УДК 004.85
DOI: 10.25140/2411-5363-2021-2(24)-114-122

Анастасія Косарева, Павло Регіда

ЗАСІБ ДЛЯ БІОМЕТРИЧНОЇ АВТЕНТИФІКАЦІЇ НА ОСНОВІ ПОВЕДІНКОВИХ ОСОБЛИВОСТЕЙ КОРИСТУВАЧА

У роботі наведено результати дослідження існуючих методів поведінкової біометрії, а саме біометрії за особливостями користувача при наборі тексту. Також у роботі описано та проаналізовано власний метод автентифікації користувача за його динамікою натискання клавіш. Запропоновано розроблений авторами метод, описаний у статті, як програмний додаток для проведення різноманітних досліджень розпізнавання особистості за її поведінковими особливостями.

Ключові слова: *біометрія; автентифікація; поведінкові особливості; динаміка натискання клавіш; нейронні мережі.*
Рис.: 9. Бібл.: 9.

Постановка проблеми. Розвиток технологій світу дозволяє людям оточити себе мобільними девайсами, які поєднують у собі багато функцій, однією з яких є зберігання персональних даних користувачів. Тому досить значущим питанням є забезпечення недоторканності цих даних ані хакерами, ані зловмисниками. Тому на додаток, а іноді на заміну паролів прийшла біометрична автентифікація. На сьогодні найбільш розповсюдженими методами є розпізнавання користувачів за відбитком пальця або за геометрією обличчя. Але й такі персональні особливості людини можливо оприлюднити. Тому дослідження на тему поведінкової біометрії є дуже необхідними, адже такі методи автентифікації не потребують від користувача додаткових зусиль та можуть доповнювати систему безпеки.

Аналіз останніх досліджень. Дослідження, що будуть розглядатися далі, пов'язані між собою спільною темою розпізнавання користувача, спираючись на його особливості під час друку тексту на віртуальній клавіатурі смартфону. Система відрізняє користувача від зловмисника не за тим, що він друкує, а за тим, як він друкує.

Сам метод розпізнавання користувача за динамікою його друку поділяється на дві групи, а саме групу статичного розпізнавання та групу динамічного розпізнавання. Статичне розпізнавання базується на автентифікації користувача під час введення статичних текстових рядків, наприклад, логіну та паролю. Динамічне розпізнавання базується на особливостях друкування рядка тексту, що є невідомим для користувача до початку набору тексту.

Однією з робіт, що будуть розглядатися, є стаття авторства Маттіаса Трояна та Френка Ортмаєра [1], у якій вони поєднали методи розпізнавання людини за її поведінковими особливостями під час друку тексту та за почерком особистості, що було імплементовано у смартфони. Було проведено два експерименти для збирання даних та їх подальшого аналізу. Для проведення експерименту використовувались пристрої на операційній системі Android, із використанням яких було необхідно вводити паролі або невеликі речення з двох слів та більше. Для першого експерименту було залучено вісімнадцять волонтерів, які мали ввести те саме речення десять разів. У другому експерименті було заохочено ще шістнадцять учасників, які мали ввести пароль вісім разів. Як класифікатор автори використовували такі технології, як дерево прийняття рішень [2], багатошаровий персептрон (MLP) [3] та Bayes Net [4].

Метою другої статті [5], на яку спирається це дослідження, є дослідження підходу, що був розроблений авторами, що заснований на автентифікації користувача за особливостями його друку на віртуальній клавіатурі смартфону. Ця реалізація поділяється на дві фази: фаза автентифікації користувача та фаза використання додатку користувачем. Перший етап валідує особистість користувача завдяки паролю та його біометричної інформації. На другому етапі, що є пасивним, вже під час користування додатку, дії людини відстежуються під час набирання тексту в полях для вводу. Цей додаток постійно працює у фоновому режимі для постійної автентифікації користувача.

Для проведення експерименту було набрано 40 волонтерів, які користувалися додатком протягом трьох тижнів, увесь цей час накопичувалася база даних про кожного з учасників. Усього було задіяно десять смартфонів на базі операційної системи Android. На першому етапі додатка вводився пароль пін-код із чотирьох символів (всього було використано 20 різних пін-кодів), на другому етапі вводиться речення до 53 слів. Із зібраних даних вилучається інформація про те, як користувач друкує текст, як натискає на клавіші віртуальної клавіатури, скільки часу на це витрачає та інше. Як класифікатори використовувались дерева рішень [2], Bayes Net [4].

Третя стаття [6] описує аналіз динаміки натискання клавіш віртуальної клавіатури смартфону, використовуючи двоетапну схему – реєстрація та автентифікація. Автори використовували два типи паролів (пін-код та багатосимвольний пароль). Головною метою цього застосунку є розпізнавання подій натискання клавіш, обчислення часу затримки між взаємодіями з екраном. Для експерименту було обрано 20 волонтерів для накопичення даних, кожен з яких мав набрати відомий йому пароль двадцять разів для етапу реєстрації та ще 10 разів для етапу автентифікації.

