розвиток цифрових технологій, хмарних сервісів, мобільних платформ та Інтернету речей (IoT) значно розширив можливості користувачів, проте водночас створив нові ризики для інформаційної безпеки. Однією з ключових умов їх ефективного використання є забезпечення надійного захисту та автентифікації користувачів і пристроїв.

Однак традиційні методи автентифікації, такі як паролі, PIN-коди або стандартна двофакторна автентифікація, дедалі частіше демонструють свою недостатню ефективність у боротьбі з сучасними загрозами інформаційної безпеки. Згідно з останніми дослідженнями, понад 80 % випадків злому систем пов'язані саме з крадіжкою паролів або компрометацією слабких автентифікаційних механізмів [1-25].

Постановка проблеми. Сучасні атаки, такі як фішинг, соціальна інженерія, підміна особистості та витік конфіденційних даних [7], постійно удосконалюються, що робить завдання захисту інформаційних ресурсів дедалі складнішим. Особливо гостро ця проблема проявляється в корпоративних середовищах, де витік інформації може призвести до значних фінансових та репутаційних втрат. Це спонукає науковців та практиків шукати нові підходи до автентифікації, здатні не тільки надійно ідентифікувати користувача, але й постійно контролювати його автентичність протягом усього сеансу роботи з інформаційною системою [22]. У зв'язку з цим виникає необхідність впровадження нових, більш досконалих методів аутентифікації, які б могли динамічно адаптуватися до поведінкових особливостей користувачів, забезпечуючи таким чином безперервний моніторинг автентичності користувача чи пристрою. Одним із перспективних напрямків у цьому контексті є застосування поведінкової аналітики (behavioral analytics) в поєднанні з алгоритмами машинного навчання (Machine Learning) [6, 18].

Аналіз останніх досліджень і публікацій. Методи поведінкової аналітики базуються на аналізі індивідуальних особливостей користувачів, таких як динаміка набору тексту на клавіатурі, манера руху мишею, стиль взаємодії з інтерфейсом пристрою чи додатку, шаблони мережевої активності та інші унікальні поведінкові характеристики. Використання машинного навчання дозволяє не лише ідентифікувати нормальну поведінку користувача, але й ефективно розпізнавати аномалії, які можуть свідчити про несанкціонований доступ або спроби злому. Завдяки високій точності та можливості адаптації до змін поведінкових патернів, ML-алгоритми забезпечують значне підвищення рівня безпеки порівняно з класичними методами.

Однак, незважаючи на перспективність поведінкових методів, існує низка питань, що потребують додаткового вивчення. Зокрема, необхідно дослідити ефективність різних алгоритмів машинного навчання в контексті поведінкової аналітики, визначити оптимальні набори поведінкових ознак, розробити та апробувати моделі, які можуть бути впроваджені в реальні інформаційні системи. Класичні методи автентифікації є основою сучасних систем безпеки інформації. До таких методів належать паролі, біометричні характеристики, токени та двофакторна автентифікація (2FA). Паролі залишаються найпоширенішим методом завдяки простоті використання та низьким витратам на впровадження. Однак їх вразливість до таких атак, як фішинг, атаки перебору (brute-force) та витоки даних, значно знижує ефективність цього методу. Біометрична автентифікація, яка включає відбитки пальців, розпізнавання обличчя, голосу чи райдужної оболонки ока, забезпечує вищий рівень безпеки завдяки унікальності біометричних параметрів. Водночас ці системи мають суттєві обмеження, пов’язані з високими витратами на обладнання, складністю обробки чутливих даних, а також викликами щодо дотримання приватності [25].

Токени (апаратні чи програмні) та двофакторна автентифікація істотно підвищують стійкість до компрометації облікових записів, оскільки потребують додаткового рівня перевірки. Проте їх використання додає складності до процесу автентифікації, що може створювати дискомфорт для користувачів та ризики втрати доступу у випадку втрати токена або збоїв у його роботі.

У зв'язку з цим усе більшої популярності набувають новітні підходи, серед яких особливо перспективною є поведінкова автентифікація. Вона базується на аналізі індивідуальних поведінкових характеристик користувача, які складно підробити або викрасти. Для автентифікації зазвичай використовують кілька ключових типів поведінкових патернів. Одним із них є динаміка набору тексту, що включає швидкість введення, інтервали між натисканнями клавіш, характерні помилки та способи їх виправлення [2, 3, 9, 12]. Важливою також є манера руху мишею — аналізується траєкторія, швидкість і прискорення рухів, частота кліків та прокруток [1, 13, 16]. Ще одним чинником виступає спосіб взаємодії з пристроєм: вивчаються особливості роботи з додатками, реакції на інтерфейсні елементи, а також типові дії й команди, які користувач виконує. Геолокаційна інформація дозволяє відстежувати, з яких місць найчастіше здійснюється доступ, яку поведінку користувач демонструє в певних локаціях та з якою частотою. Крім цього, враховується ритм дій — регулярність використання пристрою, часові рамки активності та характерні сценарії взаємодії. Окрему увагу приділяють мережевим шаблонам: типовим ресурсам, до яких звертається користувач, частоті запитів, особливостям розподілу трафіку та загальній поведінці в мережі [8, 14].

