**DOI: 10.18372/2410-7840.21.13546**
**УДК 004.056.523:57.087.1**

# МЕТОД АВТЕНТИФІКАЦІЇ КОРИСТУВАЧІВ ІНФОРМАЦІЙНИХ СИСТЕМ ЗА ЇХ РУКОПИСНИМ ПОЧЕРКОМ З БАГАТОКРОКОВОЮ КОРЕКЦІЄЮ ПЕРВИННИХ ДАНИХ

*Олександр Корченко, Анатолій Давиденко, Олена Висоцька*

*Стаття присвячена біометричній автентифікації користувачів, а саме автентифікації за рукописним почерком. В даній роботі аргументована актуальність створення системи біометричної автентифікації користувачів інформаційних систем за їх рукописним почерком. Після чого визначена множина характеристик рукописного почерку для подальшого їх використання для автентифікації. На основі проведеного аналізу обраних характеристик, визначена їх придатність для подальшого їх використання під час розпізнавання користувачів. Розроблено метод автентифікації користувачів інформаційних систем за їх рукописним почерком та метод необхідної первинної обробки зразків рукописного почерку користувачів інформаційних систем. Необхідність первинної обробки викликана специфікою використання, для динамічної передачі зображення в комп'ютер, графічного планшету (або іншого пристрою з сенсорним екраном). Ця обробка полягає в видаленні помилкових даних та в корекції даних, які будуть використовуватись для розпізнавання. В роботі виділені п'ять типів помилок та три типи корекції даних. Для поліпшення процесу розпізнавання, зображення написаної ключової фрази, для подальшого використання, розділяється на зображення окремих символів. Відповідно, до користувачів висувається умова, що символи ключової фрази, що вводиться, повинні бути написані окремо один від одного. Під час розпізнавання, аналізуються параметри не всіх точок зображення, а тільки найбільш значущих, контрольних точок. В роботі виділені три типи контрольних точок та аргументована значимість використання для цього найбільш оптимального алгоритму. Механізмом розпізнавання було обрано один з видів нейронних мереж, а саме імовірнісну нейронну мережу. На основі запропонованих методів розроблено програмне забезпечення, використовуючи яке, спочатку була сформована база даних навчальних зразків рукописного почерку користувачів інформаційної системи. Потім було проведено ряд експериментів для визначення ефективності застосування розроблених методів та для виявлення найбільш значущих, для правильного розпізнавання, налаштувань системи автентифікації користувачів. Наприкінці було зроблено висновок, що незважаючи на те, що запропоновані в даній роботі методи, дозволяють досягти досить високу імовірність правильного розпізнавання користувачів інформаційних систем, пошук більш ефективних механізмів розпізнавання та інших параметрів, які суттєво впливають на імовірність правильного розпізнавання, залишається доволі актуальною задачею.*

***Ключові слова:*** *автентифікація, розпізнавання, біометрія, рукописний почерк, інформаційні системи.*

## Вступ

Постійне зростання рівня обчислювальної техніки приводить до збільшення кола областей її використання. Залучення інформаційних систем (ІС) до таких видів діяльності, як документообіг, фінансові операції, маркетингові дослідження та інші, пов'язане з необхідністю організації безпечного доступу до цих ІС. Для цього в будь-якій подібній системі, повинні бути реалізовані функції автентифікації користувачів (АК). Існує низка відповідних методів, серед яких автентифікація за допомогою паролів, біометричних характеристик, ключів, електронного цифрового підпису, тощо [1-7]. Враховуючи зручність застосування біометричних систем, більшість користувачів ІС віддають перевагу саме їм. А враховуючи той факт, що багато сучасних пристроїв забезпечені сенсорним екраном (графічні планшети, сенсорний екран ноутбуків, планшетів, смартфонів) робить доцільним, для вирішення задачі АК, аналізувати таку біометричну характеристику користувача, як його рукописний почерк (РП) [6-9]. Таким чином, можна сказати, що створення системи АК за їх РП, яка забезпечить високий рівень показників імовірності правильно розпізнавання користувачів, є актуальною задачею.

## Постановка задачі

Метою даної роботи є розробка ефективного методу АК ІС за їх РП, який забезпечить високий рівень значень імовірності правильно розпізнавання користувачів і який можна було б використовувати в системах АК ІС в організаціях різних сфер діяльності людини, при наявності в цих організаціях пристроїв, що забезпечені сенсорним екраном.

Для вирішення поставленої задачі буде зроблено наступне:

1. Визначена множина характеристик почерку для подальшого їх використання для автентифікації.
2. Проведено аналіз обраних для автентифікації характеристик РП, для визначення їх придатності для подальшого розпізнавання.

3. Розроблено метод необхідної первинної обробки зразків РП користувачів ІС.

4. Розроблено метод АК ІС за їх РП.

5. На основі розроблених методів буде розроблено програмне забезпечення.

6. Сформована база даних навчальних зразків (БДНЗ).

7. За допомогою розробленої програми, буде проведено ряд експериментів по визначенню ефективності застосування розроблених методів та виявленню найбільш значущих налаштувань розробленої системи АК, для збільшення імовірності вірного розпізнавання.

**Вирішення поставленої задачі**

Для вирішення поставленої задачі була проаналізована низка методів АК ІС за їх РП [8-9] і далі було розроблено метод АК ІС за їх РП (АКРП) та метод первинної обробки зразків РП користувачів ІС (ПОЗРП), необхідний для виконання АК ІС за їх РП. Механізмом розпізнавання була обрана імовірнісна нейронна мережа (ІНМ) [10]. Для передачі зразка РП користувача в комп'ютер, використовувався графічний планшет (ГП), як один з варіантів пристрою з сенсорним екраном, який використовується для динамічного передавання характеристик РП користувача. Тобто в даному розділі будується функція $R$ – функція розпізнавання користувачів за паролем, який вводиться за допомогою ГП.

Для розробки методів АКРП та ПОЗРП використовується наступна модель даних:

$$\boldsymbol{US}=\{\bigcup_{p=1}^{m} US_p\}=\{US_1, US_2, \ldots, US_p, \ldots, US_x \ldots, US_m\};\ p=\overline{1,m};$$

$m$ – кількість членів множини $\boldsymbol{US}$; $\boldsymbol{US}$ – множина користувачів, які можуть спробувати отримати доступ до системи, що захищається;

$$\boldsymbol{USL}=\{\bigcup_{t=1}^{l} USL_t\}=\{USL_1, USL_2, \ldots, USL_t, \ldots, USL_d, \ldots, USL_l\};\ t=\overline{1,l};$$

$l$ – кількість легальних користувачів; $\boldsymbol{USL}$ – множина легальних користувачів даної системи; здійснюється умова, що $\boldsymbol{USL} \subset \boldsymbol{US}$;

$US_X$ – користувач ІС, який проходить автентифікацію; здійснюється умова, що $US_X \in \boldsymbol{US}$;

$USL_d$ – легальний користувач, за якого видає себе авторизована сторона, що автентифікується; здійснюється умова, що $((USL_d \in \boldsymbol{US}) \wedge (USL_d \in \boldsymbol{USL}))$;

$$\boldsymbol{USB}=\{\bigcup_{tb=1}^{lb} USB_{tb}\}=\{USL_1, USL_2, \ldots, USL_{tb}, \ldots, USL_{lb}\};\ tb=\overline{1,lb};$$

$lb$ – кількість порушників даної системи; $\boldsymbol{USB}$ – множина порушників даної системи, тобто користувачів, які в ній не зареєстровані, але намагаються отримати до неї доступ; здійснюється умова, що $((\boldsymbol{USB} \subset \boldsymbol{US}) \wedge (\boldsymbol{USB} \not\subset \boldsymbol{USL}))$;

$$\boldsymbol{T}=\{\bigcup_{az=1}^{v} T_{az}\}=\{T_1, T_2, \ldots, T_{az}, \ldots, T_v\};\ az=\overline{1,v};$$