Автори використовували два класифікатори: евклідова відстань [7] та відстань Махаланобіса [8], обидва з низькими вимогами до обробки. Вони також використовували нейронний багатошаровий перцептрон (FF-MLP) мережі [9] для підвищення ефективності підходу.

Аналізуючи підходи останніх досліджень, можна дійти висновку, що автори статей обмежувались лише декількома методами розпізнавання особи за її поведінковими особливостями та не передбачають масштабування своєї системи в майбутньому, що значно зменшує шанси програми на розвиток у якості незалежного додатку чи елемента системи захисту. Чим більше методів розпізнавання особи за її поведінковими особливостями буде імплементовано в систему, тим більшу кількість персональних маркерів буде можливо відстежувати.

Формулювання цілей. Осільки методи розпізнавання користувача за його поведінковими особливостями є мало дослідженою сферою діяльності в біометричній аутентифікації, то як одну із цілей можна виділити створення власного методу розпізнавання користувача, його аналіз та порівняння з існуючими рішеннями. Також, щоб поліпшити ситуацію у світі науки у сфері поведінкової біометрії, метою є поширення написаного методу, надання його у відкритий доступ. Буде проведено огляд використаних засобів розробки та реалізації, а також продемонстровано результати роботи додатку, результати роботи нейронної мережі та графіки, що демонструють працездатність створеного застосунку.

Виклад основного матеріалу. Як вже було зазначено раніше, у цій статті розглянуто вже існуючі методи розпізнавання людини за її поведінковими особливостями та пропонує власний метод поведінкової біометричної аутентифікації, який буде розглянуто у даному розділі. Головною метою створення цього додатка є надання його у відкритий доступ для того, щоб науковці та розробники, що зацікавлені в темі поведінкової біометрії могли б безперешкодно використовувати або удосконалювати створене рішення. Створена програма є необхідною та унікальною, адже компанії, що використовують засоби поведінкової біометрії, не поширюють використані алгоритми, бо це ставить під загрозу безпеку даних їхніх клієнтів та користувачів.

Суттю розробленого застосунку є розпізнавання користувача за його поведінковими особливостями, такими як особливості набору тексту на віртуальній клавіатурі смартфону та особливості взаємодії зі смартфоном під час користування ним. Розроблений додаток здатний розпізнавати кут нахилу смартфону до кожної з осей X, Y та Z (розраховується кут орієнтації смартфону у просторі відносно одиничного вектора кожної з

осей). Ці дані збираються під час вводу користувачем логіну та паролю та відправляються на сервер після натисканні на кнопку, якщо логін та пароль є достовірними (вважається, що користувач вже є зареєстрованим в системі та йому відомі логін та пароль). Отримані дані розподіляються за такими критеріями:

- час введення логіну;
- час введення паролю;
- час від початку заповнення полів для входу до натискання на кнопку авторизації;
- швидкість введення логіну (обчислюється як кількість введених символів за секунду);
- швидкість введення паролю (обчислюється як кількість введених символів за секунду);
- кількість натискань кнопки backspace під час введення логіну;
- кількість натискань кнопки backspace під час введення паролю;
- максимальний та мінімальний кут відносно осі X;
- максимальний та мінімальний кут відносно осі Y;
- максимальний та мінімальний кут відносно осі Z.

Отриманий стек даних формується в об'єкті, який заздалегідь вважається «достовірним», тобто за отриманими даними від користувача надалі буде натренована нейронна мережа. Отримані дані зберігаються у файлі типу json з міткою Accept при виводі результатів. Частина, сформована за даними користувача, зображена на рис. 1.


```
{
  input: {
    LoginEnter: 5.321,
    PasswordEnter: 2.492,
    LoginPasswordEnter: 8.544,
    LoginSymbPerSec: 2.78260869565217,
    PasswordSymbPerSec: 6.38977635782748,
    LoginBackSpace: 0,
    PasswordBackSpace: 0,
    AlphaMin: -100.10572500418643,
    AlphaMax: 0,
    BettaMin: 0,
    BettaMax: 33.46849725385817,
    GammaMin: -3.09186804880998,
    GammaMax: 16.7154775595426
  },
  output: { Accept: 1 }
}
```


Рис. 1. Сформований стек даних за результатами вводу користувача.

Частина input є даними, за якими тренується нейронна мережа, а частина output дає нейронній мережі відповідь на питання, чи це дійсно користувач або все ж таки самозванець, що намагається видати себе за зареєстровану особу. Повідомлення Accept означає, що користувач був успішно авторизований. Алгоритм розробленої програми, а саме частини тренування нейронної мережі на розпізнавання особи за її поведінковими особливостями, зображено на рис. 2, *а*.