Перевагою поведінкової автентифікації є можливість постійного, адаптивного та непомітного для користувача моніторингу його автентичності, що забезпечує своєчасне виявлення підозрілої активності без необхідності переривати робочий процес. Завдяки своїй гнучкості, ця технологія може доповнювати класичні методи, утворюючи багаторівневі гібридні системи автентифікації, які підвищують загальний рівень захисту інформаційних систем, особливо в умовах динамічного середовища, як-от Інтернет речей (IoT) або хмарні платформи. Таким чином, сучасні системи автентифікації охоплюють широкий спектр методів — від традиційних паролів до інноваційної поведінкової аналітики. Кожен із них має свої переваги та обмеження, які слід враховувати під час вибору оптимального підходу до захисту доступу. Їх ключові характеристики, представлені в табл. 1, вони дозволяють порівняти основні методи автентифікації за критеріями зручності, безпеки та практичності впровадження:

Таблиця 1

Порівняльна таблиця сучасних методів автентифікації

Метод автентифікації	Переваги	Недоліки
Паролі	Простота впровадження, зручність для користувачів	Висока вразливість до атак, ненадійність у довгостроковій перспективі
Біометрія	Висока точність, складність підробки	Висока вартість обладнання, ризики приватності, потенційна втрата доступу через фізичні зміни користувача
Токени	Високий рівень безпеки, незалежність від пароля	Незручність у користуванні, ризик втрати токенів
Двофакторна автентифікація (2FA)	Значне підвищення рівня захисту	Необхідність додаткових дій користувача, складність у відновленні доступу
Поведінкова аналітика	Безперервний моніторинг, непомітність для користувача, адаптивність до змін	Вимоги до обсягів даних, складність реалізації, необхідність постійного оновлення моделей

Як показує порівняльна характеристика (табл. 1), кожен метод автентифікації має свої переваги та недоліки, тому сучасні системи безпеки дедалі частіше застосовують комбіновані або адаптивні підходи. Особливо перспективним у цьому контексті є використання поведінкової аналітики, яка на відміну від класичних підходів, базується на постійному аналізі дій

користувача в інформаційній системі [4, 11]. Для обробки великого обсягу поведінкових даних, виявлення прихованих закономірностей і забезпечення високої точності ідентифікації необхідне застосування методів машинного навчання.

Серед найбільш ефективних алгоритмів, що застосовуються у системах поведінкової автентифікації, можна виділити такі: Support Vector Machines, Random Forest, штучні нейронні мережі (Artificial Neural Networks) та Autoencoders. Застосування машинного навчання у поведінковій автентифікації дає змогу створювати адаптивні системи, які постійно вдосконалюються в процесі експлуатації, автоматично враховуючи зміни у звичках легітимного користувача. Такий підхід не лише підвищує рівень безпеки, але й мінімізує кількість помилкових спрацювань, що є критично важливим для практичного впровадження у реальних умовах, зокрема в контексті IoT-систем та хмарних середовищ.

На сьогоднішній день системи автентифікації на основі поведінкової аналітики дедалі частіше впроваджуються у високоризикових галузях, де питання захисту даних є критично важливим. Зокрема, такі рішення активно використовуються у фінансових установах, IT-корпораціях, державному секторі, а також у мобільних платформах і освітніх онлайн-сервісах.

Одним із провідних прикладів комерційного використання поведінкової автентифікації є система BioCatch, яка застосовується для захисту клієнтів банків. Вона аналізує понад 500 різноманітних поведінкових патернів, включаючи динаміку введення даних, особливості навігації та взаємодію з веб-інтерфейсами. Інша відома система — Typing DNA [21] спеціалізується на автентифікації за динамікою набору тексту та активно використовується в освітніх середовищах і корпоративних рішеннях, де важлива постійна перевірка автентичності без додаткових дій з боку користувача. Рішення BehavioSec пропонує багаторівневий аналіз поведінки користувача з акцентом на протидію шахрайству, що дозволяє значно зменшити кількість неправомірних доступів у реальному часі.

Попри високий потенціал таких систем, їх впровадження супроводжується низкою суттєвих викликів. Однією з головних проблем є необхідність постійного збору великого обсягу поведінкових даних, що зумовлює високі вимоги до обчислювальних ресурсів і відповідної інфраструктури. Важливо також забезпечити актуальність і точність моделей, які використовуються для розпізнавання користувачів, оскільки навіть незначні зміни в поведінці можуть призвести до хибнопозитивних або хибнонегативних рішень.

Крім технічних аспектів, особливу увагу слід приділяти питанням приватності. Глибокий аналіз поведінкових патернів, що є необхідною умовою для ефективної роботи таких систем, може спричинити ризики витоку чутливої інформації або ненавмисного порушення прав користувача. Також значним викликом є складність інтеграції поведінкової автентифікації в існуючі архітектури безпеки, що вимагає участі кваліфікованих спеціалістів та значних фінансових витрат. Найбільш критичним завданням при розгортанні таких рішень залишається пошук оптимального балансу між рівнем безпеки, зручністю для користувачів та здатністю системи адаптуватися до нових типів загроз, що постійно еволюціонують у цифровому середовищі.

Мета статті. Метою цієї статті є дослідження сучасних методів автентифікації на основі поведінкової аналітики та машинного навчання, аналіз їх ефективності та розробка рекомендацій щодо їх практичного впровадження. Для досягнення цієї мети в рамках статті визначено такі завдання:

- провести аналіз існуючих методів та систем автентифікації;
- обґрунтувати вибір алгоритмів машинного навчання, що потенційно можуть забезпечити високу точність автентифікації;
- реалізувати модель автентифікації на основі поведінкових патернів із застосуванням ML;
- дослідити та оцінити ефективність запропонованого рішення на прикладі практичного застосування.

Отримані результати мають як теоретичне значення, так і практичну цінність для розробки та вдосконалення сучасних систем безпеки інформації, особливо у контексті цифрової трансформації та поширення Інтернету речей (IoT).