$v$ – кількість точок в зображенні ключової фрази (КФ); $\boldsymbol{T}$ – множина всіх точок зображення;

$$\boldsymbol{TS}=\{\bigcup_{c=1}^{ks} \boldsymbol{TS}_c\}=\{\boldsymbol{TS}_1, \boldsymbol{TS}_2, \ldots, \boldsymbol{TS}_c, \ldots, \boldsymbol{TS}_{ks}\};$$

$$\boldsymbol{TS}_c = \{\bigcup_{a=1}^{vc} TS_{c,a}\} = \{TS_{c,1}, TS_{c,2}, \ldots, TS_{c,a}, \ldots, TS_{c,vc}\};$$

відповідно множина $\boldsymbol{TS}$ приймає наступний вигляд: $\boldsymbol{TS} = \left\{\bigcup_{c=1}^{ks}\bigcup_{a=1}^{vc} TS_{c,a}\right\} = \{\{TS_{1,1}, TS_{1,2}, \ldots, TS_{1,a}, \ldots, TS_{1,vc}\}, \{TS_{2,1}, TS_{2,2}, \ldots, TS_{2,a}, \ldots, TS_{2,vc}\}, \ldots, \{TS_{c,1}, TS_{c,2}, \ldots, TS_{c,a}, \ldots, TS_{c,vc}\}, \ldots, \{TS_{ks,1}, TS_{ks,2}, \ldots, TS_{ks,a}, \ldots, TS_{ks,vc}\}\};\ c=\overline{1,ks};$ $ks$ – кількість символів в КФ; $\boldsymbol{TS}$ – множина точок зображень символів КФ; $a=\overline{1,vc};$ $vc$ – кількість точок в $c$-ому символі; $\boldsymbol{TS}_c$ – множина точок $c$-го символу; здійснюється умова, що $\boldsymbol{TS} \subset \boldsymbol{T}$;

$$\boldsymbol{KTS}_c=\{\bigcup_{g=1}^{w} KTS_{c,g}\}=\{KTS_{c,1}, KTS_{c,2}, \ldots, KTS_{c,g}, \ldots, KTS_{c,w}\};\ g=\overline{1,w};$$

$w$ – кількість КТ в $c$-ому символі; $\boldsymbol{KTS}_c$ – множина контрольних точок (КТ);

$$\boldsymbol{AT}=\{\bigcup_{i=1}^{n} AT_i\}=\{AT_1, AT_2, \ldots, AT_i, \ldots, AT_n\};$$

$i=\overline{1,n};$ $n$ – кількість ознак РП користувача; $\boldsymbol{AT}$ – множина ознак РП користувача;

$$\boldsymbol{AT1}=\{\bigcup_{c=1}^{ks}\bigcup_{a1=1}^{3\cdot vc} ATX1_{c,a1}\}=\{XT_{1,1}, YT_{1,1}, TPT_{1,1}, XT_{c,1}, YT_{c,1}, TPT_{c,1}, \ldots, XT_{c,a}, YT_{c,a}, TPT_{c,a}, \ldots, XT_{ks,vc}, YT_{ks,vc}, TPT_{ks,vc}\};$$

$\boldsymbol{AT1}$ – множина ознак РП, які використовуються в першому раунді АК; $\boldsymbol{XT}$, $\boldsymbol{YT}$, $\boldsymbol{TP}$ – вектори значень координат $X$ і $Y$ та типу точок зображення, відповідно; $\boldsymbol{XT} \subset \boldsymbol{AT1}$; $\boldsymbol{YT} \subset \boldsymbol{AT1}$; $\boldsymbol{TPT} \subset \boldsymbol{AT1}$;

$$\boldsymbol{AT2} = \{\bigcup_{c=1}^{ks}\bigcup_{a2=1}^{5\cdot vc+5} ATX2_{c,a2}\} = \{TPT_{1,1}, PT_{1,1}, UT_{1,1}, TMT_{1,1}, VT_{1,1}, TPT_{c,1}, PT_{c,1}, UT_{c,1}, TMT_{c,1}, VT_{c,1}, \ldots, TPT_{c,a}, PT_{c,a}, UT_{c,a}, TMT_{c,a}, VT_{c,a}, \ldots, TPT_{ks,vc}, PT_{ks,vc}, UT_{ks,vc}, TMT_{ks,vc}, VT_{ks,vc}, PL_c, KT_c, CH_c, KP_c\}$$; $\boldsymbol{AT2}$ – множина ознак РП, які використовуються в другому раунді АК; $\boldsymbol{PT}$ – вектор значень тиску, з яким користувач натискає ручкою (чи іншим подібним пристроєм) на сенсорний екран під час створення точок; $\boldsymbol{UT}$ – вектор значень кута зміни напрямку написання при створенні точок; $\boldsymbol{TMT}$ – вектор значень часу, який пройшов від початку написання символу до створення конкретної точки; $\boldsymbol{VT}$ – вектор значень швидкості переміщення ручки від попередньої точки в дану точку; $\boldsymbol{PL}$ – вектор значень площ зображень символів; $\boldsymbol{KT}$ – вектор значень кількостей точок символів, що аналізуються під час розпізнавання; $\boldsymbol{KU}$ – вектор значень кутів нахилу символів; $\boldsymbol{CH}$ – вектор значень частот точок символів, що зафіксовано системою; $\boldsymbol{KP}$ – вектор значень кількостей повторів точок в зображені символів (точок, що створені підряд, з однаковими обома координатами); $\boldsymbol{TPT} \subset \boldsymbol{AT2}$; $\boldsymbol{PT} \subset \boldsymbol{AT2}$; $\boldsymbol{UT} \subset \boldsymbol{AT2}$; $\boldsymbol{TMT} \subset \boldsymbol{AT2}$; $\boldsymbol{VT} \subset \boldsymbol{AT2}$;

$$\boldsymbol{O1}_c = \{\bigcup_{s1=1}^{3\cdot w} OX1_{s1}\} = \{XKT_1, YKT_1, TPKT_1, XTK_2, YTK_2, TPKT_2, \ldots, XKT_{c,g}, YKT_{c,g}, TPKT_{c,g}, \ldots, XKT_{c,w}, YKT_{c,w}, TPKT_{c,w}\}$$; $\boldsymbol{O1}_c$ – множина навчальних зразків написання $c$-го символу, з БДНЗ, який подається на ІНМ в першому раунді АК; $\boldsymbol{O1}_c \subset \boldsymbol{AT1}$;

$$\boldsymbol{O2}_c = \{\bigcup_{s2=1}^{5\cdot w+5} OX2_{s2}\} = \{TPKT_1, PKT_1, UKT_1, TMKT_1, VKT_1, TPKT_2, PKT_2, UKT_2, TMKT_2, VKT_2, \ldots, TPKT_{c,g}, PKT_{c,g}, UKT_{c,g}, TMKT_{c,g}, VKT_{c,g}, \ldots, TPKT_{c,w}, PKT_{c,w}, UKT_{c,w}, TMKT_{c,w}, VKT_{c,w}, PL_c, KT_c, KU_c, CH_c, KP_c\}$$; $\boldsymbol{O2}_c$ – множина навчальних зразків стилю написання $c$-го символу, з БДНЗ, який подається на ІНМ в другому раунді АК; $\boldsymbol{O2}_c \subset \boldsymbol{At2}$;

$$\boldsymbol{ON1} = \{\bigcup_{s1=1}^{3\cdot w} ONX1_{s1}\} = \{XKT_1, YKT_1, TPKT_1, XTK_2, YTK_2, TPKT_2, \ldots, XKT_{c,g}, YKT_{c,g}, TPKT_{c,g}, \ldots, XKT_{c,w}, YKT_{c,w}, TPKT_{c,w}\}$$; $\boldsymbol{ON1}$ – зразок РП, якій подається на ІНМ для розпізнавання в першому раунді АК; $\boldsymbol{ON1} \subset \boldsymbol{At1}$;

$$\boldsymbol{ON2} = \{\bigcup_{s2=1}^{5\cdot w+5} ONX2_{s2}\} = \{TPKT_1, PKT_1, UKT_1, TMKT_1, VKT_1, TPKT_2, PKT_2, UKT_2, TMKT_2, VKT_2, \ldots, TPKT_{c,g}, PKT_{c,g}, UKT_{c,g}, TMKT_{c,g}, VKT_{c,g}, \ldots, TPKT_{c,w}, PKT_{c,w}, UKT_{c,w}, TMKT_{c,w}, VKT_{c,w}, PL_c, KT_c, KU_c, CH_c, KP_c\}$$; $\boldsymbol{ON2}$ – зразок РП, якій подається на ІНМ для розпізнавання в другому раунді АК; $\boldsymbol{ON2} \subset \boldsymbol{At2}$;