На рис. 2, *б* зображено алгоритм запуску вже натренованої нейронної мережі, тобто алгоритм розпізнавання особи за його поведінковими особливостями, за якими була тренована нейронна мережа. Ці алгоритми дуже схожі, але описують різні функції: train та run, залежно від натиснутої кнопки у графічному інтерфейсі.

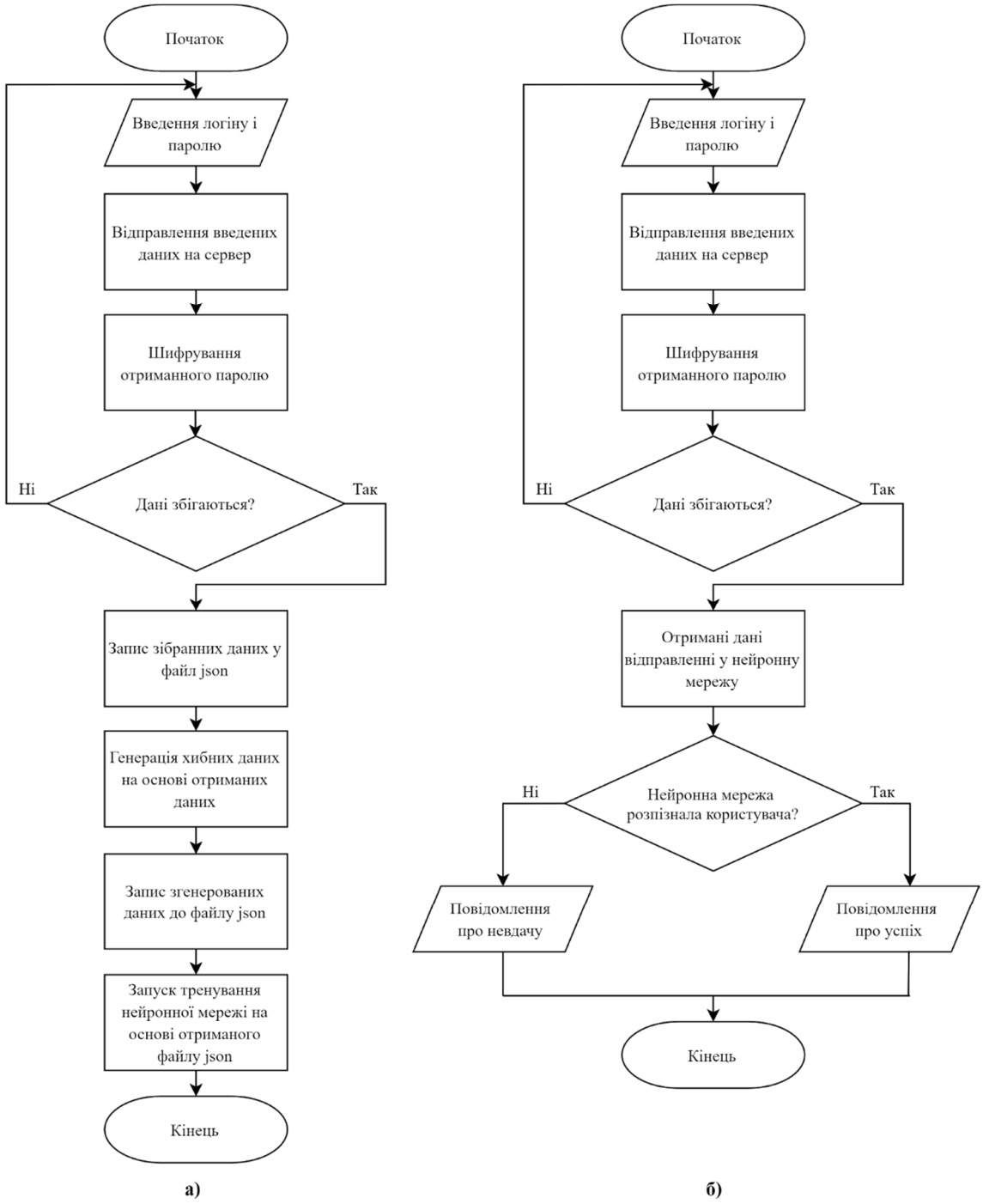

Рис. 2. Алгоритм розробленої системи розпізнавання за поведінковою біометрією: а – алгоритм тренування нейронної мережі; б – алгоритм запуску нейронної мережі

Розроблена програма є клієнт-серверним додатком для мобільних платформ, а саме Android та iOS, що робить його універсальним для більшої кількості смартфонів на ринку. Кросплатформеність обумовлена розробкою клієнтної частини програми за допомогою фреймворку мови JavaScript – React Native. Серверна частина проєкту розроблена завдяки програмній платформі node.js.

