

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інформаційних радіотехнологій і технічного захисту інформації

Кафедра Комп'ютерної інженерії та систем технічного захисту інформації

АТЕСТАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження інформативних параметрів динаміки системи
"користувач-миша" для задач ідентифікації за курсорним почерком

Виконав:
студент 2 курсу, групи СТЗІАм-19-1
Зозуля О.С.

Спеціальність 125 «Кібербезпека»
Тип програми освітньо-професійна
«Системи технічного захисту
Освітня програма інформації, автоматизація
її обробки»

Керівник доц. Горелов Д.Ю.

Допускається до захисту
Зав. кафедри

(підпис)

проф. Антіпов І.Є.

2020 р.

Харківський національний університет радіоелектроніки

Факультет	<i>Інформаційних радіотехнологій і технічного захисту інформації</i>
Кафедра	<i>Комп'ютерної інженерії та систем технічного захисту інформації</i>
Рівень вищої освіти	<i>другий (магістерський)</i>
Спеціальність	<i>125 «Кібербезпека»</i>
Тип програми	<i>освітньо-професійна</i>
Освітня програма	<i>«Системи технічного захисту інформації, автоматизація її обробки»</i>

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

«___» _____ 20 ____ р.

ЗАВДАННЯ НА АТЕСТАЦІЙНУ РОБОТУ

студентові Зозулі Олександр Сергійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження інформативних параметрів динаміки системи "користувач-миша" для задач ідентифікації за курсорним почерком

затверджена наказом по університету від «___» _____ 2020 р. № _____

2. Термін подання студентом роботи до екзаменаційної комісії 10 грудня 2020 р.

3. Вихідні дані до роботи Дослідити інформативність ознак курсорного почерку для задач біометричної ідентифікації користувачів комп'ютерних мереж.

Для досягнення поставленої мети необхідно розв'язати наступні задачі:

1) провести аналітичний огляд сучасних рішень і методів статичної та динамічної аутентифікації;

2) провести пошук відкритих баз даних параметрів курсорного почерку та обрати одну з них для подальших досліджень;

3) на основі обраної бази даних експериментально дослідити точність аутентифікації, як функцію використаних інформативних ознак курсорного почерку.

4. Перелік питань, що потрібно опрацювати в роботі

1. Технології статичної біометричної аутентифікації

2. Принципи стабільної динамічної біометрії

3. Технології динамічної біометричної аутентифікації

4. Пошук відповідної до цілей досліджень бази даних параметрів курсорного почерку

5. Розробка схеми експерименту та проведення досліджень

6. Висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів)

1. Методи ідентифікації користувачів комп'ютерних мереж

за інформаційним почерком. А4. Ел.ф.

2. Датасети динаміки системи «користувач-миша»,

що знаходяться у вільному доступі. А4. Ел.ф.

3. Інформативні характеристики динаміки системи

«користувач-миша». А4. Ел.ф.

4. Схема експерименту у Orange. А4. Ел.ф.

5. Результати проведених досліджень. А4. Ел.ф.

6. Висновки. А4. Ел.ф.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Огляд технологій статичної біометричної аутентифікації	02.09.20 – 20.09.20	
2	Аналіз принципів стабільної динамічної біометричної аутентифікації	21.09.20 – 10.10.20	
3	Огляд технологій динамічної біометричної аутентифікації	11.10.20 – 10.11.20	
4	Проведення експериментальних досліджень	11.11.20 – 30.11.20	
5	Перевірка роботи на антиплагіат	01.12.20 – 05.12.20	
6	Представлення атестаційної роботи на кафедрі	10.12.2020	

Дата видачі завдання

02 вересня 2020 р.

Студент

(підпис)

Керівник роботи

(підпис)

(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 100 с., 40 рис., 6 табл., 33 джерела, 2 додатки.

БІОМЕТРИЧНІ СИСТЕМИ, ІДЕНТИФІКАЦІЯ, КЛАСИФІКАЦІЯ ДАНИХ, DATA MINING, ІНФОРМАЦІЙНИЙ ПОЧЕРК, ДИНАМІКА КОМП'ЮТЕРНОЇ МИШІ, ТРАЄКТОРІЯ КУРСОРУ.

Об'єкт дослідження – біометричні системи ідентифікації особи.

Предмет дослідження – аутентифікація за динамікою системи «користувач – комп'ютерна миша».

Метою цієї роботи є підвищення інформаційної безпеки комп'ютерних систем на основі аналізу курсорного почерку.

У роботі проведено аналітичний огляд методів біометричної аутентифікації, що базуються на клавіатурному та курсорному почерку. За допомогою датасету «The Balabit Mouse Challenge Data Set» та програмного забезпечення Orange було проведено експериментальні дослідження інформативності параметрів курсорного почерку. Проаналізовано особливості динаміки миші трьох типів дій миші (mouse movement, point and click, drag and drop). Проаналізовано 38 параметрів, що характеризують динаміку та траєкторії рухів курсору миші. Набір даних оцінювали за двома сценаріями: використовуючи лише навчальну частину набору даних та використовуючи весь набір даних. Для сценарію, коли система перевіряє поточного користувача на приналежність до класу «зловмисник», використовуючи «mouse movement» дії та параметри, що враховують прискорення, ривок та форму траєкторії курсору, значення параметру «True Negative» склало 99%.

ABSTRACT

Master thesis: 100 p., 6 tables, 40 fig., 33 sources, 2 annex.

BIOMETRIC SYSTEMS, IDENTIFICATION, DATA CLASSIFICATION, DATA MINING, USER BEHAVIOR ANALYTICS, MOUSE CLICKSTREAM, MOUSE ACTION, CURSOR TRAJECTORY.

The objects of the research are the biometric systems of personal identification.

The subject of the research is the authentication by the personal mouse dynamics usage pattern.

The objective of the work is explore the possibility of using mouse dynamics in computer networks' user identification tasks.

An analytical review of the keystroke and mouse dynamics based biometric authentication methods has been completed in this work. Using "The Balabit Mouse Challenge Data Set" and Orange software, has been conducted experimental studies of the informativeness of the computer mouse dynamics features. Has been analyzed mouse dynamics features of three types of mouse actions (mouse movement, point and click, drag and drop). Has been analyzed 38 features extracted from individual mouse actions. The dataset was evaluated using two scenarios: using only the training part of the data set, and using the whole data set. For the scenario of the system checking the current user for belonging to the "attacker" class using "mouse movement" actions and features that take into account acceleration, jerk and the shape of the mouse cursor trajectory, the "True Negative" value 99% was obtained.