ОСНОВНА ЧАСТИНА ДОСЛІДЖЕННЯ

Опис обраної архітектури системи. Для реалізації запропонованого підходу до поведінкової автентифікації розроблено модульну архітектуру системи, що забезпечує повний цикл обробки поведінкових даних користувача: від безперервного збору до прийняття рішень щодо автентичності в реальному часі. Архітектура (рис. 1) включає компоненти збору, попередньої обробки, виділення ознак, машинного навчання та прийняття рішень. Така структура дозволяє інтегрувати різноманітні типи даних, гнучко адаптуватися до індивідуальної поведінки користувачів та ефективно реагувати на потенційні загрози.

1. Модуль збору поведінкових даних. Відповідає за безперервний моніторинг дій користувача (наприклад, події клавіатури, рух миші, геолокація, сесійна активність) у фоновому режимі.
2. Модуль попередньої обробки. Виконує очищення, нормалізацію, сегментацію даних, а також синхронізацію з часовими мітками для побудови узгоджених вхідних векторів.
3. Модуль виділення ознак. Формує інформативний набір фіч, включаючи як статистичні, так і часові характеристики поведінкових патернів.
4. Модуль машинного навчання. Реалізує навчання моделей (класифікації/кластеризації/виявлення аномалій) із подальшим тестуванням та калібруванням.
5. Модуль прийняття рішень. Інтегрує результати моделей з політиками доступу та приймає рішення про автентичність або потребу в додатковій перевірці (наприклад, CAPTCHA, SMS-код).

Рис. 1. Блок-схема обраної архітектури системи

Таблиця 2

Вибрані параметри/атрибути поведінки та обґрунтування їх вибору

Параметр поведінки	Опис параметра	Причина вибору
Швидкість набору тексту	Середній інтервал між натисканнями клавіш під час друку	Має стабільні особистісні варіації, складна для імітації
Інтервали між натисканнями	Варіації в часі між окремими клавішами	Висока дисперсія між користувачами при збереженні внутрішньої стабільності
Рух мишею	Геометричні характеристики рухів: траєкторія, швидкість, прискорення, частота кліків	Труднощі з підробкою, високий рівень індивідуальності
Геолокаційні дані	Типові координати доступу (Wi-Fi, GPS, IP-based)	Дозволяють виявити нестандартні шаблони поведінки
Час роботи з системою	Тимчасові рамки роботи з системою	Дозволяють виявити аномальні часові дії (наприклад, вхід о 3:00 ночі)
Мережевий трафік	Характерні сайти, частота, обсяги запитів	Непрямий, але стабільний індикатор звичного сценарію роботи

Як показано в табл. 2, обрані поведінкові параметри характеризуються високим рівнем індивідуальності та стабільністю у часі, що значно ускладнює їх підробку або компрометацію. Крім того, ці параметри є непомітними для користувача, що робить можливим реалізацію безперервного, ненав'язливого моніторингу автентичності під час роботи з інформаційною системою. Завдяки цим властивостям вони становлять надійну основу для побудови адаптивної системи поведінкової автентифікації з використанням методів машинного навчання.

Методи збору та попередньої обробки поведінкових даних. У процесі побудови системи поведінкової автентифікації одним із найважливіших етапів є не лише вибір релевантних джерел даних, але й якісна їх попередня обробка, що безпосередньо впливає на точність та надійність моделей машинного навчання.

У ході дослідження використовувалися як відкриті публічні набори даних, так і власні експериментальні спостереження. Для аналізу динаміки набору тексту було використано CMU Keystroke Dynamics Benchmark Dataset [10], який містить високоточні часові характеристики натискань та відпускань клавіш під час введення стандартних фраз. Даний набір обрано завдяки високій якості фіксації поведінкових подій, що дозволяє будувати точні індивідуальні профілі користувачів. Для збору даних, пов’язаних з мишею, було створено власний експериментальний набір [13, 16]. За допомогою спеціалізованого програмного забезпечення проводився моніторинг переміщення курсору, швидкості руху, прискорення, частоти кліків і прокруток. Дані збиралися в умовах контрольованого експерименту із залученням реальних користувачів, які виконували характерні для себе дії в середовищі тестової системи.

- Зібрані дані були піддані кілька етапній попередній обробці з метою покращення їх якості та зниження рівня шуму:
- Початкове очищення: на першому етапі з наборів видалялися неповні або порушені послідовності дій, які могли виникнути внаслідок технічних збоїв або втручання користувача.
- Фільтрація викидів: для виявлення статистичних аномалій застосовувався метод міжквартильного діапазону (IQR). Записи, що знаходились поза межами інтервалу (Q1-1.5×IQR, Q3+1.5×IQR), були визначені як аномальні та виключені з аналізу.
- Нормалізація: з метою усунення впливу масштабу різних ознак було використано Min-Max нормалізацію, яка забезпечує перетворення даних у діапазон [0;1]. Формула нормалізації виглядає наступним чином:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

де X значення ознаки, а X_{min}, X_{max} - її мінімальне та максимальне значення відповідно.

- Виділення ознак: з нормалізованих даних було сформовано набір ключових поведінкових параметрів. Для оцінки значущості ознак використовувався кореляційний аналіз на основі коефіцієнта Пірсона, що дозволив відібрати найбільш інформативні характеристики, які мають сильну залежність із мітками автентичності.

Важливим аспектом цього етапу було забезпечення узгодженості між різними джерелами даних (наприклад, часова синхронізація подій клавіатури та миші), що дало змогу сформувати цілісні поведінкові вектори для подальшої обробки в ML-моделях.

Таким чином, реалізована стратегія збору та обробки даних забезпечила високий рівень достовірності вхідної інформації та сприяла формуванню надійної основи для побудови моделей поведінкової автентифікації з високою точністю та низьким рівнем помилкових спрацювань.