Основою розробленого проєкту є нейронна мережа Brain.js. Саме цей інструмент аналізує отримані дані, розподіляючи їх на дві категорії – Accept і Reject, а також виконує розпізнавання особи за отриманими даними, формуючи відповідь запиту клієнта. Структурна схема розробленого застосунку зображена на рис. 3.

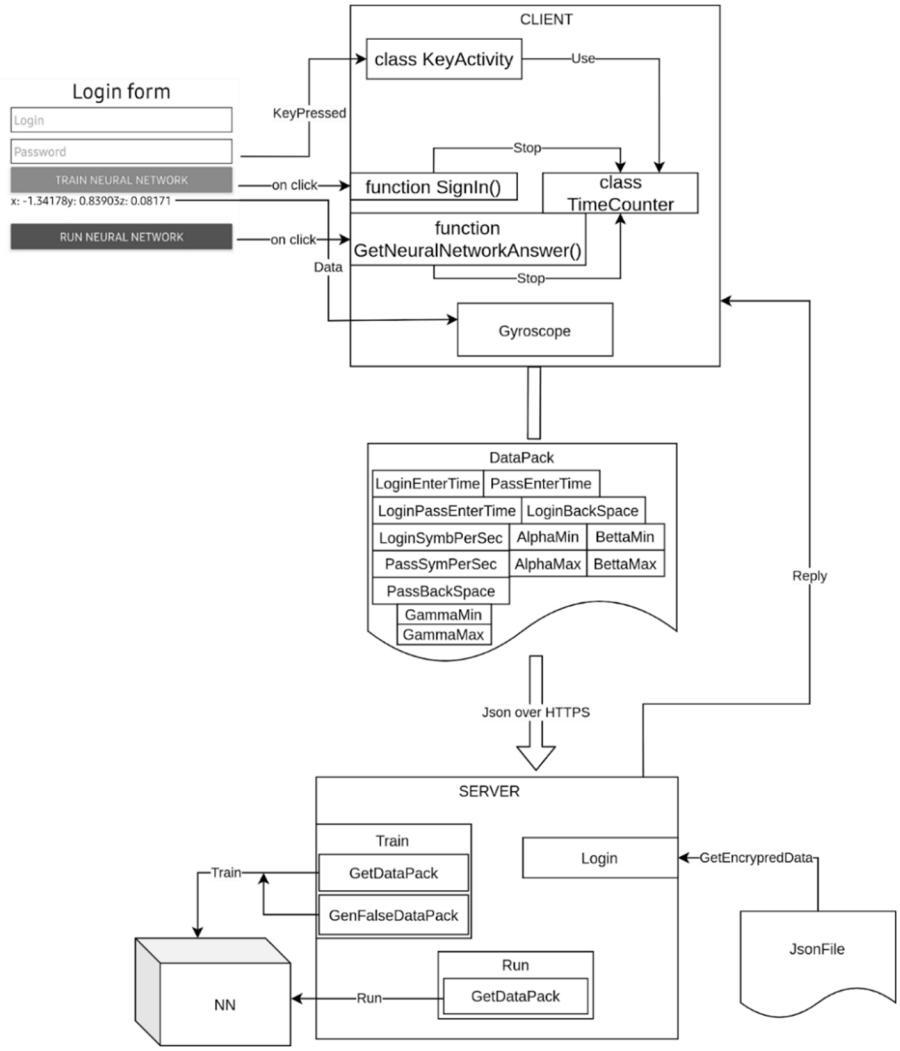

Рис. 3. Структурна схема розробленого клієнт-серверного додатку

У результаті роботи можливо отримати дві відповіді, а саме Accept і Reject. Результати обох випадків зображені на рис. 4. Також рис. 4 ілюструє графічну складову частини клієнта, а саме дві кнопки, які відповідають за тренування нейронної мережі та її запуску відповідно. Є два окремих текстових поля для вводу паролю та логіну. Також на екрані присутні значення кутів нахилу смартфону відносно кожної з осей координат (X, Y і Z). Ці значення представлені в радіанах.