ЗМІСТ

Перелік скорочень та термінів	7
Вступ	8
1 Технології статичної біометричної аутентифікації	10
2 Технології динамічної біометричної аутентифікації	18
2.1. Принципи стабільної динамічної біометрії	18
2.2. Аутентифікація за рукописним почерком	23
2.3. Аутентифікація за клавіатурним почерком	31
2.4. Голосова аутентифікація	41
2.5. Аутентифікація за динамікою системи «користувач – комп’ютерна миша»	46
3 Дослідження інформативних параметрів динаміки системи «користувач-комп’ютерна миша» для задач біометричної ідентифікації	48
3.1. The Balabit Mouse Challenge Data Set	48
3.2. Схема експерименту та результати проведених досліджень	56
Висновки	79
Перелік джерел посилання	80
Додаток А. Комплект графічних матеріалів	84
Додаток Б. Публікації за темою атестаційної роботи	98

ПЕРЕЛІК СКОРОЧЕНЬ ТА ТЕРМІНІВ

CA (Classification accuracy) – точність класифікації;

EER (Equal Error Rate) – коефіцієнт рівної імовірності помилок 1 і 2-го роду;

FAR (False Acceptance Rate) – помилка другого роду – випадок надання системою доступу неавторизованому користувачеві;

FN (False Negative) – ситуація, коли зареєстрованого користувача ідентифіковано як зловмисника і в доступі відмовлено;

FP (False Positive) – ситуація, коли зловмисника ідентифіковано як зареєстрованого користувача та дозволено отримати доступ;

FRR (False Rejection Rate) – помилка першого роду – доступ заборонений користувачеві, зареєстрованому в системі;

TN (True Negative) – ситуація, коли зловмисника правильно не ідентифіковано і в доступі відмовлено;

TP (True Positive) – ситуація, коли зареєстрованого користувача ідентифікований вірно і дозволено доступ;

АРП – автоматичне регулювання підсилення;

ЧНК – час натискання клавіші;

ШНМ – штучна нейронная мережа.

ВСТУП

Правильна організація розмежування доступу до автоматизованого робочого місця і аутентифікація користувачів є необхідною умовою коректної роботи інформаційних систем і, як наслідок, життєздатності автоматизованих бізнес-процесів. На технічному рівні забезпечення гарантії легальності автоматизованого робочого місця користувача зводиться до вирішення ряду питань його ідентифікації.

Своєчасність ідентифікації користувача, який виконує роботу за автоматизованим робочим місцем, є важливою умовою досягнення ефективності результатів його роботи. При використанні того чи іншого методу ідентифікації необхідно розглядати і приймати до уваги загрози інформаційній безпеці, пов'язані з можливістю зловмисника отримати доступ до автоматизованого робочого місця, видаючи себе за легального користувача.

У традиційних системах розмежування доступу ідентифікація користувачів здійснюється одноразово при проходженні процедури авторизації. Однак сценарій використання автоматизованих робочих місць припускає, що легальний користувач може регулярно залишати робоче місце, а облікові дані можуть бути загублені або викрадені. Подібні загрози, що виникають в комп'ютерних системах, є актуальними і вказують на низьку ефективність традиційних методів парольної ідентифікації або ідентифікації на основі технічних пристроїв у вирішенні проблеми виявлення підміни користувачів автоматизованих робочих місць. Практичне використання даних методів не дозволяє гарантувати присутність користувача на робочому місці, так як користувач може бути відсутнім після аутентифікації, тобто під час активної сесії.

Для вирішення даної проблеми розпізнавання користувача актуально проводити не тільки на етапі авторизації, але і з певною періодичністю під час сеансу роботи в комп'ютерній системі. При цьому ефективність розпізнавання користувача буде залежати від використовуваного підходу до вирі-

шення задачі ідентифікації. Так, наприклад, класичні паролльні системи ідентифікації не в змозі повністю вирішити дану проблему, так як пароль, що зберігається на фізичному або цифровому носії, може бути втрачено або викрадено зловмисником. Очевидно, що для розпізнавання користувача необхідно використовувати ідентифікатори, які не можна передати, втратити або викрасти, тобто біометричні ознаки.

Розглянемо критерії вибору методів біометричної ідентифікації, використання яких дозволить ефективно вирішувати задачі виявлення підміни користувачів в комп'ютерних системах:

1) відсутність необхідності в апаратному забезпеченні крім стандартного робочого місця користувача (в іншому випадку методика розпізнавання користувача буде занадто дорогою);

2) можливість розпізнавання користувача непомітно для нього (даним критерієм задовольняють динамічні біометричні методи.

Таким чином, найбільш переважними методами розпізнавання користувачів в комп'ютерних системах є динамічні методи біометричної ідентифікації. З урахуванням того, що в сучасних інформаційних системах використовуються графічні інтерфейси, а основну частину часу користувач використовує комп'ютерну мишу, одним з найбільш перспективних методів є ідентифікація по динаміці системи користувач-комп'ютерна миша.

Метою роботи є підвищення інформаційної безпеки комп'ютерних мереж на основі аналізу динаміки системи користувач-комп'ютерна миша.

Для досягнення поставленої мети необхідно розв'язати наступні задачі:

1) провести аналіз принципів стабільної динамічної біометрії;

2) провести пошук відкритих датасетів параметрів курсорного почерку та обрати один з них для подальших досліджень;

3) на основі обраного датасету запропонувати інформативні параметри курсорного почерку для створення математичної моделі класифікатора користувачів за динамікою системи «користувач – комп'ютерна миша»;

4) проведення експериментальних досліджень.

1 ТЕХНОЛОГІЇ СТАТИЧНОЇ БІОМЕТРИЧНОЇ АУТЕНТИФІКАЦІЇ

Людині властиво розвиватися, міняти свої звички, зовнішність і поведінку, однак у неї є й ознаки, що залишаються незмінними протягом всього життя, наприклад, малюнок відбитка пальця або кровеносних судин очного дна. Усі ці властивості можуть використовуватися біометричними системами як засіб ідентифікації. І залежно від цього методи біометричної ідентифікації діляться на дві великі групи – статичні й динамічні.

Технології статичної біометрії – пристрої і біометричні програми, побудовані на аналізі статичних (незмінних) фізіологічних характеристик людини. До числа цих характеристик входять [1]: відбитки пальців; форма і геометрія обличчя; форма і будова черепа; сітківка ока; райдужна оболонка ока; геометрія долоні, кисті руки або пальця; термографія обличчя, термографія руки; малюнок вен на долоні або пальці руки; ДНК; запах тіла; форма вуха.

Аутентифікація за відбитком пальця. Дактилоскопія – найбільш популярна технологія біометричної аутентифікації, заснована на скануванні і розпізнаванні відбитків пальців.

Даний метод активно підтримується правоохоронними органами, з метою залучення в свої архіви електронних зразків. Також, метод сканування відбитків пальців легкий у використанні і надійний з погляду на універсальність даних. Головним пристроєм цього методу біометричної аутентифікації є сканер, який сам по собі має невеликі розміри і є відносно недорогим в ціні. Така аутентифікація здійснюється досить швидко за рахунок того, що система не вимагає розпізнавання кожної лінії рисунка папілярних ліній і порівняння її зі зразками, що знаходяться в базі. Системі досить визначити збіги в масштабних блоках і проаналізувати роздвоєння, розриви та інші спотворення ліній (мінуції).

Унікальність кожного відбитка дозволяє використовувати даний метод

біометричної аутентифікації як в криміналістиці або в процесах серйозних бізнес-операцій, так і в побуті. Останнім часом з'явилась безліч ноутбуків з вбудованим сканером відбитків пальців, клавіатур, комп'ютерних мишей, а також смартфонів для аутентифікації користувача.