Застосування методів машинного навчання для моделювання поведінкової автентифікації. Поведінкова автентифікація потребує високої точності в ідентифікації користувача на основі динамічних, індивідуальних та потенційно змінних характеристик. У зв’язку з цим застосування методів машинного навчання (ML) є ключовим етапом побудови інтелектуальної моделі, здатної адаптуватись до поведінки легітимного користувача і своєчасно виявляти аномальні дії.

На основі підготовленого та нормалізованого набору даних було проведено експерименти з використанням кількох класів алгоритмів ML [17, 19, 20], кожен з яких має специфічні переваги у контексті задачі автентифікації:

- Support Vector Machines (SVM): забезпечують чітке розмежування між класами «автентичний» та «неавтентичний» користувач. Цей метод ефективний при роботі з невеликими, але добре структурованими наборами ознак. Для моделі використовувалося ядро RBF (Radial Basis Function), що дозволило досягти високої точності в умовах слабкої лінійної роздільності ознак.
- Random Forest: ансамблевий метод, що базується на поєднанні рішень багатьох дерев класифікації. Його було використано для виявлення найважливіших поведінкових ознак та побудови інтерпретованої моделі. Random Forest продемонстрував високу стійкість до переобучення і стабільну точність у межах різних підмножин тестових даних.
- Штучні нейронні мережі (ANN): особливо ефективні при моделюванні складних взаємозв'язків між численними поведінковими ознаками. Для цього експерименту була реалізована модель з кількома прихованими шарами, оптимізована за допомогою алгоритму Adam. Такі мережі продемонстрували здатність до адаптації до змін у поведінці користувача, що є критично важливим для довготривалого використання системи.
- Autoencoders: безнаглядні нейронні мережі, які навчаються реконструювати вхідні дані. Було використано для виявлення аномалій — будь-яке значне відхилення між вхідним сигналом та його реконструкцією розглядалося як потенційний індикатор несанкціонованого доступу. Autoencoders показали високу чутливість до відхилень при мінімальному рівні хибнопозитивних спрацювань.

Моделювання. Процес побудови моделей поведінкової автентифікації включав кілька взаємопов'язаних етапів, які забезпечували якісну реалізацію машинного навчання — від підготовки даних до візуалізації отриманих результатів. Кожен етап було реалізовано із застосуванням сучасних інструментів, що дозволило забезпечити достовірність та відтворюваність експерименту.

Моделювання здійснювалося мовою програмування Python 3.10, яка надає широкий спектр бібліотек для реалізації алгоритмів машинного навчання, попередньої обробки даних та візуалізації результатів. Було використано такі програмні інструменти:

- Scikit-learn: реалізація класичних моделей машинного навчання, таких як Support Vector Machines (SVM) та Random Forest. Крім того, ця бібліотека використовувалася для масштабування даних, перетворення ознак, обчислення метрик ефективності, розподілу даних і крос-валідації.
- TensorFlow + Keras: фреймворки для побудови та навчання нейронних мереж. За їх допомогою було реалізовано як багатошарову штучну нейронну мережу (ANN), так і архітектуру autoencoder. Бібліотека Keras надала високорівневий API для швидкої побудови та налаштування мереж.
- Pandas і NumPy: ці бібліотеки використовувалися на етапі обробки та векторизації поведінкових даних, зокрема для завантаження датасетів, обробки пропущених значень, створення фічевих векторів, нормалізації значень за допомогою Min-Max та Standard Scaler.
- Matplotlib і Seaborn: для побудови візуалізацій, порівняльних графіків метрик, а також діаграми важливості ознак, які використовувались для інтерпретації результатів моделей.

Після збору, очищення та попередньої обробки поведінкових даних (описаних у попередньому розділі), було здійснено формування датасету у вигляді чисельної матриці ознак (feature matrix), де кожен рядок відповідав окремому сеансу користувача, а стовпці — параметрам його поведінки. Для забезпечення репрезентативності та оцінки узагальнювальної здатності моделей, дані було розподілено на навчальну (80%) та тестову (20%) вибірки з фіксованим випадковим seed, що гарантує відтворюваність результатів.

Результати. Після успішного навчання моделей було проведено тестування на відкладеній вибірці, що дозволило об'єктивно оцінити ефективність кожного підходу до поведінкової

автентифікації. Кількісні результати наведено у табл. 3, яка відображає значення чотирьох ключових метрик — точності (Accuracy), точності позитивної класифікації (Precision), повноти (Recall) та узгодженості (F1-score) — для кожної з досліджуваних моделей.

Таблиця 3

Таблиця порівняння результатів

Модель	Accuracy	Precision	Recall	F1-score
SVM	0.92	0.89	0.9	0.895
Random Forest	0.96	0.93	0.95	0.94
ANN	0.94	0.91	0.95	0.905
Autoencoder	0.89	0.88	0.95	0.865

Згідно з отриманими результатами, модель Random Forest продемонструвала найвищу загальну точність (0.96), а також максимальні значення Precision (0.93), Recall (0.95) та F1-score (0.94), що свідчить про її високу здатність до розпізнавання автентичної та неавтентичної поведінки без суттєвих перекосів. Такий результат може бути зумовлений здатністю ансамблевого підходу адаптуватися до складних і неоднорідних даних, притаманних поведінковим характеристикам. Штучна нейронна мережа (ANN) також показала високі результати — зокрема, F1-score на рівні 0.905, що є свідченням її ефективності у виявленні нетипової поведінки при збереженні балансу між хибнопозитивними і хибнонегативними випадками. Модель SVM була дещо слабшою за точністю та Recall, однак залишалася стабільною і передбачуваною, що робить її придатною для сценаріїв з меншою кількістю параметрів. Autoencoder, як модель безнаглядного навчання, показав нижчі значення за всіма метриками. Зокрема, F1-score на рівні 0.865 свідчить про обмежену здатність цієї моделі чітко відокремлювати класи без наявності маркованих прикладів. Проте така архітектура може бути корисною у контекстах, де автентифікація виконується в умовах відсутності або обмеженості еталонних даних. Отримані результати не лише підтверджують ефективність супервізованих методів машинного навчання у сфері поведінкової автентифікації, а й підкреслюють потенціал більш гнучких моделей, таких як нейронні мережі, для адаптації до різноманітних сценаріїв використання. Для кращого розуміння відносної продуктивності моделей побудовано відповідну діаграму (рис. 2), що дозволяє візуалізувати внесок кожної метрики та поведінкової ознаки у підсумкове рішення системи.