$ED\_T$ – це параметр, який означає максимальну довжину лінії (в точках), яку необхідно вважати випадковою;

$MAXXT_1$, $MAXXT_{ks}$, $MAXXT_c$, $MAXYT_1$, $MAXYT_{ks}$, $MAXYT_c$ – максимальні значення векторів значень $\boldsymbol{XT}$ та $\boldsymbol{YT}$ 1-го, $ks$-го та $c$-го символу КФ;

$MINXT_1$, $MINXT_{ks}$, $MINXT_c$, $MINYT_1$, $MINYT_{ks}$, $MINYT_c$ – мінімальні значення векторів значень $\boldsymbol{XT}$ та $\boldsymbol{YT}$ 1-го, $ks$-го та $c$-го символу КФ;

$XMAX$ і $YMAX$ – максимально можлива кількість точок робочої області обраного розміру за осями $X$ та $Y$;

$UG$ – кут, на який необхідно повернути зображення кожного символу, для нормалізації кута нахилу їх осей координат;

$M_c$ – коефіцієнт по символьного масштабування с-го символу.

В роботі, для підвищення якості розпізнавання користувачів, запропоновано розділити процес АК на два раунди:

1. розпізнавання КФ, що написана;
2. розпізнавання стилю написання КФ.

Відповідно функція $R$ має вигляд:

$$R = \begin{cases} 1, \text{ при } (US_X \in \boldsymbol{USL}) \wedge (US_X = USL_d); \\ 0, \text{ при } ((US_X \in \boldsymbol{USL}) \wedge (US_X \neq USL_d)) \\ \qquad \vee (US_X \in \boldsymbol{USB}). \end{cases}$$

Якщо КФ є секретом, тобто імовірність її несанкціонованого отримання мінімальна, тоді для методу автентифікації достатнім є виконання тільки першого раунду розпізнавання.

В роботі розроблено метод реалізації першого раунду розпізнавання за РП та запропоновані рекомендації для реалізації другого раунду розпізнавання за РП.

В даній роботі був розроблений метод МЗ первинної обробки зразків РП користувачів ІС, необхідний для виконання АК ІС за їх РП.

**Метод ПОЗРП складається з наступних двох етапів, виконання яких необхідно на різних стадіях роботи методу АКРП**:

Етап 1. Попередній відбір даних, які будуть використовуватись для розпізнавання, або іншими словами, видалення помилкових даних із зразків РП користувачів ІС;

Етап 2. Корекція даних, які будуть використовуватись для розпізнавання зразків РП користувачів ІС.

**Етап 1. Попередній відбір даних, які будуть використовуватись для розпізнавання, або іншими словами, видалення помилкових даних із зразків РП користувачів ІС.** Для РП властива деяка нестабільність. ІНМ досить непогано справляється з цією проблемою, але якщо в зразку присутні дані, які є випадковими відхиленнями (помилками), які нічого не характеризують і будуть тільки погіршувати якість розпізнавання, тоді їх необхідно видалити з зразку; а якщо таких помилок занадто багато в одному зразку, або вони занадто грубі, тоді цей зразок треба видалити з БДНЗ (або не аналізувати його під час розпізнавати). Ці помилки викликані специфікою використання ГП для АК за РП. Проаналізувавши накопичені навчальні зразки (НЗ), в даній роботі було виділені п'ять типів помилок.

В даній роботі видалення цих помилок виконується на різних стадіях виконання АК ІС за їх РП і складається з п'яти кроків.

**Крок 1. Видалення помилок 1-го типу.** Помилка 1-го типу – це послідовність точок з нульовим тиском (крім першої подібної точки з кожної послідовності). Виникають якщо користувач провів ручкою над ГП на невеликій відстані, не доторкнувшись до нього. Ці помилки усуваються за рахунок зберігання в зразку РП лише тих пакетів з нульовим тиском, які є першими в послідовності таких пакетів. Виконується на етапі формування БДНЗ або зразка, що розпізнається.

**Крок 2. Видалення помилок 2-го типу.** Помилка 2-го типу – це випадкові точки (невелика кількість). Виникають якщо користувач випадково доторкнеться ручкою до ГП. Точки (в діапазоні від точки ($az$+1) по наступну точку з нульовим тиском) визнаються випадковими і видаляються із зразка, що аналізується, якщо виконується умова: $(PT_{az} = 0) \wedge ((PT_{az+1} = 0) \vee (PT_{az+2} = 0) \vee (PT_{az+3} = 0) \vee \ldots \vee (PT_{az+ED\_T+1} = 0))$, де $ED\_T$ налаштовується для кожної ІС і вказується в налаштуваннях системи.

**Крок 3. Видалення помилок 3-го типу.** Помилка 3-го типу – це повтори, тобто послідовність точок, що йдуть підряд, у яких значення координат по обом осям ($X$ і $Y$) не змінилися (крім випадку, коли у одній з точок нульовий тиск). Виникають якщо пакет даних передався в комп'ютер в зв'язку з зміною не координат по якоїсь з осей ($X$ і $Y$), а іншого параметру. Дані по точці ($az$ +1) повинні видалятися із зразка, що аналізується, якщо виконується наступна умова: $(XT_{az} = XT_{az+1}) \wedge (YT_{az} = YT_{az+1}) \wedge (PT_{az} \neq 0) \wedge (PT_{az+1} \neq 0)$.

**Крок 4. Видалення помилок 4-го типу.** Помилка 4-го типу – це випадкові невеликі загини (зазвичай, з гострим кутом) на початку ліній. Виникають або з вини інертності ГП, або через тремтіння руки користувача. Для визначення закінчення подібної помилки перевірявся напрям зміни координат ручки по осям $X$ і $Y$ і, якщо напрям хоча б по одній осі змінився – це означає, що помилка (загин) закінчився і далі вже йде сам символ. Відповідно, $az$ -та точка є останньою точкою загину, якщо виконується наступна умова: $(\text{sgn}(XT_{az+2} - XT_{az+1}) \neq \text{sgn}(XT_{az+1} - XT_{az})) \vee (\text{sgn}(YT_{az+2} - YT_{az+1}) \neq \text{sgn}(YT_{az+1} - YT_{az}))$. Довжину загину, який є помилкою (а не частиною символу, або відмінністю РП користувача), необхідно настроювати.

**Крок 5. Видалення помилок 5-го типу.** Помилка 5-го типу - неякісний зразок, який відкидається через неможливість розбивки зображення КФ на задану кількість зображень символів. Виникають якщо або вводиться невірна КФ, або якщо користувач має малий досвід роботи з ГП, або якщо користувач які-небудь символи написав не окремо, а разом. При наявності такої помилки, зображення КФ не можна розділити на задану кількість зображень окремих символів, тому такі зразки вилучаються на етапі розділення зображення КФ на зображення окремих символів КФ.