До недоліків даного методу біометричної аутентифікації можна віднести наступне. Через використання складних алгоритмів розпізнавання дрібних папілярних ліній, система аутентифікації може демонструвати збої при недостатньому контакті пальця зі сканером. Також даний метод погано захищений від спуфінг-атак (ситуацій, коли зловмисник успішно маскується під зареєстрованого користувача шляхом фальсифікації біометричних даних і отримує незаконний доступ).

За принципом роботи, що використовуються для аутентифікації сканери, діляться на три види:

1) оптичні сканери, що функціонують на технології відбиття світла, або за принципом просвіту світлом. З усіх видів, оптичне сканування не здатне розпізнати муляж, однак, завдяки своїй вартості і простоті, саме оптичні сканери найбільш популярні;

2) напівпровідникові сканери – підрозділяються на радіочастотні, ємнісні, термочутливі і чутливі до тиску сканери. Теплові (термосканери) і радіочастотні сканери краще за всіх здатні розпізнати справжній відбиток і не допустити аутентифікацію за допомогою муляжу пальця. Напівпровідникові сканери вважаються більш надійними, ніж оптичні;

3) ультразвукові сканери. Даний вид пристроїв є найскладнішим і дорогим. За допомогою ультразвукових сканерів можна здійснювати аутентифікацію не тільки за відбитками пальців, але і по деяким іншим біометричними параметрами, таким як частота пульсу.

Аутентифікація за сітківкою ока. Даний метод стали використовувати ще в 50-х роках минулого століття. У той час, як раз, була вивчена і визначена унікальність рисунка кровоносних судин очного дна.

Сканери сітківки ока мають досить великі габарити і більш високу ці-

ну, ніж сканери відбитків пальців. Однак, надійність такого виду аутентифікації набагато вище дактилоскопії, що і виправдовує затрати. Особливості рисунка кровоносних судин очного дна такі, що він не повторюється навіть у близнюків. Тому, така аутентифікація має максимальний захист. Підробити біометричні дані для такого сканеру також практично неможливо. Збої при розпізнаванні очного рисунка незначно малі – приблизно, один на мільйон випадків. Якщо, у користувача немає серйозних очних захворювань (наприклад, катаракти), він може впевнено використовувати систему аутентифікації за сітківкою ока для захисту доступу до сховищ, приватних кабінетів і надсекретних об'єктів.

Сканування сітківки ока передбачає використання інфрачервоного низько інтенсивного випромінювання, яке підсвічує кровоносні судини очного дна через зіницю. Сигнал відображає кілька сотень характерних точок, які записуються в шаблон. Найсучасніші сканери замість інфрачервоного світла використовують лазер м'якої дії.

Для проходження даної аутентифікації, людина повинна максимально наблизити до сканера обличчя (око повинно бути не далі 1.5 см від пристрою), зафіксувати його в одному положенні і направити погляд на дисплей сканера, на спеціальну мітку. Близько сканера, в такому положенні, доводиться перебувати приблизно хвилину. Саме стільки багато часу потрібно сканеру для здійснення операції сканування, після чого, системі знадобиться ще кілька секунд для порівняння отриманого зразка з встановленим шаблоном. Тривале знаходження в одному положенні і фіксація погляду на спалах світла і є найбільшими недоліками використання даного виду аутентифікації. Плюс – через відносно тривале сканування сітківки і обробку результатів, даний пристрій неможливо встановлювати для аутентифікації великої кількості людей (наприклад, на прохідний).

Аутентифікація за райдужною оболонкою ока. Даний метод аутентифікації заснований на розпізнаванні унікальних особливостей райдужної оболонки ока.

Схожий на мережу, складний рисунок рухомої діафрагми між задньою і передньою камерами ока – це і є унікальна райдужна оболонка. Даний рисунок формується ще до народження людини і особливо не змінюється протягом усього життя. Надійності аутентифікації методом сканування райдужної оболонки ока сприяє відмінність рисунка райдужки у лівого та правого очей людини. Така технологія, практично, виключає помилки і збої в процесі аутентифікації.

Однак, складно назвати пристрої, що зчитують рисунок райдужної оболонки – сканерами. Це, швидше за все, спеціалізована камера, яка робить 30 знімків в секунду. Потім оцифровується один із кадрів і перетворюється в спрощену форму, з якої відбираються близько 200 характерних точок і інформація по ним записується в шаблон. Це куди більш надійно, ніж сканування відбитків пальців – для формування таких шаблонів використовуються всього лише 60-70 характерних точок.

Даний вид аутентифікації передбачає додатковий захист від підробки – в деяких моделях пристроїв для визначення «живого ока» змінюється потужність потоку світла, спрямованого в око і система відстежує реакцію і визначає чи змінюється розмір зіниці.

Дані сканери вже широко використовуються, наприклад, в аеропортах багатьох країн для аутентифікації співробітників під час перетину зон обмеженого доступу, а також, непогано зарекомендували себе в Англії, Німеччині, США та Японії під час експериментального використання з банкоматами. Слід зазначити, що в процесі аутентифікації за райдужною оболонкою ока, на відміну від сканування сітківки, сканер-камера може перебувати на відстані від 10 см до 1 метра від ока, а процес сканування і розпізнавання проходить набагато швидше.

Вартість завжди була найбільшим стримуючим фактором перед впровадженням технології, але зараз системи ідентифікації за райдужною оболонкою стають доступнішими для різних компаній. Прихильники технології заявляють про те, що розпізнавання райдужної оболонки ока дуже скоро стане

загальноприйнятою технологією ідентифікації в різних областях.

Аутентифікація за геометрією руки. Даний метод біометричної аутентифікації передбачає вимір певних параметрів кисті людини, наприклад: довжина, товщина і вигин пальців, загальна структура кисті, відстань між суглобами, ширина і товщина долоні.

Руки людини не є чимось унікальним, тому для надійності даного виду аутентифікації необхідно комбінувати розпізнавання відразу за кількома параметрами. Імовірність помилок при розпізнаванні геометрії кисті становить близько 0,1%, а це значить, що при ударі, артриті та інших захворюваннях і пошкодженнях кисті, швидше за все, пройти аутентифікацію не вдасться. Тому даний метод біометричної аутентифікації не підходить для забезпечення безпеки об'єктів високого ступеня секретності.

Однак, даний метод знайшов широке поширення, завдяки тому, що він зручний для користувачів по цілому ряду причин. По-перше, пристрій для розпізнавання параметрів руки не призводить до дискомфорту і не забирає багато часу (весь процес аутентифікації здійснюється за кілька секунд). По-друге, ні температура, ні забрудненість, ні вологість кисті не впливають на процедуру аутентифікації. Також, зручний даний метод і тим, що для розпізнавання кисті можна використовувати зображення низької якості – розмір шаблону, що зберігається в базі всього 9 байт. Процедура порівняння кисті користувача з встановленим шаблоном дуже проста і легко може бути автоматизована.

Пристрої даного виду біометричної аутентифікації можуть мати різний зовнішній вигляд і функціонал – одні сканують лише два пальця, інші роблять знімок всієї руки, а деякі сучасні пристрої за допомогою інфрачервоної камери сканують вени і по їх зображенню здійснюють аутентифікацію. Даний метод вперше був використаний на початку 70-х років минулого століття. Сьогодні подібні пристрої можна зустріти в аеропортах і різних підприємствах, де необхідно формувати достовірні відомості про присутність того, чи іншого працівника, обліку робочого часу та інших процедур контролю.