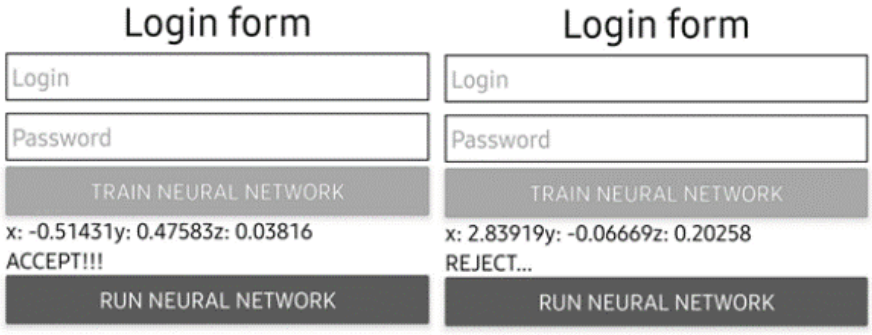

Рис. 4. Графічний інтерфейс клієнта з двома варіантами відповіді системи залежно від отриманих даних

Було проведено декілька експериментів у розробленій програмі, щоб оцінити показники помилок першого та другого роду, доцільність використання великої кількості тренувальних даних, залежність часу, що витрачається на тренування, та ймовірності похибки залежно від кількості тренувальних даних. Результати цих експериментів будуть представлені нижче.

Для першого експерименту було вирішено порівнювати відгуки системи на різну кількість вхідних даних. Нейронна мережа відповідає на отриманий пакет даних для запуску функції, результати чого зафіксовані та відображені у вигляді графіків залежності ймовірностей відповіді Accept та відповіді Reject. Відповіді системи з кількістю тренувальних даних до 100 входжень були досить поширені, та не давали чіткої відповіді, але починаючи з 500 входжень та далі, система поводила себе дедалі краще. На рис. 5 зображені графіки, на яких система спочатку мала відсіяти реальних користувачів (рис. 5, *а*), а потім намагатися розпізнати зловмисників (рис. 5, *б*).

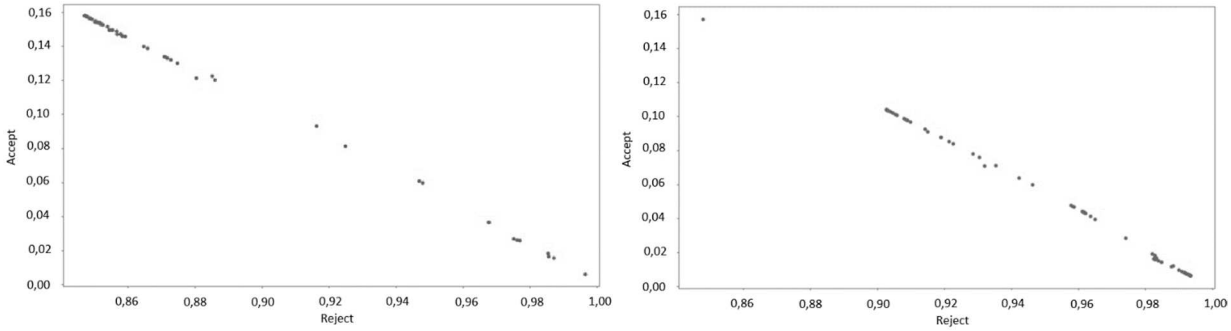

Рис. 5. Результати експерименту на 1000 входжень:
а – розпізнавання користувача; б – уникнення самозванця

Звичайно, чим більше входжень на тренування нейронної мережі (ітерацій), тим краще вона поводить себе в реальному використанні. Для наступного експерименту було використано 3500 вхідних даних як для тренування нейронної мережі, так і для її запуску. Ці результати були наближені до ідеальних. Більшість входжень на функції запуску наближаються до одиниці – максимальної ймовірності дозволу чи відторгнення. Результати цього експерименту зображені на рис. 6.

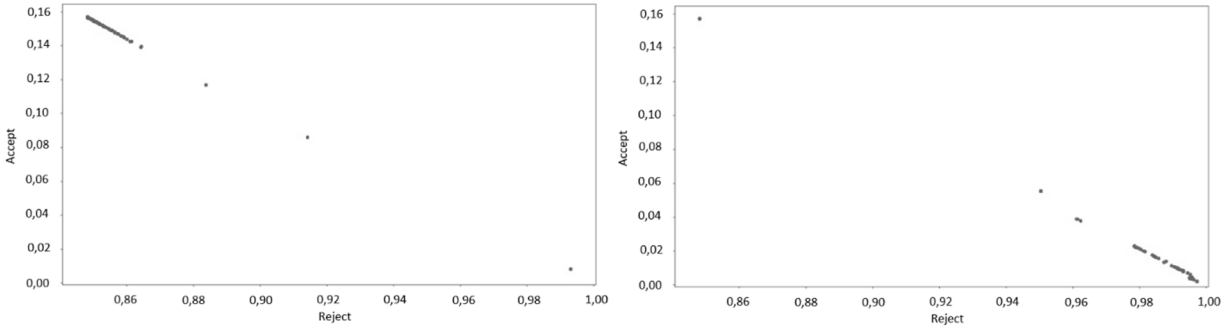

Рис. 6. Результати експерименту на 3500 входжень:
а – розпізнавання користувача; б – уникнення самозванця

За отриманими даними було сформовано графіки, зображені на рис. 7, 8 та 9. А саме графік залежності ймовірності похибки відповіді нейронної мережі від кількості даних для тренування, графік залежності часу тренування нейронної мережі від кількості даних для тренування та графік, що ілюструє залежності часу тренування від ймовірності похибки на основі кількості даних для тренування відповідно.