Деякі з цих типів помилок є ними тільки в першому раунді АК, а в другому раунді є характерними особливостями стилю письма користувача. Наприклад, помилки 4-го типу можуть мути особливостями стилю письма, а помилка 5-го типу є помилками і в першому, і в другому раунді автентифікації.

**Етап 2. Корекція даних, які будуть використовуватись для розпізнавання зразків РП користувачів ІС.** На основі проведеного аналізу накопичених НЗ, було прийнято рішення о доцільності проведення корекції даних, що аналізуються, яка необхідна для підвищення імовірності вірного

розпізнавання, а в деяких випадках для того щоб АК взагалі була можливою. Перед першим і другим раундами АК корекція різна. В першому раунді розпізнавання необхідність корекції обумовлена потребою привести зразки написання всіх символів під один шаблон. В даній роботі виконувались три різновиди корекції даних, під час яких вектори ***XT***, ***YT*** перетворюються в вектори ***XTP***, ***YTP***. Перетворення відбувається на трьох кроках.

**Крок 6.** Корекція 1-го типу – посимвольний поворот зображень символів для нормалізації кута нахилу їх осей координат. Значення *UG* обчислюється за наступною формулою:

$$UG = arctg(\frac{MAXYT_{ks} - MAXYT_1}{MAXXT_{ks} - MAXXT_1}).$$

Поворот виконується за наступними формулами:

$$SX = \frac{MAXXT_c - MINXT_c}{2};$$

$$SY = \frac{MAXYT_c - MINYT_c}{2};$$

$$XTP_{c,a} = round((XT_{c,a} - SX + MINXT_c)\cdot \cos(UG) + (YT_{c,a} - SY + MINYT_c)\cdot \sin(UG)) + round(SX + MINXT_c);$$

$$YTP_{c,a} = round((YT_{c,a} - SY + MINYT_c)\cdot \cos(UG) - (XT_{c,a} - SX + MINXT_c)\cdot \sin(UG)) + round(SY + MINYT_c).$$

**Крок 7.** Корекція 2-го типу – посимвольний зсув зображень кожного символу в центр робочої області обраного розміру. Зсув виконується за наступними формулами:

$$XTP_{c,a} = XTP_{c,a} + round(\frac{XMAX - (MAXXT_c - MINXT_c)}{2}) - MINXT_c;$$

$$YTP_{c,a} = YTP_{c,a} + round(\frac{YMAX - (MAXYT_c - MINYT_c)}{2}) - MINYT_c.$$

**Крок 8.** Корекція 3-го типу – посимвольне пропорційне масштабування (розтягування/стискання) зображень кожного символу на всю робочу область обраного розміру.

Коефіцієнт масштабування обраховувався за наступною формулою:

$$M_c = \min(\min(\frac{0 - (SX + MINXT_c)}{MINXT_c - (SX + MINXT_c)}, \frac{XMAX - (SX + MINXT_c)}{MAXXT_c - (SX + MINXT_c)}), \min(\frac{0 - (SY + MINYT_c)}{MINYT_c - (SY + MINYT_c)}, \frac{YMAX - (SY + MINYT_c)}{MAXYT_c - (SY + MINYT_c)})).$$

Масштабування виконується за наступними формулами:

$$XTP_{c,a} = round((XTP_{c,a} - (SX + MINXT_c))\cdot M_c + (SX + MINXT_c));$$

$$YTP_{c,a} = round((YTP_{c,a} - (SY + MINYT_c))\cdot M_c + (SY + MINYT_c)).$$

Якщо в першому раунді АК ці функції не виконувати, тоді буде виконуватися розпізнавання місця розміщення символів, їх розміру і куту нахилу, а не самих символів.

В другому раунді АК, з трьох типів корекції можлива, але не обов'язкова, тільки корекція 2-го типу.

**Розроблений метод АКРП складається з десяти етапів.**

**Етап 1. Попереднє формування множини ознак РП користувача ІС.** Для АК ІС аналізується множина ознак його РП $AT=\{\bigcup_{i=1}^{n} At_i\}$, яка описана в моделі даних. Ця множина складається з векторів ознак, значення яких передаються з ГП – це ***XT***, ***YT***, ***PT*** та з векторів ознак, які обраховуються на наступних етапах даного методу і є похідними від тих, що передані з ГП – це ***TPT***, ***UT***, ***TMT***, ***VT***, ***PL***, ***KT***, ***KU***, ***CH***, ***KP***. В кожному з перерахованих раундів АК, множина ознак ***AT***, яка використовується для розпізнавання зразків, різна. В першому раунді АК, для розпізнавання КФ в роботі використовується множина ***AT1***; $AT1 \subset AT$. В другому раунді, для розпізнавання стилю написання КФ, використовується множина ***AT2***, в яку входять вектори ознак не тільки точок зображення, а й параметри загальні для усього символу; $AT2 \subset AT$. В залежності від співвідношення надійності системи, що вимагається, до припустимого об'єму ресурсів, що задіяні, можна використовувати всі перераховані ознаки, або деякі з них.

**Етап 2. Налаштування параметрів, які є найбільш критичними при АК ІС за їх РП.** Налаштування найбільш критичних параметрів має дуже велике значення для збільшення імовірності вірного розпізнавання користувачів ІС за їх РП. Визначення переліку цих параметрів здійснювалося за допомогою аналізу результатів проведених експериментів. Цими параметрами є: необхідна кількість символів КФ; мінімальна необхідна кількість навчальних зразків в БДНЗ для кожного користувача; помилки яких типів необхідно видаляти і при яких умовах (довжина послідовності точок, які повинні вважатися випадковими та довжина лінії до її загину, яка повинна вважатися помилкою); які типи корекції даних проводити для другого раунду АК (для першого корекція всіх трьох типів обов'язкова); з яких типів точок формувати множину КТ; довжина частини лінії, на якій треба обрати тільки одну КТ 2-го типу.

**Етап 3. Реєстрація користувачів ІС або формування БДНЗ користувачів ІС.** В залежності від заданого режиму роботи, отриманий зразок РП користувача ІС, буде або зареєстрований в БДНЗ, або відправлений на автентифікацію. Якщо в БДНЗ недостатньо НЗ для користувача, що розпізнається, тоді його не допускають до етапу автентифікації, а направляють на етап формування БДНЗ. Однією з особливостей принципу роботи ГП, яка враховувалася при розробці зазначеної системи АК, є той факт, що пакет даних з параметрами точки передається в комп'ютер, якщо змінилось значення хоча б одного параметру. Внаслідок цього в систему передаються зайві пакети даних, які необхідно видалити. Для видалення помилок одного з типів на даному етапі виконується Крок 1, Етапу 1, методу ПОЗРП. Крім параметрів точок зображення КФ, в даній роботі, в БДНЗ фіксуються "логін" користувача, дата та час створення точки. До користувачів висувається вимога, що символи КФ повинні бути написані не разом, а окремо один від одного.

Для отримання даних по точкам зображення з ГП, враховувались характеристики ГП, що використовувався для динамічної передачі зображення в комп'ютер; використовувався інтерфейс для роботи з ним (Wintab) та єдина модель біометричних технологій розпізнавання (BioAPI).

**Етап 4. Видалення помилок 2, 3, 4 типів.** Для виконання цієї задачі виконуються Крок 2-Крок 4, Етапу 1, методу ПОЗРП.

**Етап 5. Умовне розділення зображення КФ на зображення окремих символів.** Доцільність цього етапу аргументується наступними причинами:

1. Легше накопичити БДНЗ для $N$ символів, ніж для $\sum_{ks=1}^{mks} ks^N$ можливих КФ (або $ks^N$, якщо $ks$ відоме), де $N$ – кількість елементів множини можливих символів КФ; $ks$ – довжина КФ; $mks$ – максимально можлива довжина КФ.