Аутентифікація за геометрією обличчя. Широке застосування мультимедійних технологій, велика кількість відеокамер на вокзалах, аеропортах, площах, вулицях, дорогах і інших місцях скупчення людей, сприяли широкому поширенню даного способу ідентифікації і аутентифікації. В останні роки ще більший імпульс до розвитку цей напрям одержав з поширенням технологій відеоконференцій в мережі Internet. Орієнтація на стандартні відеокамери персональних комп'ютерів робить цей клас біометричних систем порівняно дешевим.

Алгоритм функціонування системи розпізнавання досить простий. Зображення обличчя зчитується звичайною відеокамерою та аналізується. Програмне забезпечення порівнює введений портрет з еталоном, що зберігається в пам'яті. Деякі системи додатково архівують зображення, що аналізуються в процесі аутентифікації, для можливого в майбутньому розбору конфліктних ситуацій. Вельми важливо також те, що біометричні системи цього класу потенційно здатні виконувати безперервну ідентифікацію (аутентифікацію) користувача комп'ютера протягом всього сеансу його роботи. Більшість алгоритмів дозволяє компенсувати наявність окулярів, капелюхів і бороди у досліджуваного індивіда.

Основними проблемами, з якими стикаються розробники даного класу біометричних систем, є зміна освітленості, варіації положення голови користувача, виділення інформативної частини портрета (гасіння фону). З цими проблемами вдається впоратися, автоматично виділяючи на обличчі особливі точки і потім вимірюючи відстані між ними. На обличчі виділяють контури очей, брів, носа, підборіддя. Відстані між характерними точками цих контурів утворюють вельми компактний еталон конкретного обличчя, що легко піддається масштабуванню. Задача виділення контурів характерних деталей обличчя легко може бути вирішена для плоских двовірних зображень з фронтальним підсвічуванням, але такі біометричні системи зломисник може легко обійти за допомогою плоского зображення обличчя-оригіналу. Для двовірних систем виготовлення муляжу-фотографії – це нескладна технічна

задача.

Істотні технічні труднощі при виготовленні муляжу виникають при використанні тривимірних біометричних систем, що реєструють перепади яскравості відбитого світла при відновленні тривимірного зображення обличчя. Такі системи здатні компенсувати невизначеність розташування джерела освітлення по відношенню до ідентифікованого обличчя, а також невизначеність положення обличчя по відношенню до відеокамери. Обійти системи цього класу можна тільки об'ємною маскою, яка точно відтворює оригінал.

Аутифікація по термограмою обличчя. Більш надійним різновидом ідентифікації за геометрією обличчя є ідентифікація за «тепловим портретом» обличчя людини в інфрачервоному діапазоні. В даній технології здійснюється термографія ідентифікованого обличчя, що виявляє унікальність розподілу артерій на обличчі, що постачають шкіру теплою кров'ю. Проблема підсвічування для цього класу біометричних пристроїв не існує, так як вони сприймають тільки температурні перепади обличчя і можуть працювати в повній темряві. На результати ідентифікації не впливають перегрів обличчя, його переохолодження, природне старіння людини, пластичні операції, так як вони не змінюють внутрішнє розташування судин. Методом лицьової термографії можна розрізняти однайцевих близнюків, бо кровоносні судини на їх обличчях мають досить суттєві відмінності. Дистанційне зчитування з будь-якої відстані незалежно від освітленості забезпечує високу пропускну здатність і вандалозахищеність. Метод розрахований на використання спеціалізованої відеокамери далекого інфрачервоного діапазону, що і визначає його високу вартість.

Наведемо порівняльні таблиці, де якісні параметри різних біометричних методів зведені разом. У стовпчиках вказані ті критерії, яким повинен відповідати будь-біометричний метод, і якісна оцінка кожного біометричного методу за цими критеріями.

Таблиця 1.1 – Характеристики статичних біометричних методів

№	Метод ідентифікації	Біометрична характеристика	Тип методу	Рекомендований рівень безпеки	Застосування методу
1	Відбиток пальця (опт.)	Папілярний рисунок на пальці	Статичний / контактний	Середній	Універсальний
2	Відбиток пальця (ел.)	Папілярний рисунок на пальці	Статичний / контактний	Середній	Універсальний
3	Відбиток пальця (УЗ)	Папілярний рисунок на пальці	Статичний / контактний	Середній	Секретні приміщення
4	Геометрія руки	Форма і об'єм кисті руки	Статичний / контактний	Середній	Приміщення, що охороняються
5	Портрет	Геометрія обличчя	Статичний / безконтактний	Середній	Приміщення, що охороняються
6	3D-портрет	Об'ємна геометрія обличчя	Статичний / безконтактний	Середній	Приміщення, що охороняються
7	Райдужка оболонки ока	Форма і рисунок райдужної оболонки ока	Статичний / безконтактний	Високий	Секретні приміщення
8	Сітківка ока	Розташування судин на сітківці ока	Статичний / безконтактний	Надвисокий	Секретні приміщення
9	Термограма обличчя	Розташування вен і температура обличчя	Статичний / безконтактний	Високий	Приміщення, що охороняються
10	Відбиток долонь	Розташування папілярних ліній на долоні	Статичний / контактний	Середній	Приміщення, що охороняються

Таблиця 1.2 – Оцінка статичних біометричних методів

№	Простота	Стійкість до підробки	Помилка розпізнавання % (порядок)		Поширеність методу	Вартість обладнання	Швидкість, сек
			FAR	FRR			
1	3	Погана	$10^{-3} - 10^{-1}$	$0.01 - 10$	Велика	Низька	0.015
2	3	Погана	$10^{-3} - 10^{-1}$	$0.01 - 10$	Велика	Низька	0.055
3	4	Середня	$10^{-4} - 10^{-2}$	$0.001 - 1$	Велика	Середня	0.015
4	3	Низька	$10^{-3} - 10^{-2}$	$0.01 - 1$	Середня	Середня	1-5
5	3	Погана	$10^{-2} - 1$	$0.1 - 10$	Середня	Середня	1-5
6	4	Середня	$10^{-4} - 10^{-2}$	$0.001 - 0.1$	Дуже мала	Середня	1-10
7	4	Хороша	$10^{-8} - 10^{-5}$	$10^{-3} - 10^{-1}$	Середня	Висока	1-10
8	5	Відмінна	$10^{-8} - 10^{-7}$	$10^{-3} - 10^{-1}$	Мала	Дуже висока	> 20
9	4	Відмінна	$10^{-4} - 10^{-2}$	$0.001 - 0.1$	Дуже мала	Висока	1-10
10	3	Середня	$10^{-4} - 10^{-2}$	$0.001 - 0.1$	Мала	Середня	0.15

2 ТЕХНОЛОГІЇ ДИНАМІЧНОЇ БІОМЕТРИЧНОЇ АУТЕНТИФІКАЦІЇ

2.1. Принципи стабільної динамічної біометрії.