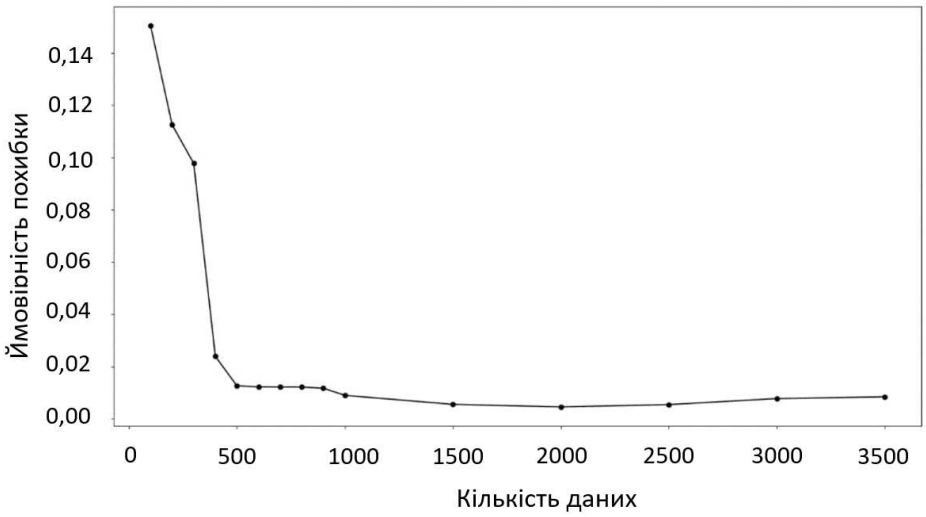

Рис. 7. Графік залежності ймовірності похибки від кількості даних для тренування

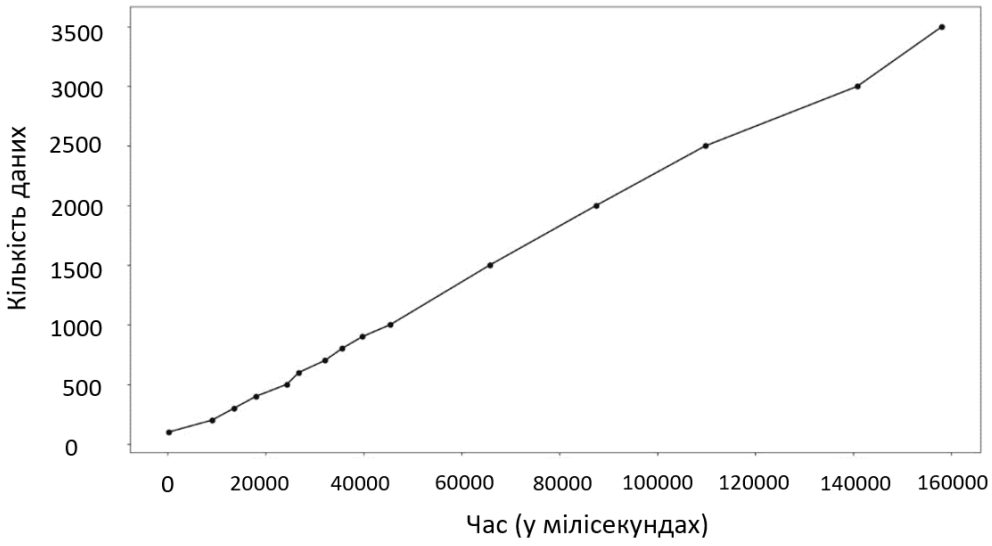

Рис. 8. Графік залежності часу тренування нейронної мережі від кількості даних для тренування