2. Об'єкт, що класифікується, легше перевірити чи належність він до одного з $N$ класів, ніж до одного з $\sum_{ks=1}^{mks} ks^N$ класів.

3. Значно зменшується об'єм ресурсів, що використовуються під час розпізнавання, за рахунок зменшення довжини зразка РП. Тобто, враховуючи те, що зображення одного символу в середньому складається з 100 точок, то якщо розділення не виконувати, тоді зразок РП буде складатися приблизно з $300 \cdot ks$ характеристик в першому раунді автентифікації та з $505 \cdot ks$ характеристик в другому раунді, а якщо розділення виконувати, тоді відповідно приблизно з 300 та 505 характеристик.

Розділення зображення КФ на зображення окремих символів виконувалось на основі аналізу значення тиску ($PT_{az}$) в пакетах даних о точках, що передаються з ГП, а саме – точка, в якої $PT_{az}$ =0, використовувалась, як розмежував між зображеннями символів. Значення $ks$ задається при налаштуванні системи АК ІС за їх РП.

Якщо після розділення, кількість отриманих зображень виявляється менша, ніж $ks$, тоді це помилка 5 типу і для її видалення виконується Крок 5, Етапу 1, методу ПОЗРП.

Якщо після розділення, кількість отриманих зображень виявляється більша, ніж $ks$, тобто деякі символи складаються з декількох частин, тоді ці частини з'єднувались, за рахунок знаходження сусідніх зображень, між границями яких мінімальна відстань.

Після розділення зображення КФ на зображення окремих символів КФ, з множини $T$ буде сформована множина $TS$.

**Етап 6. Корекція даних, які будуть використовуватись для розпізнавання зразків РП користувачів ІС.** Для виконання цієї задачі виконуються Етап 2, методу ПОЗРП.

**Етап 7. Формування множини контрольних точок.** Необхідність цього етапу пояснюється тим, що $c_v \approx 100$, відповідно множина $AT1$ буде складатися з $3 \cdot ks \cdot c_v$ елементів (в нашому випадку ≈1800), а множина $AT2$ відповідно з $5 \cdot ks \cdot c_v$ елементів (в нашому випадку ≈3000), у зв'язку з чим для виконання розпізнавання, витрачаються занадто великі ресурси. В зв'язку з цим в даній роботі аналізуються характеристики не всіх точок, а тільки найбільш значимих для кожного символу – контрольних точок (КТ). При цьому для кожного символу, з множини $TS_c$ виділяється множина $KTS_c$, яка і буде використовуватися для розпізнавання. Відповідно $XTP \rightarrow XKT$, $YTP \rightarrow YKT$, $TPT \rightarrow TPKT$, $PT \rightarrow PKT$, $UT \rightarrow UKT$, $TMT \rightarrow TMKT$, $VT \rightarrow VKT$. Від правильності розстановки КТ значно залежить імовірність вірного розпізнавання об'єктів. В даній роботі виділяються наступні 3 типи КТ:

1. Початкові та кінцеві точки кожної лінії – це точки торкання ручкою ГП та точки відриву ручки від ГП. На рис. 1 ці точки показані червоним кольором (точки 1 та 15). Для визначення таких точок використовуються збережені пакети даних с нульовим тиском ($PT_{c,a}$=0). Умова того, що $a$-та точка є початковою або кінцевою КТ:

$$((PT_{c,a} \neq 0) \wedge ((a = 1) \vee (a = vc) \vee (PT_{c,a-1} = 0) \vee (PT_{c,a+1} = 0))).$$

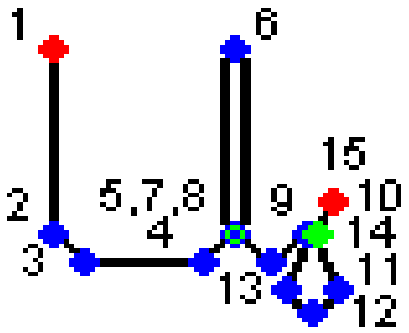


Рис. 1. Приклад розстановки КТ

2. Кутові точки ліній – це точки, які знаходяться на вигині лінії. На рис. 1 ці точки показані синім кольором (точки 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13). Вигином лінії будемо називати зміну напряму лінії, яку можна визначити за зміною знака зміни координат по одній з осей (або по обом). Є декілька варіантів розміщення точок, при яких утворюються вигини ліній. Умова того, що $a$-та точка є кутовою КТ:

$$((a \neq 1) \wedge (a \neq vc) \wedge (PT_{c,a} \neq 0) \wedge (PT_{c,a-1} \neq 0) \wedge (PT_{c,a+1} \neq 0) \wedge (c_a = c_{a-1} = c_{a+1}) \wedge ((\text{sgn}(XT_{c,a} - XT_{c,a-1}) \neq \text{sgn}(XT_{c,a+1} - XT_{c,a})) \vee (\text{sgn}(YT_{c,a} - YT_{c,a-1}) \neq \text{sgn}(YT_{c,a+1} - YT_{c,a}))));$$

де sgn – математична функція, яка визначає знак виразу. При визначенні кутових КТ, за вказаним критерієм, виникає певна кількість хибних кутових КТ. Для усунення від процесу розпізнавання таких точок, в розробленій системі АК ІС за їх РП створена функція відбору найбільш значимих кутових КТ. На кожному відрізку лінії довжиною $D$ (підбирається) визначається тільки одна, найбільш значима, кутова КТ, а інші відкидаються. При відборі аналізується траєкторія руху ручки по ГП під час переміщення між двома сусідніми кутовими точками зображення символу. Якщо виконується наступна умова:

$$\sum_{cd=cdn}^{cdk} \sqrt{(XTP_{c,cd} - XTP_{c,cd+1})^2 + (YTP_{c,cd} - YTP_{c,cd+1})^2} < D,$$

де $cdn$ і $cdk$ – номера, під якими зберігаються в множині $TS_c$, відповідно дві точки, які згодом визначились як сусідні кутові КТ), тоді одну з двох точок необхідно видалити з множини КТ. В якості критерію відбору того, яку з двох кутових КТ треба залишити, в даній роботі використовується кут вигину лінії ($KUT$) в точці, що аналізується, – точка, в якій $KUT$ менше, але не дорівнює нулю, повинна залишатися.

3. Точки перетинання ліній – це точки, які знаходяться на перетинанні ліній. На рис. 1 ці точки показані зеленим кольором (точки 8 та 14). Пошук ускладнює той факт, що точки зображення знаходяться не деякій відстані друг від друга (відстань між точками, зазвичай, більше 1 пікселю), тому в даній роботі були виділені три типи КТ перетинання. В першому випадку умовою того, що $a$-та точка є точкою перетинання ліній, є:

$$(TS_{c,j} \in \boldsymbol{TS}_c) \wedge (TS_{c,j} = \overline{TS_{c,1}, TS_{c,a-1}}) \wedge (PT_{c,a} \neq 0) \wedge (XTP_{c,j} = XTP_{c,a}) \wedge (YTP_{c,j} = YTP_{c,a}) \wedge (PT_{c,j} \neq 0).$$

В другому випадку умовою того, що $a$-та точка є точкою перетинання ліній, є:

$$(TS_{c,j} \in \boldsymbol{TS}_c) \wedge (TS_{c,j+1} \in \boldsymbol{TS}_c) \wedge (TS_{c,j} = \overline{TS_{c,1}, TS_{c,vc-1}}) \wedge (PT_{c,a} \neq 0) \wedge (TS_{c,a} \in [TS_{c,j} TS_{c,j+1}]) \wedge (PT_{c,j} \neq 0) \wedge (PT_{c,j+1} \neq 0) \wedge (TS_{c,a} \neq TS_{c,j}) \wedge (TS_{c,a} \neq TS_{c,j+1}).$$