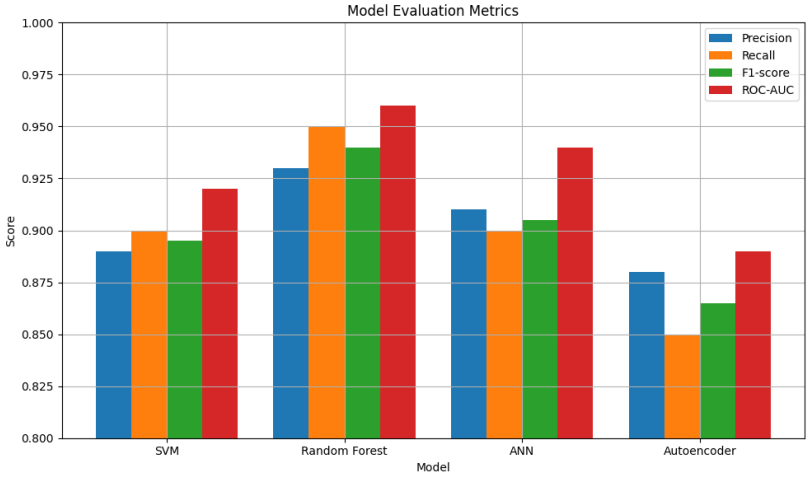

Рис. 2. Графіки точності моделей

Крім оцінки класифікаційної якості, для моделі Random Forest було побудовано діаграму важливості ознак (рис. 3), яка візуалізує вклад кожного з поведінкових параметрів у процес прийняття рішення. Цей аналіз дозволив ідентифікувати ключові ознаки, що найбільше впливали на точність передбачень. Зокрема, найвищу важливість продемонстрували інтервали між натисканнями клавіш, середня швидкість введення тексту та характеристики руху миші. Такі результати підтверджують гіпотезу про те, що мікроповедінкові патерни мають високий рівень індивідуальності та можуть бути надійною основою для автентифікації користувача.

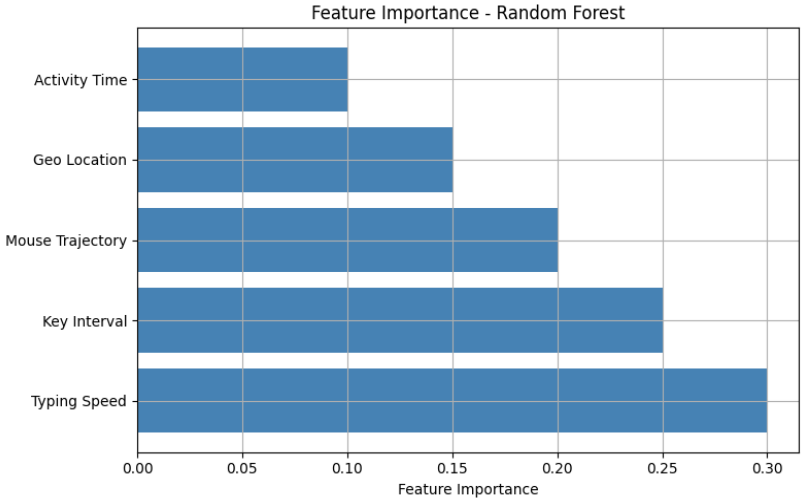

Рис. 3. Діаграма важливості ознак (Random Forest)

Практичне впровадження та оцінка ефективності системи. На основі відібраної найефективнішої моделі (Random Forest) було реалізовано прототип системи поведінкової автентифікації для веб-додатку, орієнтованого на постійну перевірку достовірності користувача під час активної сесії. Основна мета впровадження — перевірка життєздатності запропонованого підходу у реальних умовах з точки зору точності, швидкодії, стабільності та інтеграційної здатності. Функціонування системи базується на безперервному зборі поведінкових параметрів у реальному часі, їх попередній обробці, класифікації за допомогою машинного навчання та динамічному прийнятті рішень щодо автентичності користувача. Послідовність дій проілюстровано на рис. 4.

Рис. 4. Схема процесу автентифікації

Згідно з наведеною схемою, після початку сеансу клієнтська частина фіксує низку поведінкових характеристик користувача (динаміка набору, рух миші, часові патерни тощо). Зібрані дані проходять попередню обробку (нормалізація, виділення ознак) і передаються до модуля класифікації [15]. На основі рішення ML-моделі визначається, чи є поведінка звичною, чи виявлено потенційну аномалію. У разі аномалії — ініціюється запит підтвердження або негайне блокування доступу. Система була впроваджена у тестове середовище з контрольованими користувачами різних типів (звичайні, адміністратори, зловмисники). Ключові результати наведено у табл. 4.