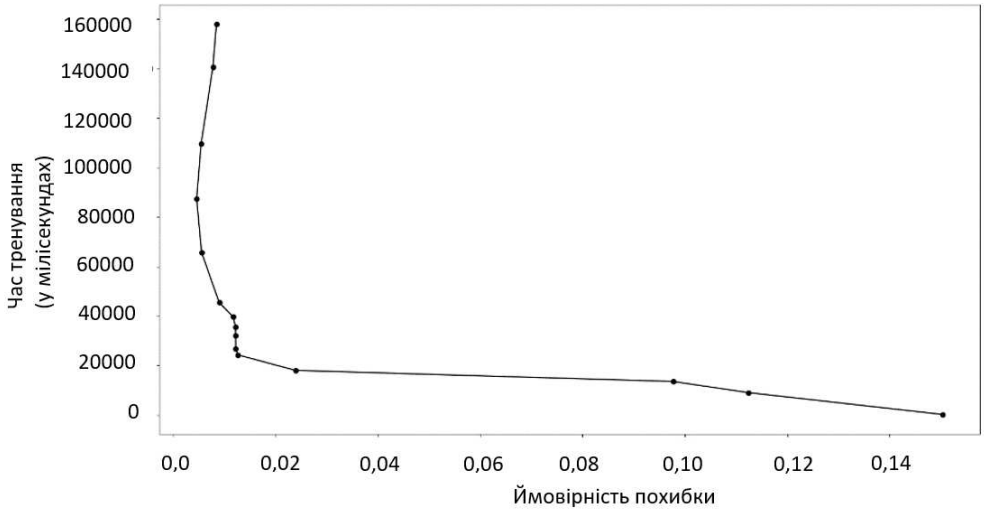

Рис. 9. Графік відповідності часу тренування нейронної мережі до ймовірності похибки

З отриманих даних можна зробити висновок, що кількість входжень від 2000 до 2500 є оптимальною, бо використовує середню кількість ресурсів комп'ютера, займає в середньому 9 секунд на тренування та надання відповіді. Звичайно, ймовірність похибки не є ідеальною, але чим менше цей показник, тим більше часу витрачається на тренування нейронної мережі. Згідно з графіками на рис. 7, 8, 9 максимальний допустимий показник ймовірності похибки є 2000 входжень.

Висновки. У цій роботі було висвітлено проблему недостатньої дослідженості теми розпізнавання осіб за їхніми поведінковими особливостями, та надано рішення як публікація коду цієї системи у відкритому доступі, що дає можливість для інших науковців та розробників, які зацікавлені в цій темі, використовувати розроблений додаток у власних дослідницьких цілях, а також продовжувати розробку з додаванням інших методів поведінкової біометрії та поліпшення алгоритму навчання нейронної мережі. Були розглянуті засоби реалізації розробленого додатку, стек використаних технологій, такі як мова програмування, фреймворк для розроблення кросплатформенних мобільних додатків, програмне забезпечення для підтримання власного серверу, засіб тренування та запуску нейронних мереж та інше. Розроблена програма була детально проаналізована, про що свідчать надані результати проведених експериментів. Запропонований додаток демонструє досить непогані результати, спираючись на отримані графіки результатів, що свідчить про необхідність продовження розробки й додавання та інших поведінкових методів біометричної аутентифікації.

Програму необхідно розвивати як платформу для тестування різноманітних методів поведінкової біометрії, для аналізування спроб зловмисників відтворити відомі йому особливості користувача. Також, слід імплементувати реальну базу даних для зберігання даних для кожного із зареєстрованих користувачів.

Список використаних джерел

1. Trojahn M., Ortmeier F. Toward mobile authentication with keystroke dynamics on mobile phones and tablets. *Advanced Information Networking and Applications Workshops (WAINA), 27th International Conference on. IEEE.* 2013. Pp. 697–702). URL: https://doi.org/10.1109/WAINA.2013.36.

2. Quinlan J. R. Bagging, boosting, and c4. 5. *AAAI/IAAI*. 1996. 1. Pp. 725–730. URL: https://www.aaai.org/Papers/AAAI/1996/AAAI96-108.pdf.

3. Pal S. K., Mitra S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks.* 1992. Vol. 3. Pp. 683–697. URL: https://doi.org/10.1109/72.159058.

4. Friedman N., Geiger D., Goldszmidt M. (1997). Bayesian network classifiers. *Machine learning.* Vol. 29. Pp. 131–163. URL: http://www.cs.technion.ac.il/~dang/journal_papers/friedman1997Bayesian.pdf.

5. Buchoux A., Clarke N. L. Deployment of keystroke analysis on a smartphone. *Australian Information Security Management Conference.* 2008. P. 48. URL: https://www.researchgate.net/publication/49282754_Deployment_of_Keystroke_Analysis_on_a_Smartphone.

6. Draffin B., Zhu J., Zhang J. Keysens: passive user authentication through micro-behavior modeling of soft keyboard interaction. *International Conference on Mobile Computing, Applications, and Services.* 2013. Vol. 130. Pp. 184–201. URL: https://dx.doi.org/10.1007/978-3-319-05452-0_14.

7. Danielsson P. E. Euclidean distance mapping. *Computer Graphics and image processing.* 1980. Vol. 14. Pp. 227–248. URL: https://doi.org/10.1016/0146-664X(80)90054-4.

8. De Maesschalck R., Jouan-Rimbaud D., Massart D. L. The mahalanobis distance. *Chemometrics and intelligent laboratory systems.* 2000. Vol. 50. Pp. 1–18. URL: https://doi.org/10.1016/S0169-7439(99)00047-7.