Поряд зі статичними методами біометричної ідентифікації особистості на практиці використовуються методи динамічної біометрії. До них в першу чергу слід віднести найбільш поширені методи ідентифікації особистості за голосом і рукописним почерком [2]. Очевидно, що методи динамічної біометрії мають як переваги, так і недоліки.

В якості першої переваги слід зазначити відносно низьку вартість систем динамічної біометрії. Останнє обумовлено тим, що вони можуть бути побудовані з використанням стандартних вже наявних на комп'ютері засобів введення інформації та/ або мультимедіа. В цьому випадку вартість системи буде визначатися тільки вартістю програмного забезпечення.

Ще однією перевагою динамічної біометрії є потенційна можливість зміни біометричного зразка користувача за рахунок швидкої зміни відтвореного слова (парольної фрази). Статичний зразок особистості (відбиток пальця) неможна зберегти в таємниці або змінити, він належить людині. Інша справа зразки динамічної біометрії – вони легко піддаються зміні і можуть бути збережені користувачем в таємниці. Останнє робить динамічну біометрію більш привабливою для віддаленої аутентифікації (ідентифікації) особистості у відкритому інформаційному просторі. Динамічна біометрія принципово сильніше статичної біометрії, так як за рахунок таємниці біометричних зразків теоретично може бути забезпечений будь-який заданий рівень інформаційної безпеки. Можливість збереження в таємниці біометричних зразків набагато важливіша за будь-які досяжні технічні характеристики біометрії відкритих зразків. Кращі системи біометрії відкритих зразків (рисунок вен, очного дна) забезпечують рівень ймовірності помилок другого роду порядку 10^{-9} , в той час як збереження в таємниці біометричного способу дає ймовірність помилок цього типу на рівні $10^{-20} \dots 10^{-40}$ [3].

Важливою властивістю динамічної біометрії є анонімність її зразків. Перехоплення зловмисником зразку динамічної біометрії не дає йому можливості знайти за цим зразком його власника. Навіть якщо цей зразок відкритий, то в його еталоні зберігається тільки частина індивідуальної інформації, за якою неможна повністю відновити первинний зразок. Знаючи відбиток пальця можна знайти його власника. Знаючи частину коефіцієнтів ряду Фур'є (параметри динамічної біометрії), неможна однозначно відновити криві динаміки почерку або особливостей голосу. Анонімність даних для динамічної біометрії забезпечується набагато простіше, ніж для статичної біометрії.

Як недолік цього класу систем слід відзначити те, що на їх роботу впливає психофізичний стан особистості (переляк, стрес, психотропні препарати). Проте, системи динамічної біометрії є ефективними тривалий час, існуючі дані свідчать про стабільне збереження відпрацьованих рухових навичок у людини протягом декількох десятиліть.

Питання про те, чому взагалі працюють системи динамічної біометрії і чому вони працюють стійко протягом тривалих періодів часу, вкрай важливий. Мабуть, єдиним правдоподібним поясненням цього є величезна надмірність рухових можливостей людини плюс обмеженість її обчислювальних ресурсів (головного і спинного мозку) [4].

У техніці добре відомо, що в реальному часі без особливих проблем вирішуються задачі, розмірність яких збігається з розмірністю обчислювальної машини. Мозок людини утворений з нейронів, розташованих в тривимірному просторі (кожен з нейронів має кілька входів і один вихід). Проте, виходячи з сучасних уявлень фізіологів [5] про мозок, він є двовимірною машиною. Саме тому відомі сьогодні нейронні мережі розглядаються тільки як двовимірні обчислювальні машини.

Якщо оцінити розмірність розв'язуваних мозком людини динамічних задач в повсякденному житті, то у першому наближенні вона дорівнює кількості м'язів, задіяних при тих чи інших діях. Наприклад, під час письма ручкою задіяні м'язи більшості пальців і частина м'язів передпліччя [6]. Загальна

кількість задіяних м'язів більше 50, але найбільш істотний вплив роблять приблизно 10 м'язів. Отже, в процесі рукописного підпису людина вирішує в реальному часі десятимірну задачу управління (відповідна цій ситуації механічна модель наведена на рис. 2.1).

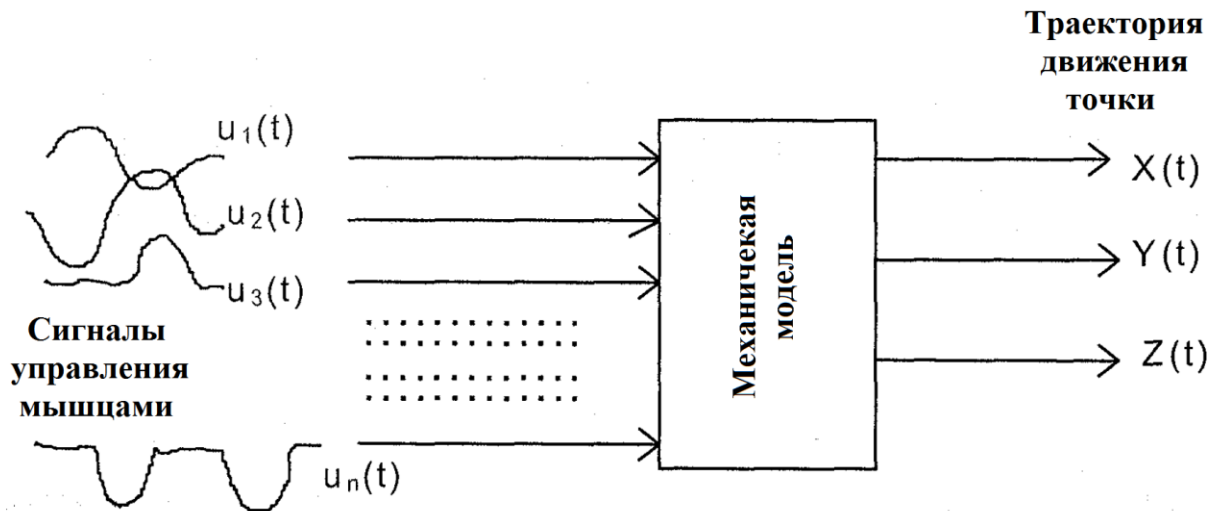


Рисунок 2.1 – Багатовимірна механічна модель управління людиною положенням матеріальної точки в просторі

При роботі з клавіатурою (сліпе друкування тексту), додатково включаються ще приблизно 20 м'язів плеча і плечового пояса на кожній руці. Виходить, що при сліпому друку двома руками потенційно можуть бути задіяні 140 м'язів. Виходячи з припущення, що найбільший вплив роблять тільки 20 % від загального числа м'язів, можна отримати $n = 0.2 \cdot 140 = 28$, тобто приблизно 28-мірну задачу управління.

При відтворенні мови беруть участь 44 м'язи грудей, 9 м'язів живота і черевної порожнини, 28 м'язів обличчя і щелеп, 12 м'язів язика, 9 м'язів глотки, 6 м'язів м'якого піднебіння і м'язи гортані. Таким чином, задіяно близько 120 м'язів, що при гіпотезі 20% групи найбільшого впливу дає 24-мірну задачу управління. Також слід зазначити, що вирішення задачі мовотворення є найбільш складним, бо для цієї задачі керуючі сигнали повинні вироблятися найбільш швидко – частота основного тону знаходиться в межах від 60 до 600 Герц в порівнянні з більш м'якими режимами управління пальцями рук,

для яких характерні істотно менші частоти керуючих сигналів (спектр частот для підпису від 0.1 до 10 Гц).