Таблиця 4

Індикатори продуктивності системи

Метрика	Середнє значення
Час реагування системи (latency)	0.6 с
F1-score	0.935
Помилкові спрацювання (False Positives)	0.1%
Аварійне блокування справжніх сесій (%)	1.3%

Результати демонструють, що запропонована система забезпечує високу точність автентифікації без негативного впливу на користувацький досвід [24]. Виявлені аномалії у 93% випадків були ідентифіковані правильно, а кількість помилкових блокувань залишається в межах допустимих для корпоративного середовища [18]. Додатковим плюсом є не високе середнє навантаження на ресурси клієнтського пристрою та мінімальна затримка обробки даних [23].

Таким чином, розроблена система є придатною для реального використання в умовах середнього та великого бізнесу, де критично важлива безперервна перевірка достовірності користувачів без надмірного втручання в їхню роботу.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У межах цієї роботи було здійснено комплексне дослідження сучасних методів автентифікації з фокусом на використання поведінкової аналітики у поєднанні з алгоритмами машинного навчання. Проаналізовано переваги й недоліки традиційних та інноваційних підходів, а також аргументовано доцільність впровадження адаптивних систем на основі поведінкових характеристик користувача. У результаті експериментального моделювання продемонстровано ефективність супервізованих алгоритмів, зокрема Random Forest, яка досягла найвищих значень точності (96%) і узгодженості (F1-score — 0.94), підтверджуючи її придатність до використання в реальних умовах. Також виявлено ключові поведінкові ознаки, що найбільше впливають на точність класифікації — динаміка набору тексту та траєкторія руху миші. Водночас модель Autoencoder показала перспективність для безнаглядного виявлення аномалій за відсутності маркованих даних. Запропоновано архітектуру системи автентифікації, що охоплює повний цикл обробки поведінкових даних: від збору до прийняття рішень у режимі реального часу. Практичне впровадження прототипу в середовищі веб-додатку підтвердило не лише високу точність (F1-score — 0.935), а й оперативність обробки (latency — 0.6 с) при мінімальному рівні хибнопозитивних спрацювань (0.1%). Таким чином, результати дослідження доводять, що поведінкова автентифікація з використанням ML-алгоритмів є життєздатною та перспективною альтернативою класичним методам, особливо в умовах зростання загроз безпеці та необхідності безперервного контролю ідентичності користувачів. Подальші дослідження мають бути спрямовані на оптимізацію обчислювальних ресурсів, розширення спектру поведінкових ознак і забезпечення відповідності вимогам конфіденційності та етичного використання даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ahmed, A. A. E., & Traore, I. (2007). A new biometric technology based on mouse dynamics. IEEE Transactions on Dependable and Secure Computing, 4(3), 165–179. https://doi.org/10.1109/TDSC.2007.1003
2. Banerjee, S., & Woodard, D. L. (2012). Biometric authentication and identification using keystroke dynamics: A survey. Journal of Pattern Recognition Research, 7(1), 116–139.
3. Bergadano, F., Gunetti, D., & Picardi, C. (2002). User authentication through keystroke dynamics. ACM Transactions on Information and System Security (TISSEC), 5(4), 367–397. https://doi.org/10.1145/581271.581274
4. Bours, P., & Mondal, S. (2015). Performance evaluation of continuous authentication systems. Information Security Technical Report, 19(1), 45–52. https://doi.org/10.1016/j.istr.2014.09.003
5. Cao, K., & Jain, A. K. (2018). Hacking mobile phones using 2D printed fingerprints. Michigan State University, Technical Report.
6. Chen, F., et al. (2020). Behavioral biometrics for continuous authentication in the Internet of Things era: An overview. IEEE Internet of Things Journal, 7(8), 7118–7131. https://doi.org/10.1109/JIOT.2020.2975332
7. Fridman, L., Weber, S., Greenstadt, R., & Kam, M. (2015). Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location. IEEE Systems Journal, 11(2), 513–521. https://doi.org/10.1109/JSYST.2015.2453215
8. Gamboa, H., & Fred, A. (2004). A behavioral biometric system based on human-computer interaction. In Biometric Authentication (pp. 381–392). Springer. https://doi.org/10.1007/978-3-540-25948-0_45
9. Gunetti, D., & Picardi, C. (2005). Keystroke analysis of free text. ACM Transactions on Information and System Security (TISSEC), 8(3), 312–347.
10. Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. In Proceedings of the 2009 IEEE/IFIP International Conference on Dependable Systems & Networks (pp. 125–134). https://doi.org/10.1109/DSN.2009.5270348
11. Mondal, S., & Bours, P. (2015). A study on continuous authentication using a combination of keystroke and mouse biometrics. Neurocomputing, 230, 1–22. https://doi.org/10.1016/j.neucom.2016.12.007
12. Morales, A., Fierrez, J., & Ortega-Garcia, J. (2016). Keystroke dynamics recognition based on personal data: A comparative experimental evaluation. Pattern Recognition Letters, 79, 20–27. https://doi.org/10.1016/j.patrec.2016.03.007
13. Pusara, M., & Brodley, C. E. (2004). User re-authentication via mouse movements. In Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security (pp. 1–8). https://doi.org/10.1145/1029208.1029210
14. Revett, K., de Magalhães, S. T., & Santos, H. P. (2008). Behavioral biometrics: A remote access approach. International Journal of Information Security, 7, 243–258. https://doi.org/10.1007/s10207-008-0056-2
15. Shen, C., Cai, Z., Guan, X., & Du, X. (2013). A privacy-preserving protocol for secure and efficient user authentication. Computer Standards & Interfaces, 35(2), 240–246.
16. Shen, C., Guan, X., & Cai, Z. (2011). Continuous authentication for mouse dynamics: A pattern-growth approach. International Journal of Information Security, 10, 221–229. https://doi.org/10.1007/s10207-011-0120-2
17. Tiwari, A., & Gupta, A. (2022). A hybrid model for user authentication using keystroke and mouse dynamics with machine learning. Journal of Intelligent & Fuzzy Systems, 42(2), 1283–1296. https://doi.org/10.3233/JIFS-210883
18. Traore, I., Ahmed, A., & Woungang, I. (2013). Behavioral biometrics for continuous and transparent authentication. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(3), 531–546. https://doi.org/10.1109/TSMC.2012.2218613
19. Yang, Z., Hu, Y., Yu, Z., Wang, Y., & Li, J. (2020). Multi-modal behavioral biometrics authentication based on deep learning. IEEE Access, 8, 24690–24700. https://doi.org/10.1109/ACCESS.2020.2969195
20. Zhang, C., Sun, J., & Li, J. (2018). A deep learning-based behavior modeling approach for user identification in cloud systems. Future Generation Computer Systems, 86, 95–104. https://doi.org/10.1016/j.future.2018.03.028
21. Barchenko, N., Tolbatov, V., Lavryk, T., Obodiak, V., Shelehov, I., Tolbatov, A., Gnatyuk, S., & Tolbatova, O. (2022). Mathematical Model for Adaptive Technology in E-learning Systems. International Journal of Modern Education and Computer Science (IJMECS), 14(4), 1–15.