В третьому випадку умовою того, що $p$-та точка є точкою перетинання ліній, є:

$$(TS_{c,a} \in \boldsymbol{TS}_c) \wedge (TS_{c,a+1} \in \boldsymbol{TS}_c) \wedge (TS_{c,j} \in \boldsymbol{TS}_c) \wedge$$
$$(TS_{c,j+1} \in \boldsymbol{TS}_c) \wedge (TS_{c,j} = \overline{TS_{c,2}, TS_{c,a-2}}) \wedge$$
$$(TS_{c,p} \notin \boldsymbol{TS}_c) \ \wedge (TS_{c,p} \in [TS_{c,a} TS_{c,a-1}])$$
$$\wedge (TS_{c,p} \in [TS_{c,j} TS_{c,j-1}]) \wedge (PT_{c,a} \neq 0) \wedge$$
$$(PT_{c,a-1} \neq 0) \wedge (PT_{c,j} \neq 0) \wedge (PT_{c,j-1} \neq 0) \wedge$$
$$(c_j = c_{j-1} = c_a = c_{a-1}).$$

Для зменшення ресурсів, які використовувалися, не погіршив при цьому якість розпізнавання, в розробленому методі АК ІС за їх РП, забезпечувалось дотримання того, щоб одна і та сама точка не зберігалась більше ніж один раз в множині КТ. При правильному розміщенні КТ в кожному символі їх кількість буде приблизно однаковою, що є необхідною умовою для подальшого розпізнавання.

**Етап 8. Виконання першого раунду АК ІС за їх РП, а саме розпізнавання КФ, що написана.**

Процес розпізнавання в першому та другому раундах АК за їх РП різні. Задачу АК, тобто розпізнавання образів, в даному випадку, можна звести до задачі класифікації образів. В якості механізму класифікації образів, в даному випадку, була обрана ІНМ [10 ].

Як вже було сказано раніше, в першому раунді АК розпізнається кожний символ окремо і, відповідно, зразок написання кожного символу складається з множини ознак $\boldsymbol{AT1}$ для кожного символу з множини $\boldsymbol{KTS}_c$ Відповідно для розпізнавання на ІНМ подається зразок $\boldsymbol{ON1}$. Архітектура ІНМ обумовлена тим, що ця мережа повинна розпізнавати кожний написаний на ГП символ. Обрана мережа складається з наступних чотирьох шарів: вхідний шар, шар зразків, шар підсумовування, вихідний шар. Кількісно архітектуру мережі, виходячи з задачі, яка розв'язується, можна визначити наступним чином:

1. Число вхідних елементів дорівнює числу ознак. В якості ознак, в даному випадку, використовуються по три параметра кожної зафіксованої КТ, відповідно, кількість вхідних елементів дорівнює $3 \cdot w$.

2. Число елементів шару зразків дорівнює числу НЗ. Кількість НЗ дорівнює сумі кількостей НЗ написання всіх символів (для яких є зразки написання в БДНЗ) з $w$-ою кількістю КТ (тому що для розпізнавання використовуються тільки ті НЗ написання символів, в яких зафіксована така ж кількість КТ, як і в зразку, що розпізнається). Відповідно, кількість елементів шару зразків дорівнює $\sum_{cu=1}^{ks} kol\_u_{cu}$, де $kol\_u$ – масив, елементи якого дорівнюють кількостям НЗ для кожного з $ks$ символів.

3. Число елементів шару підсумовування дорівнює числу класів. При розпізнавання написаних на ГП символів класом є символ з $w$-ою кількістю КТ, для яких є НЗ в БДНЗ. Відповідно, кількість класів дорівнює $ks$.

4. Вихідний шар завжди складається з одного елементу. Він визначає клас (в даному випадку символ), до якого з найбільшою імовірністю належить невідомий зразок.

Активність елементу шару зразків обраховується за наступною формулою:

$$AK1_{nz} = \exp\left(\frac{\sum_{s1=1}^{3 \cdot w} ONX1_{s1} \cdot OX1_{s1,nz} - 1}{\sigma^2}\right),$$

де $\sigma$ – ширина функції активності.

Для використання цієї формули, вектор вхідних даних повинен бути нормалізовано, тобто всі ознаки з множин $\boldsymbol{ON1}$ та $\boldsymbol{O1}$ необхідно нормалізувати за наступною формулою:

$$OZN_{s1} = \frac{OZ_{s1}}{\sqrt{\sum_{s1=1}^{3 \cdot w} OZ_{s1}^2}},$$

де $OZN_{s1}$ – нормалізована ознака, $OZ_{s1}$ – ненормалізована ознака.

Якщо всі символи КФ розпізнані правильно, тоді приймається рушення, що КФ написана правильна. Як аргументовано далі, в залежності від необхідного рівня безпеки, рішення про успішне розпізнавання можна приймати не обов'язково при вірному розпізнаванні всіх символів КФ, а наприклад, при розпізнаванні 5 з 6 символів. В даній роботі, у випадку коли правильно розпізнана вся КФ, приймалось рішення про успішне проходження АК ІС за їх РП.

**Етап 9. Первинна обробка зразків РП користувачів ІС, за допомогою розробленого методу, яка необхідна для виконання другого раунду АК ІС за їх РП.** Необхідні на цьому етапі дії описані в розробленому методі ПОЗРП.

**Етап 10. Виконання другого раунду АК ІС за їх РП, а саме розпізнавання стилю написання КФ.**

В другому раунді АК, в залежності від вимог до системи, в якості параметрів, що аналізуються, можна використовувати всі, або деякі ознаки множини ознак *AT2* (наведених раніше параметрів КТ і параметрів окремих символів). Тобто для розпізнавання на ІНМ подається зразок *ON2*. Кількісно архітектура мережі в другому раунді АК відрізняється від мережі в першому раунді і виглядає наступним чином:

1. Число вхідних елементів дорівнює кількості ознак і відповідно дорівнює $5 \cdot w+5$.
2. Число елементів шару зразків дорівнює числу НЗ стилю написання КФ для $l$ користувачів ІС.
3. Число елементів шару підсумовування дорівнює числу класів і відповідно дорівнює $l$.
4. Вихідний шар завжди складається з одного елементу. Він визначає клас (в даному випадку користувача), якому з найбільшою імовірністю належить невідомий зразок.

Активність елементу шару зразків обраховується за наступною формулою:

$$AK2_{nz} = \exp\left(\frac{\sum_{s1=1}^{5 \cdot w+5} ONX2_{s1} \cdot OX2_{s1,nz} - 1}{\sigma^2}\right).$$

Для використання цієї формули, так як і в першому раунді розпізнавання, необхідно виконати нормалізацію всіх ознак з множин *ON2* та *O2*.

**Експериментальна частина**

Для визначення ефективності використання розроблених в даній роботі методів, було проведено ряд експериментів. Для цього на всі описані в роботі дії були створені алгоритми, на базі яких потім розроблено програмне забезпечення АК ІС за їх РП. За допомогою розробленого програмного забезпечення була накопичена необхідна БДНЗ, зразки з якої потім і використовувалися під час проведення експериментів. Експерименти проводились тільки по реалізації першого раунду АК ІС за їх РП, тобто по розпізнаванню КФ, що написана.

На імовірність правильного розпізнавання зразків, значний вплив має правильність вибору КТ зображення, чиї характеристики надалі використовуються для розпізнавання. А при використанні для розпізнавання ІНМ (як в даній роботі), обов'язковою вимогою є те, що у всіх зразках, що задіяні в розпізнаванні (і навчальних, і тих, що розпізнаються), повинна бути однакова кількість ознак, що аналізуються, і відповідно однакова кількість КТ. Але за результатами експериментів можна зробити висновок, що не тільки в різних буквах виявляється різна кількість КТ, але й в одній і тій же точці, при різному написанні, виявляється різна кількість КТ. Тому в даній роботі, для кожного символу НЗ розділяються на групи, в залежності від кількості КТ, а під час розпізнавання з БДНЗ відбираються тільки ті зразки, в яких така сама кількість КТ, як і зразку, що розпізнається. Враховуючи все сказане, в роботі було проведено аналіз кількості зафіксованих КТ в різних символах (рис. 2). За результатами аналізу можна сказати, що при використанні правильного алгоритму формування множини КТ, розкид значень кількості КТ, які часто повторюються, досить малий. Тому для економії ресурсів, що використовуються, є сенс відкидати ті зразки РП, кількість КТ, в яких не попадає в діапазон характерних значень. Цей діапазон треба підбирати.