Таким чином, динамічні задачі управління, що вирішує людина в повсякденному житті, мають розмірність набагато більше, ніж обчислювальна розмірність мозку і людина нездатна швидко і точно виконувати зовсім нові для неї складні рухи. Для підсвідомого здійснення складних рухів, людина повинна довго шукати (підбирати) вдалі рішення складних динамічних задач і далі запам'ятовувати їх шляхом тренувань (повторень). Такий спосіб вирішення складних динамічних задач управління дуже трудомісткий і природа використовує його всього один раз. Після того, як протягом декількох років у індивідуума вироблені хода, почерк, голос тощо природа не може собі дозволити розкоші перенавчання заново. Наприклад, навчитися керувати велосипедом важко. Однак, як тільки людина навчиться керувати велосипедом, не замислюючись і не напружуючись, вона вже ніколи не втратить ці навички.

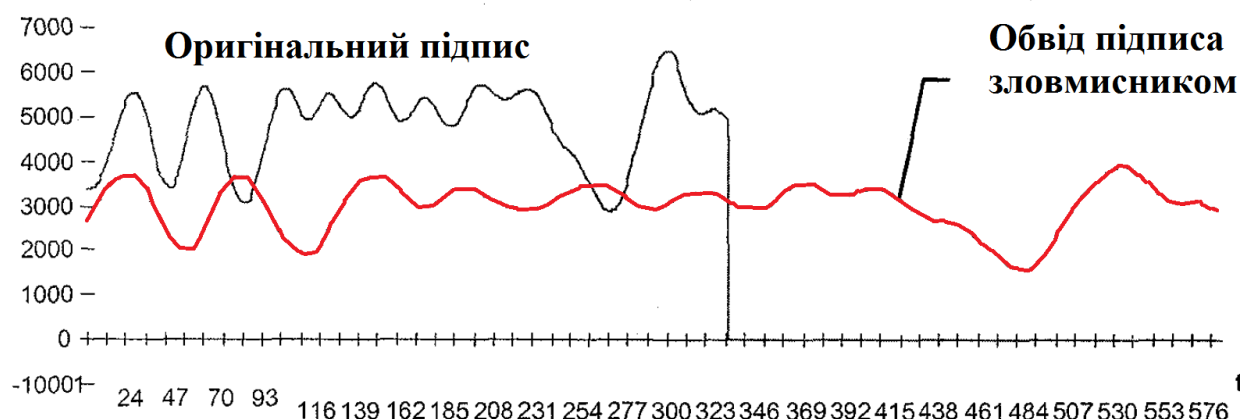


Рисунок 2.2 – Коливання пера автора і злоумисника, який обводить підпис

Однією з переваг підсвідомих рухів є істотне зниження часу їх виконання. Цей ефект добре ілюструє рис. 2.2, де зображено графіки коливання пера, що отримано при відтворенні свого підпису автором та злоумисником, який обводить цей же підпис. Досить примітивні дії обводу злоумисник виконує зі швидкістю, приблизно в два рази нижче, ніж швидкість складних дій, виконуваних автором. Якщо злоумисник намагається якомога точніше відтворити траєкторію чужого підпису, то його швидкість падає ще більше,

ніж при примітивному обводі.

Слід зазначити, що підсвідомі рухи стабільні, якщо в них не втручається більш високий свідомий рівень. В процесі втручання свідомості високого рівня відбувається переривання «підпрограми управління» до моменту прийняття рішення. Ця ситуація проілюстрована рис. 2.3, де зображено координати коливання пера автора підпису, який на мить замислився про правильність відтвореної ним конфігурації підпису. Переривання підпрограми управління рухами відповідає горизонтальній ділянці для обох координат.



Рисунок 2.3 – Ефект втручання свідомості у підсвідому програму управління

Ще одним наслідком багатовимірності задачі управління рухами є унікальність динаміки рухів кожної людини. Якщо кількість входів точно збігається з розмірністю загальної задачі (числом виходів на рис. 2.1), то оптимальні рішення задачі управління виявляються дуже схожими. Якщо ж кількість входів істотно перевищує кількість виходів, то подібні результати можна досягти абсолютно різними сумами керуючих впливів. З одного боку, подібна надмірність є гарантією працездатності системи в цілому при поганій роботі частини м'язів. З іншого боку, через велику надмірність задачі кожен індивідум знаходить своє рішення задачі управління. Однакові результати різні люди можуть отримати, працюючи різними групами м'язів.

Підводячи підсумки, слід зазначити наступне.

1. Динамічна біометрія потенційно сильніше статичної через можливість збереження в таємниці її біометричних зразків і швидкої їх зміни за необхідності.

2. Динамічна біометрія дешевше статичної, так як може використовувати стандартні пристрої введення інформації.

3. Стабільність динамічної біометрії пов'язана з величезними труднощами початкового навчання людини та існуючою в природі заборонаю на перенавчання.

2.2. Аутентифікація за рукописним почерком.

Рукописний підпис – це один з класичних способів ідентифікації особистості, застосовуваний вже кілька століть в багатьох сферах людської діяльності. Даний метод ідентифікації особистості побудований на тому, що автор сам придумує собі факсиміле і добре відпрацьовує його шляхом тренувань. Бажано, щоб авторське факсиміле мало істотні відмінності за формою від класичного написання букв у вигляді додаткових елементів (розчерків, повернень, накладення букв).

Відповідно до [7, 8] проблему ідентифікації особистості за її факсимільним підписом слід розглядати як дві незалежні проблеми:

1) ідентифікацію особистості тільки за статичним зображенням підпису;

2) ідентифікація автора за параметрами динаміки відтворення підпису в момент проходження процедури ідентифікації, впродовж якої проводиться спостереження за індивідуальними особливостями звичних для автора підсвідомих рухів.

Зазначені вище дві постановки задачі мають досить суттєві відмінності, але обидві ці задачі можуть вирішуватися паралельно і незалежно. У першій постановці задачі мова йде про порівняння зображень, відтворених раніше в невідомій послідовності. У другій постановці задачі є дані про параметри ко-

ливання пера при відтворенні підпису в тривимірному просторі. Якщо користуватися декартовою системою координат (X, Y, Z) , то дані про динаміку відтворення підпису отримують у вигляді двох функцій часу коливань пера в площині графічного планшета $X(t), Y(t)$ і ще однієї функції – варіації тиску пера на графічний планшет $Z(t)$.

Перша постановка задачі дуже складна для рівня розвитку сучасних інформаційних технологій і, як правило, автоматичні банківські системи ідентифікації за статичним зображенням підпису працюють істотно гірше досвідченого експерта. Як правило, ці системи тільки напіваавтоматичні – вони полегшують роботу експерта, даючи йому відповідні чисельні характеристики близькості фрагментів дослідного зображення підпису до оригіналу, але остаточне рішення приймає людина. Відомі на сьогодні технічні рішення першого варіанту постановки задачі по імовірнісним характеристикам істотно гірше, ніж статистика роботи досвідчених людей (експертів). Саме з цієї причини фірми-виробники не дають статистичних даних про помилки першого і другого роду для ідентифікації автора тільки за статичним зображенням підпису.