22. Lavrov, E., Tolbatov, A., Pasko, N., & Tolbatov, V. (2017). Cybersecurity of distributed information systems: The minimization of damage caused by errors of operators during group activity. Proceedings of the 2nd International Conference on Advanced Information and Communication Technologies (AICT 2017), Lviv, Ukraine, 83–87.
23. Lavrov, E., Tolbatov, A., Pasko, N., & Tolbatov, V. (2017). Ergonomic reserves for improving reliability of data processing in distributed banking systems. Proceedings of the 2nd International Conference on Advanced Information and Communication Technologies (AICT 2017), Lviv, Ukraine, 79–82.
24. Gnatyuk, S., Barchenko, N., Azarenko, O., Tolbatov, A., Obodiak, V., & Tolbatov, V. (2019). Ergonomic Support for Decision-Making Management of the Chief Information Security Officer. Proceedings of the 1st International Conference on Cyber Hygiene and Conflict Management in Global Information Networks (CyberConf 2019), Lviv, Ukraine, 459–471. Retrieved from http://ceur-ws.org/Vol-2588/
25. Legenkyi, M., Piankivska, L., & Tolbatov, A. (2024). Legal basis for cybersecurity in Ukraine under martial law. CH&CMiGIN, 3925. Retrieved from https://ceur-ws.org/Vol-3925/short10.pdf

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (30), 2025
Спеціальний випуск ISSN 2663 – 4023
Special issue

Ievgenii Sokyrka
postgraduate
State University "Kyiv Aviation Institute", Kyiv, Ukraine
ORCID: 0000-0003-2356-7213
sokyrka_evgeniy@yahoo.com

Ivan Kukulevskyi
postgraduate
State University "Kyiv Aviation Institute", Kyiv, Ukraine
ORCID: 0000-0002-2830-3607
ivan.kukulevskyi@outlook.com

Andrii Tolbatov
PhD, Associate Professor,
Doctoral Researcher at the Faculty of Cybersecurity, Computer and Software Engineering
State University "Kyiv Aviation Institute", Kyiv, Ukraine
ORCID: 0000-0002-9785-9975
tolbatov@ukr.net

AUTHENTICATION METHODS USING BEHAVIORAL ANALYTICS AND MACHINE LEARNING FOR INTERNET OF THINGS DEVICES

Abstract. The growing complexity of cyber threats has highlighted the limitations of traditional authentication methods, including passwords, tokens, and standard two-factor authentication (2FA). In the Internet of Things (IoT) environment, these methods are particularly vulnerable due to limited computational resources, the dynamic nature of connections, and the need for seamless user–device interaction. In response to these challenges, behavioral analytics and machine learning (ML) are gaining increasing attention as they enable the development of adaptive, continuous, and user-transparent authentication systems. This study focuses on behavioral authentication methods, including keystroke dynamics, mouse movement patterns, geolocation data, session activity, and network traffic analysis. A modular architecture is proposed that integrates both supervised and unsupervised ML algorithms, such as Support Vector Machines (SVM), Random Forest, Artificial Neural Networks (ANN), and autoencoders. Based on a combination of public and experimental datasets, extensive preprocessing and feature engineering were applied to identify the most informative behavioral characteristics of users and devices. Experimental results showed that the Random Forest model achieved the highest accuracy (96%) and F1-score (0.94), while the deployed prototype system provided fast response times (0.6 s) and a low false positive rate (0.1%) in a real-time web environment. These findings confirm the practical applicability of behavioral authentication methods for IoT, where classical approaches are often ineffective. At the same time, several key implementation challenges were identified: the need for large volumes of training data, ensuring the privacy of behavioral patterns, integration into heterogeneous IoT ecosystems, and maintaining a balance between performance and accuracy. Promising directions for further research include optimizing algorithms for resource-constrained devices and applying federated learning to minimize the risks of data leakage. Thus, behavioral analytics combined with ML forms a new paradigm of authentication, capable of providing a high level of information security in the context of the rapid expansion of IoT.