9. Bishop C. M. Neural networks for pattern recognition. Oxford university press, 1995.

References

1. Trojahn, M., & Ortmeier, F. (2013). Toward mobile authentication with keystroke dynamics on mobile phones and tablets. *Advanced Information Networking and Applications Workshops (WAINA), 27th International Conference on. IEEE* (pp. 697–702). https://doi.org/10.1109/WAINA.2013.36.

2. Quinlan, J. R. (1996). Bagging, boosting, and c4. 5. *AAAI/IAAI,* (1), 725–730. https://www.aaai.org/Papers/AAAI/1996/AAAI96-108.pdf.

3. Pal, S. K., & Mitra, S. (1992). Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on Neural Networks,* (3), 683–697. https://doi.org/10.1109/72.159058.

4. Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine learning,* (29),131-163. http://www.cs.technion.ac.il/~dang/journal_papers/friedman1997Bayesian.pdf.

5. Buchoux, A., & Clarke, N. L. (2008). Deployment of keystroke analysis on a smartphone. *Australian Information Security Management Conference,* (p. 48). https://www.researchgate.net/publication/49282754_Deployment_of_Keystroke_Analysis_on_a_Smartphone.

6. Draffin, B., Zhu, J., & Zhang, J. (2013). Keysens: passive user authentication through micro-behavior modeling of soft keyboard interaction. *International Conference on Mobile Computing, Applications, and Services,* (130), 184–201. https://dx.doi.org/10.1007/978-3-319-05452-0_14.

7. Danielsson, P. E. (1980). Euclidean distance mapping. *Computer Graphics and image processing,* (14), 227–248. https://doi.org/10.1016/0146-664X(80)90054-4.

8. De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The mahalanobis distance. *Chemometrics and intelligent laboratory systems,* (50), 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7.

9. Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.

UDC 004.85

Anastasia Kosareva, Pavlo Rehida

TOOL FOR BIOMETRIC AUTHENTICATION BASED ON USER BEHAVIORAL FEATURES

Biometric authentication is one of the most common ways to identify a person. However, the topic of behavioral biometrics is almost unstudied, and those authentication methods can significantly increase the level of security of personal data, but it is still not implemented in modern devices due to lack of an existing researches and scientific works.

It is a very important issue to ensure the integrity of personal data by hackers or attackers, so password recognition is being supplemented by biometric authentication. Today, the most common methods are to recognize users by fingerprint or facial geometry. But even such personal characteristics can be made public. Therefore, research on behavioral biometrics is essential, as such authentication methods do not require additional effort from the user and can complement the security system.

There are currently such behavioral biometrics methods that use keyboard typing dynamics, keystroke type strength, typing speed, and other criteria studied in described papers by authors such as Abdulaziz Ali Alzubaidi, Yugal Kalita, Troyan and Ortmeyer, and others.

Since the methods of user recognition by its behavioral characteristics are not studied enough in the field of biometric authentication, one of the goals is to create authors' own method of user recognition, analysis and comparison with existing solutions.

The aim of the paper is to create a platform for analyzing various methods of behavior authentication and its spreading through open source policy. This will help researchers and scientists to conduct their own experiments and improve the level of protection by behavioral biometrics methods.

The created method of user recognizing by their behavioral features is described, namely by means of dynamics of keystrokes, speed of typing, angle of inclination of a smartphone during its use. The neural network, which is trained for personality recognition, responds to the generated data packet to run the function, the results of which are recorded and displayed in the form of graphs of the probabilities of the Accept response and the Reject response.

The system of user recognition by their behavioral features, which is provided in open source, is developed, the created program is analyzed and results of its work in the form of graphs are provided. It was concluded that the maximum allowable error rate is 2000 data occurrences.

Keywords: *biometrics; authentication; behavioral characteristics; keystroke dynamics; neural networks.*

Косарева Анастасія Сергіївна – студентка, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (просп. Перемоги, 37, м. Київ, 03056, Україна).
Kosareva Anastasiia – Student, National Technical University of Ukraine «Kyiv Polytechnic Institute named after Igor Sikorsky» (37 Pobedy Av., 03056 Kyiv, Ukraine).
E-mail: asya.kosareva666@gmail.com
ORCID: https://orcid.org/0000-0002-9439-5984

Регіда Павло Геннадійович – асистент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (просп. Перемоги, 37, м. Київ, 03056, Україна).
Rehida Pavlo – Assistant, Department of Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (37 Pobedy Av., 03056 Kyiv, Ukraine).
E-mail: pavel.regida@gmail.com
ORCID: https://orcid.org/0000-0002-6591-7069

Косарева А., Регіда П. Засіб для біометричної автентифікації на основі поведінкових особливостей користувача. *Технічні науки та технології.* 2021. № 2(24). С. 114-122.