У другій постановці задачі автоматизована система ідентифікації має істотно більше інформації в порівнянні з експертом, тому вони за своїми статистичними характеристиками виявляються істотно краще експертів. Ілюстрацією цього можуть служити ситуації обводу чужого підпису (рис. 2.2) і зупинок для прийняття свідомого рішення під час підробки (рис. 2.3). Експерт не володіє подібною динамічною інформацією і тільки з цієї причини він виявляється гірше автоматизованої системи.

Практично всі існуючі сьогодні комерційні системи ідентифікації особистості цього класу побудовані в основному на аналізі динаміки відтворення підпису. Однак в цих системах можуть бути присутніми і фрагменти більш складного аналізу статичних зображень підпису. Комерційні системи слід розділити на одно-, дво- і трьохкоординатні. Відповідно ці системи відрізняються тим, що аналізують одну криву, дві кривих або три кривих

$X(t), Y(t), Z(t)$. Однокоординатні системи можуть бути побудовані шляхом аналізу будь-якої з цих часових функцій, забезпечуючи ймовірність помилок першого і другого роду на рівні 0.1. Двокоординатні системи використовують будь-яку пару функцій часу з трійки $X(t), Y(t), Z(t)$ і дозволяють досягти рівня ймовірності помилок близько 0.01.

Слід зазначити, що деякі з систем біометричної ідентифікації за рукописним підписом використовують не самі проекції функцій $X(t), Y(t), Z(t)$ на осі координат, а їх першу або другу похідну. Останнє обумовлено тільки типом використовуваного датчика, чутливого до похідної, і практично не впливає на якість і обсяг дослідної інформації. Штучне нарощування кількості аналізованих функцій за рахунок вимірювання їх самих і їх же похідних недоцільно, так як всі лінійні перетворення первісних дають сильно корельовані дані. Зокрема, кореляція біометричних даних, отриманих з первісної $X(t)$, і біометричних даних, отриманих з її похідної $dX(t)/dt$, близька до одиниці [9,10].

Слід зазначити, що помилка першого роду або помилкова відмова справжнього автору з ймовірністю 0.01 – це цілком прийнятна характеристика для вимог сьогодення. Інша справа – помилки другого роду або хибний пропуск злоумисника. Для ряду практичних застосувань ймовірність помилок другого роду на рівні 0.01 є непринятно великою величиною. Для багаторазового зниження цієї помилки другого роду використовують ідентифікацію особистості за динамікою відтворення біометричного пароля. В якості біометричного пароля зазвичай використовується деяке секретне слово – пароль. Автором зберігається в секреті як саме слово-пароль, так і його особливості написання. У цьому випадку головною виявляється не сама біометрія, а знання користувача, тобто пароль. Особливості біометрії множаться на таємницю рукописно відтвореного слова.

Як приклад біометричного пароля може бути використано слово «Захист», графік якого і відповідні криві коливання пера по двох осях $Y(t), X(t)$ представлені на рис. 2.4 – 2.6. Функції $Y(t), X(t)$ на цих рисунках мають по 8

синхронних точок звернення до нуля, відповідних відриву пера автором від поверхні графічного планшета при відтворенні їм слова-пароля.