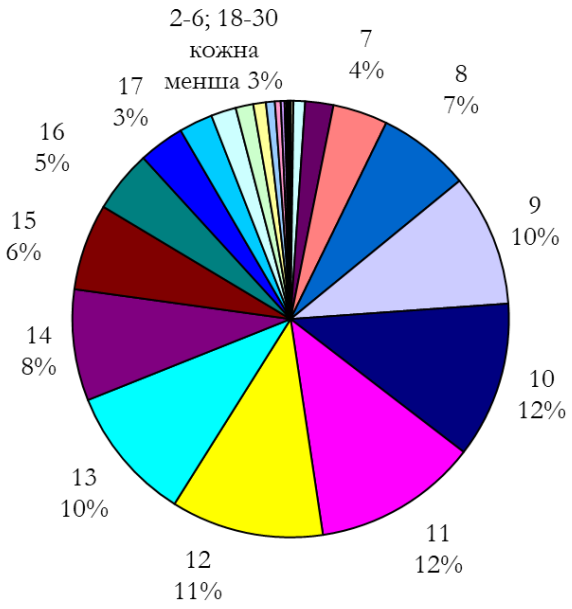


Рис. 2. Розкид значень кількості КТ в одному символі

Експерименти для визначення імовірності правильного розпізнавання проводилися при наступних умовах: в системі 7 користувачів; для кожного з них в БДНЗ накопичено приблизно по 1000 НЗ; в КФ 6 символів; видаляються помилки 2-го типу (лінія з п'яти та менше точок), помилки 3-го типу, помилки 4-го типу (загини з п'яти та менше точок); виконуються всі три типи корекції даних, використовуються КТ трьох типів, але з КТ 2-го типу обирались тільки найбільш значущі на відрізках з 200 точок. Результати експериментів

показані на рис. 3. Невисока імовірність розпізнавання у деяких користувачів обумовлюється згладжуванням ними характерних особливостей деяких букв, тобто ці люди різні букви пишуть практично однаково. Крім того, експерименти показали, що деякі символи розпізнаються набагато гірших за інших. Це обумовлюється схожістю деяких букв. В залежності від рівня безпеки, що вимагається, можна приймати рішення про успішне розпізнавання у випадку, якщо вірно розпізнанні не всі символи, а наприклад 5 з 6. Це збільшить імовірність успішного проходження процедури АК ІС за їх РП. Також причиною не дуже високої імовірності правильного розпізнавання, є те що при розпізнаванні кожного символу, використовуються не всі зразки його написання, а тільки ті, в яких кількість КТ така ж, як і у зразка, що розпізнається, тобто збільшення БДНЗ збільшить імовірність правильного розпізнавання.

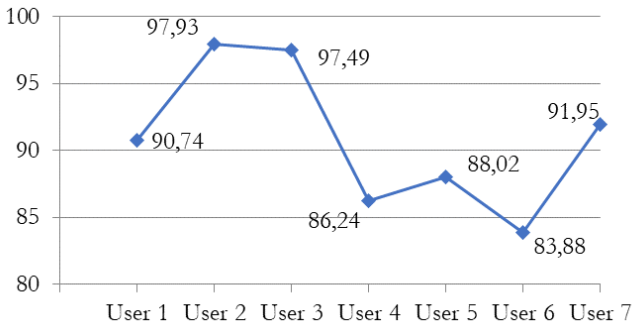


Рис. 3. Імовірність правильного розпізнавання користувачів ІС за їх РП

**Висновки**

Наукова новизна даної роботи полягає в наступному:

1. Запропоновано метод первинної обробки зразків рукописного почерку, в якому за рахунок видаленню помилкових даних п'яти типів та проведенню корекції даних трьох типів, досягається збільшення імовірності правильної розпізнавання користувачів ІС та, завдяки зменшенню кількості ознак в зразках, що аналізуються, зменшення затрачуваних ресурсів.

2. Удосконалено метод автентифікації користувачів ІС за їх рукописним почерком, який за рахунок автоматизації процесу відбору контрольних точок в зразках РП, чиї характеристики аналізуються; виконання первинної обробки зразків РП; використання для розпізнавання імовірнісної нейронної мережі, збільшує імовірність вірного розпізнавання користувачів ІС.

Практична цінність даної роботи полягає в наступному:

1. На основі запропонованого методу розпізнавання користувачів за їх рукописним почерком, з використанням методу обробки навчальних даних, створено програмне забезпечення для реалізації біометричної автентифікації користувачів ІС за їх рукописним почерком, що дозволяє збільшити ступень багатофакторності автентифікації ІС при наявності стандартних сенсорних засобів вводу графічної інформації та збільшити імовірність вірного розпізнавання користувачів ІС.

2. На основі результатів проведених експериментів, за допомогою розробленого програмного забезпечення, здійснено вибір конфігураційних параметрів, налаштування яких є найбільш критичними для збільшення імовірності вірного розпізнавання користувачів ІС та отримана оцінка імовірності вірного розпізнавання користувачів ІС за обраними біометричними характеристиками, яку забезпечує використання імовірнісної нейронної мережі.

Незважаючи на те, що запропоновані в даній роботі методи, дозволяють досягти досить високу імовірність правильного розпізнавання користувачів ІС, пошук більш ефективних механізмів розпізнавання та інших параметрів, які суттєво впливають на імовірність правильного розпізнавання, залишається доволі актуальною задачею.

**ЛІТЕРАТУРА**

[1]. Н. Кошева, Н. Мазниченко, "Ідентифікація користувачів інформаційно-комп'ютерних систем: аналіз і прогнозування підходів", *Системи обробки інформації*, № 6(113), С. 215-223, 2013.

[2]. O. Vysotska, A. Davydenko, "Keystroke Pattern Authentication of Computer Systems Users as One of the Steps of Multifactor Authentication", *In: Hu Z., Petoukhov S., Dychka I., He M. (eds) Advances in Computer Science for Engineering and Education II. ICCSEEA 2019. Advances in Intelligent Systems and Computing*, vol. 938, pp. 356-368, Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-16621-2_33.

[3]. S. Kazmirchuk, A. Ilyenko, S. Ilyenko, "Digital Signature Authentication Scheme with Message Recovery Based on the Use of Elliptic Curves", *In: Hu Z., Petoukhov S., Dychka I., He M. (eds) Advances in Computer Science for Engineering and Education II. ICCSEEA 2019. Advances in Intelligent Systems and Computing*, vol. 938, pp. 279-288, Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-16621-2_26.

[4]. А. Ільєнко, С. Ільєнко, "Програмний модуль з використанням процедури формування та верифікації електронного цифрового підпису", *Наукоємні технології*, Т. 39, № 3, С. 345-354, 2018.

[5]. О. Висоцька, "Моніторинг роботи користувачів комп'ютерних систем за допомогою технологій розпізнавання за клавіатурним почерком", *Моделювання та інформаційні технології. Збірник наукових праць інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України*, вип. 84, С. 119-125, 2018.

[6]. O. Vysotska, A. Davydenko, "Authentication of information systems users, based on the analysis of their handwriting", *Scientific and Practical Cyber Security Journal (SPCSJ)*, №2(4). pp. 51-63, 2018.