Keywords: behavioral biometrics; machine learning; user authentication; anomaly detection; cybersecurity.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Ahmed, A. A. E., & Traore, I. (2007). A new biometric technology based on mouse dynamics. IEEE Transactions on Dependable and Secure Computing, 4(3), 165–179. https://doi.org/10.1109/TDSC.2007.1003
2. Banerjee, S., & Woodard, D. L. (2012). Biometric authentication and identification using keystroke dynamics: A survey. Journal of Pattern Recognition Research, 7(1), 116–139.
3. Bergadano, F., Gunetti, D., & Picardi, C. (2002). User authentication through keystroke dynamics. ACM Transactions on Information and System Security (TISSEC), 5(4), 367–397. https://doi.org/10.1145/581271.581274
4. Bours, P., & Mondal, S. (2015). Performance evaluation of continuous authentication systems. Information Security Technical Report, 19(1), 45–52. https://doi.org/10.1016/j.istr.2014.09.003
5. Cao, K., & Jain, A. K. (2018). Hacking mobile phones using 2D printed fingerprints. Michigan State University, Technical Report.
6. Chen, F., et al. (2020). Behavioral biometrics for continuous authentication in the Internet of Things era: An overview. IEEE Internet of Things Journal, 7(8), 7118–7131. https://doi.org/10.1109/JIOT.2020.2975332
7. Fridman, L., Weber, S., Greenstadt, R., & Kam, M. (2015). Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location. IEEE Systems Journal, 11(2), 513–521. https://doi.org/10.1109/JSYST.2015.2453215
8. Gamboa, H., & Fred, A. (2004). A behavioral biometric system based on human-computer interaction. In Biometric Authentication (pp. 381–392). Springer. https://doi.org/10.1007/978-3-540-25948-0_45
9. Gunetti, D., & Picardi, C. (2005). Keystroke analysis of free text. ACM Transactions on Information and System Security (TISSEC), 8(3), 312–347.
10. Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. In Proceedings of the 2009 IEEE/IFIP International Conference on Dependable Systems & Networks (pp. 125–134). https://doi.org/10.1109/DSN.2009.5270348
11. Mondal, S., & Bours, P. (2015). A study on continuous authentication using a combination of keystroke and mouse biometrics. Neurocomputing, 230, 1–22. https://doi.org/10.1016/j.neucom.2016.12.007
12. Morales, A., Fierrez, J., & Ortega-Garcia, J. (2016). Keystroke dynamics recognition based on personal data: A comparative experimental evaluation. Pattern Recognition Letters, 79, 20–27. https://doi.org/10.1016/j.patrec.2016.03.007
13. Pusara, M., & Brodley, C. E. (2004). User re-authentication via mouse movements. In Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security (pp. 1–8). https://doi.org/10.1145/1029208.1029210
14. Revett, K., de Magalhães, S. T., & Santos, H. P. (2008). Behavioral biometrics: A remote access approach. International Journal of Information Security, 7, 243–258. https://doi.org/10.1007/s10207-008-0056-2
15. Shen, C., Cai, Z., Guan, X., & Du, X. (2013). A privacy-preserving protocol for secure and efficient user authentication. Computer Standards & Interfaces, 35(2), 240–246.
16. Shen, C., Guan, X., & Cai, Z. (2011). Continuous authentication for mouse dynamics: A pattern-growth approach. International Journal of Information Security, 10, 221–229. https://doi.org/10.1007/s10207-011-0120-2
17. Tiwari, A., & Gupta, A. (2022). A hybrid model for user authentication using keystroke and mouse dynamics with machine learning. Journal of Intelligent & Fuzzy Systems, 42(2), 1283–1296. https://doi.org/10.3233/JIFS-210883
18. Traore, I., Ahmed, A., & Woungang, I. (2013). Behavioral biometrics for continuous and transparent authentication. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 43(3), 531–546. https://doi.org/10.1109/TSMC.2012.2218613
19. Yang, Z., Hu, Y., Yu, Z., Wang, Y., & Li, J. (2020). Multi-modal behavioral biometrics authentication based on deep learning. IEEE Access, 8, 24690–24700. https://doi.org/10.1109/ACCESS.2020.2969195
20. Zhang, C., Sun, J., & Li, J. (2018). A deep learning-based behavior modeling approach for user identification in cloud systems. Future Generation Computer Systems, 86, 95–104. https://doi.org/10.1016/j.future.2018.03.028
21. Barchenko, N., Tolbatov, V., Lavryk, T., Obodiak, V., Shelehov, I., Tolbatov, A., Gnatyuk, S., & Tolbatova, O. (2022). Mathematical Model for Adaptive Technology in E-learning Systems. International Journal of Modern Education and Computer Science (IJMECS), 14(4), 1–15.

22. Lavrov, E., Tolbatov, A., Pasko, N., & Tolbatov, V. (2017). Cybersecurity of distributed information systems: The minimization of damage caused by errors of operators during group activity. Proceedings of the 2nd International Conference on Advanced Information and Communication Technologies (AICT 2017), Lviv, Ukraine, 83–87.
23. Lavrov, E., Tolbatov, A., Pasko, N., & Tolbatov, V. (2017). Ergonomic reserves for improving reliability of data processing in distributed banking systems. Proceedings of the 2nd International Conference on Advanced Information and Communication Technologies (AICT 2017), Lviv, Ukraine, 79–82.
24. Gnatyuk, S., Barchenko, N., Azarenko, O., Tolbatov, A., Obodiak, V., & Tolbatov, V. (2019). Ergonomic Support for Decision-Making Management of the Chief Information Security Officer. Proceedings of the 1st International Conference on Cyber Hygiene and Conflict Management in Global Information Networks (CyberConf 2019), Lviv, Ukraine, 459–471. Retrieved from http://ceur-ws.org/Vol-2588/
25. Legenkyi, M., Piankivska, L., & Tolbatov, A. (2024). Legal basis for cybersecurity in Ukraine under martial law. CH&CMiGIN, 3925. Retrieved from https://ceur-ws.org/Vol-3925/short10.pdf