[7]. O. Vysotska, A. Davydenko, "The usage of handwriting recognition systems of information systems users for their authentication", *La science et la technologie à l'ère de la société de l'information: coll. de papiers scientifiques «ΛΌГОΣ» avec des matériaux de la conf. scientifique et pratique internationale, Bordeaux, 3 mars, 2019. Bordeaux : OP «Plateforme scientifique européenne»*, vol. 9, pp. 48-51, 2019.

[8]. T. Furukawa, "The New Method of Identification of Handwriting Using Volumes of Indentations", *2012 International Conference on Frontiers in Handwriting Recognition*, pp. 163-168. DOI: 10.1109/ICFHR. 2012.281.

[9]. A. Gattal, Y. Chibani, "Segmentation and Recognition Strategy of Handwritten Connected Digits Based on the Oriented Sliding Window", *2012 International Conference on Frontiers in Handwriting Recognition*, pp. 297-301. DOI: 10.1109/ICFHR.2012.265.

[10]. Р. Каллан, *Основные концепции нейронных сетей, Пер. с англ.,* М.: Издательский дом "Вильямс", 2001, 290 с.

## МЕТОД АУТЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ ИНФОРМАЦИОННЫХ СИСТЕМ ПО ИХ РУКОПИСНОМУ ПОЧЕРКУ С МНОГОШАГОВОЙ КОРРЕКЦИЕЙ ПЕРВИЧНЫХ ДАННЫХ

Статья посвящена биометрической аутентификации пользователей, а именно аутентификации по рукописному почерку. В данной работе аргументирована актуальность создания системы биометрической аутентификации пользователей информационных систем по их рукописному почерку. После чего определено множество характеристик рукописного почерка для дальнейшего их использования для аутентификации. На основе проведенного анализа выбранных характеристик, определена их пригодность для дальнейшего использования во время распознавания пользователей. Разработан метод аутентификации пользователей информационных систем по их рукописному почерку и метод необходимой первичной обработки образцов рукописного почерка пользователей информационных систем. Необходимость первичной обработки вызвана спецификой использования, для динамической передачи изображения в компьютер, графического планшета (или другого устройства с сенсорным экраном). Эта обработка заключается в удалении ошибочных данных и в коррекции данных, которые будут использоваться для распознавания. В работе выделены пять типов ошибок и три типа коррекции данных. Для улучшения процесса распознавания, изображение написанной ключевой фразы, для дальнейшего использования, разделяется на изображения отдельных символов. Соответственно, к пользователям выдвигается условие, что символы вводимой ключевой фразы, должны быть написаны отдельно друг от друга. Во время распознавания, анализируются параметры не всех точек изображения, а только наиболее значимых, контрольных точек. В работе выделены три типа контрольных точек и аргументирована значимость использования для этого наиболее оптимального алгоритма. Механизмом распознавания было выбрано один из видов нейронных сетей, а именно вероятностную нейронную сеть. На основе предложенных методов разработано программное обеспечение, используя которое, сначала была сформирована база данных учебных образцов рукописного почерка пользователей информационных систем. Затем, было проведено ряд экспериментов для определения эффективности применения разработанных методов и для выявления наиболее значимых, для правильного распознавания, настроек системы аутентификации пользователей. В завершение был сделан вывод, что, несмотря на то, что предложенные в данной работе методы, позволяют достичь достаточно высокую вероятность правильного распознавания пользователей информационных систем, поиск более эффективных механизмов распознавания и других параметров, которые существенно влияют на вероятность правильного распознавания, остается весьма актуальной задачей.

**Ключевые слова:** аутентификация, распознавание, биометрия, рукописный почерк, информационные системы.

## AUTHENTICATION METHOD OF INFORMATION SYSTEMS USER BY THEIR HANDWRITING WITH MULTISTAGE CORRECTION OF PRIMARY DATA

The article is devoted to the biometric authentication of users, namely authentication by handwriting. In this paper, the relevance of creating a biometric authentication system of information systems users by their handwriting was argued. After that, there were determined a lot of characteristics of handwriting for their further use for authentication. Based on the analysis of the selected characteristics, there was determined their suitability for further use during user recognition. The method of authentication of information systems users by their handwriting and the method of the necessary primary processing of handwriting samples of information systems users were developed. Primary processing need is caused by the specific use, for the dynamic transfer of images to a computer, a graphic tablet (or other touch screen device). This processing is to remove erroneous data and correction of data to be used for recognition. There are

five types of errors and three types of data correction in the work. To improve the recognition process, the image of the written key phrase, for further use, is divided into images of individual characters. Accordingly, the user is forced by the condition that the characters of the key phrase entered should be written separately from each other. During the recognition, the parameters of not all points of the image, but only the most significant, control points are analyzed. There are three types of control points in the work and the reasoned significance of using the most optimal algorithm for this. One of the types of neural networks, namely the probabilistic neural network, was chosen as the recognition mechanism. On the basis of the proposed methods, the software was developed, using which, first, a database of training samples handwriting of information systems was formed. Then, a series of experiments was conducted to determine the effectiveness of the application of the developed methods and to identify the most significant, for proper recognition, user authentication system settings. In the end, it was concluded that, despite the fact that the methods proposed in this paper allow to achieve a sufficiently high probability of correct recognition of users of information systems, the search for more efficient recognition mechanisms and other parameters that significantly affect the probability of correct recognition remains very actual objective.

**Keywords:** authentication, recognition, biometrics, handwriting, information systems.

**Корченко Олександр Григорович**, доктор технічних наук, професор, лауреат Державної премії України в галузі науки і техніки, завідувач кафедри безпеки інформаційних технологій Національного авіаційного університету, візит-професор Університету в Бельсько-Бялій (Гуманітарно-технічна академія в Бельсько-Бялій, м. Бельсько-Бяла, Польща), провідний науковий співробітник Національної академії СБ України.
E-mail: icaocentre@nau.edu.ua
Orcid ID: 0000-0003-3376-0631.
**Корченко Александр Григорьевич**, доктор технических наук, профессор, лауреат Государственной премии Украины в области науки и техники, заведующий кафедрой безопасности информационных технологий Национального авиационного университета, визит-профессор Университета в Бельско-Бялой (Гуманитарно-техническая академия в Бельско-Бялой, г. Бельско-Бяла, Польша), ведущий научный сотрудник Национальной академии СБ Украины.
**Korchenko Oleksandr**, Dr Eng (Information security), professor, laureate of the State Prize of Ukraine in Science and Technology, Head of IT-Security Academic Department, National Aviation University, Visit-Professor at The University of Bielsko-Biala (Akademia Techniczno-Humanistyczna, Bielsko-Biała, Poland), Leading Researcher of the National Academy of SS of Ukraine.

**Давиденко Анатолій Миколайович**, кандидат технічних наук, старший науковий співробітник, провідний науковий співробітник відділу Теорії моделювання Інституту проблем моделювання в енергетиці ім. Г.Є. Пухова НАН України.
E-mail: davydenko@ipme.kiev.ua.
Orcid ID: 0000-0001-6466-1690.
**Давиденко Анатолий Николаевич**, кандидат технических наук, старший научный сотрудник, ведущий научный сотрудник отдела Теории моделирования Института проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины.
**Davydenko Anatoly**, Candidate of Technical Sciences, Senior Researcher, Leading Researcher of Department of Modelling Theory, Pukhov Institute for Modelling in Energy Engineering of NAS of Ukraine.

**Висоцька Олена Олександрівна**, асистент кафедри комп'ютеризованих систем захисту інформації Національного авіаційного університету.
E-mail: Lek_Vys@ukr.net.
Orcid ID: 0000-0002-9543-1385.
**Высоцкая Елена Александровна**, ассистент кафедры компьютеризированных систем защиты информации, Национального авиационного университета.
**Vysotska Olena**, teacher of the department of computerized information security systems of the National Aviation University.