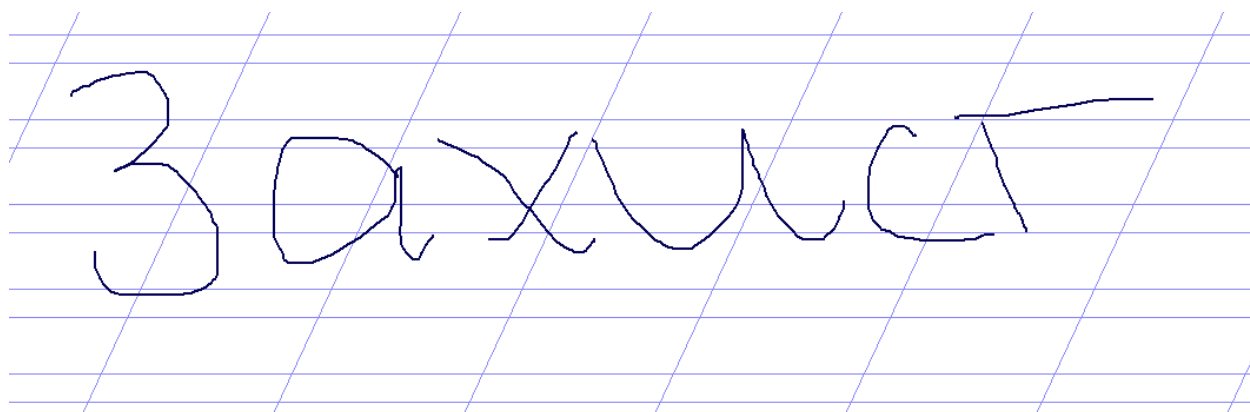


Рисунок 2.4 – Траєкторія відтворення слова-пароля

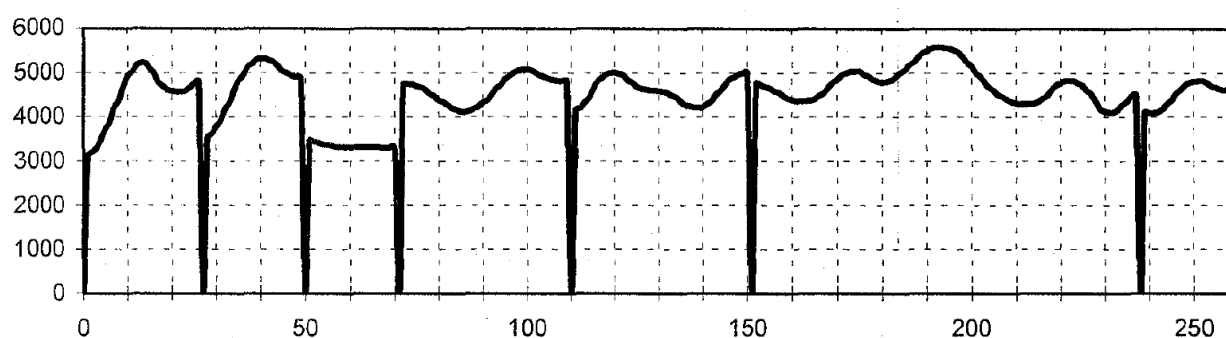


Рисунок 2.5 – Графік функції $Y(t)$ коливань пера по вертикалі

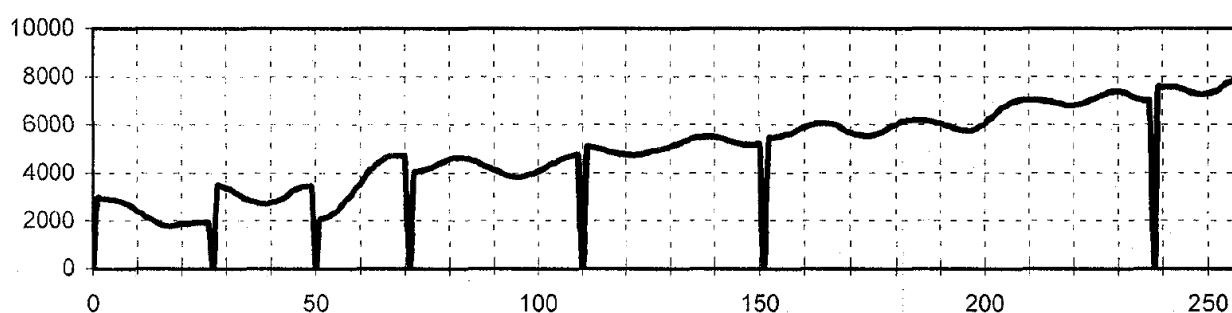


Рисунок 2.6 – Графік функції $X(t)$ горизонтальних коливань пера

Зазвичай використовувані при ідентифікації динамічні параметри отримують шляхом розрахунків деяких лінійних функціоналів по повній реалізації підпису (глобальні динамічні параметри) або по його фрагментах (локальні динамічні параметри), наприклад, різних буквах парольного слова.

У якості лінійних функціоналів зазвичай використовують ортогональні функціонали Фур'є, Уолша та Хаара. Вибір ортогональних перетворень обумовлений тим, що вони дають дані з меншою внутрішньою кореляцією в порівнянні з іншими, неортогональними функціоналами.

Наприклад, для функціоналу Фур'є глобальні динамічні параметри будуть мати сенс коефіцієнтів тригонометричного ряду Фур'є, що обчислюються наступним чином:

$$\begin{aligned} a_{yk} &= \frac{1}{T} \int_0^T Y(t) \cdot \cos\left(k \frac{2\pi}{T} t\right) dt, & b_{yk} &= \frac{1}{T} \int_0^T Y(t) \cdot \sin\left(k \frac{2\pi}{T} t\right) dt; \\ a_{xk} &= \frac{1}{T} \int_0^T X(t) \cdot \cos\left(k \frac{2\pi}{T} t\right) dt, & b_{xk} &= \frac{1}{T} \int_0^T X(t) \cdot \sin\left(k \frac{2\pi}{T} t\right) dt; \\ a_{zk} &= \frac{1}{T} \int_0^T Z(t) \cdot \cos\left(k \frac{2\pi}{T} t\right) dt, & b_{zk} &= \frac{1}{T} \int_0^T Z(t) \cdot \sin\left(k \frac{2\pi}{T} t\right) dt; \end{aligned} \quad (2.1)$$

де T – повний час відтворення підпису.

Зазвичай враховується скінченна кількість коефіцієнтів ряду Фур'є – зазвичай не більш перших $N = 20$ коефіцієнтів.

Лінійність використовуваних ортогональних функціоналів дозволяє вкрай просто здійснювати операцію масштабування сигналів по осях X, Y, Z . Для визначення невідомих масштабів тільки-но відтвореної паролльної фрази досить вирішити наступні лінійні рівняння:

$$\begin{aligned} \mu_y \cdot \sqrt{\sum_{k=1}^N (a_{yk}^2 + b_{yk}^2)} &= C_y; \\ \mu_x \cdot \sqrt{\sum_{k=1}^N (a_{xk}^2 + b_{xk}^2)} &= C_x; \\ \mu_z \cdot \sqrt{\sum_{k=1}^N (a_{zk}^2 + b_{zk}^2)} &= C_z; \end{aligned} \quad (2.2)$$

де C_x, C_y, C_z – це значення квадратних коренів в цих же рівняннях, що відповідають першій відтвореній паролній фразі.

Для приведення даних до масштабу першої відтвореної паролної фрази досить помножити значення обчислених лінійних функціоналів на отримані з рівнянь (2.2) масштабні коефіцієнти.

Локальні динамічні параметри обчислюються так само, як і глобальні, з тією лише різницею, що період інтегрування береться рівним деякому локальному відрізку часу. Зокрема, стосовно до рис. 2.4, локальні динамічні параметри можуть бути обчислені на сьомі інтервалах торкання пера графічного планшета, між вісьмома відривами пера. Крім відривів пера від планшета для дроблення підпису на фрагменти можуть використовуватися перетини фрагментів траєкторії підпису, точки зміни напрямку руху, кутові точки букв, що мають розрив похідної [11, 12].

Зазвичай кількість виділених локальних ділянок підпису в першому наближенні виявляється пропорційною кількості букв в рукописно відтворюваному слові (коефіцієнт пропорційності знаходиться в межах від 1 до 2.4 в залежності від використаних алгоритмів фрагментування). На кожній з локальних ділянок обчислюється порівняно невелика кількість параметрів ($M = 2 \div 6$). Таким чином для $N = 16$ та $M = 4$ для слова з 5 букв треба розрахувати 216 динамічних параметрів, з яких 96 є глобальними коефіцієнтами ряду Фур'є, і 120 параметрів є локальними коефіцієнтами. Слід підкреслити, що співвідношення між глобальними і локальними динамічними параметрами має бути розумно збалансовано. Глобальні динамічні параметри описують схожість підпису зі зразком в цілому, а локальні динамічні параметри дозволяють оцінювати схожість відтворення кожної з букв підпису.

Біометричний еталон для цього класу систем формується на етапі навчання і повинен відображати як стабільну, так і нестабільну складові біометричних даних ідентифікованої особистості. Біометричний зразок підпису (слова-пароля) може формуватися кількома способами залежно від прийнятого в системі вирішального правила. Зокрема, якщо рішення в системі

приймається нейронною мережею, то роль біометричного зразка буде грати структура штучної нейронної мережі, вид і параметри функцій збудження, а також значення вагових коефіцієнтів нейронів. У разі, коли рішення приймається детермінованим вирішальним правилом, роль біометричного зразка можуть грати вектори значень математичних сподівань і дисперсій вимірюваних динамічних параметрів.

Якщо система аналізу підпису оцінює тільки глобальні параметри, проблем з формуванням біометричного зразка підпису не виникає, бо кількість оцінюваних параметрів незмінна. Інша справа – системи, що враховують локальні особливості відтворення підпису (слова-пароля). При несвідомому відтворенні підпису з певною частотою з'являються нові додаткові фрагменти підпису або, навпаки, зникають (зливаються) раніше розділені сусідні фрагменти. В результаті правило розбиття підпису на фрагменти виявляється неоднозначним. В такому разі при навчанні системи фіксується кілька варіантів підпису. Ця ситуація з неоднозначною фрагментацією підпису ілюстрована на рис. 2.7, деображено три часто найчастіших варіанти підпису конкретної людини. З наведеного рисунка видно, що I варіант розмітки підпису відрізняється від II варіанту меншим числом виділених локальних фрагментів. Алгоритм розмітки в I варіанті об'єднав в один 4-й фрагмент два фрагмента (4 і 5), присутніх в інших варіантах. Третій варіант розмітки відрізняється від попередніх варіантів появою нової 7-ї ділянки підпису.

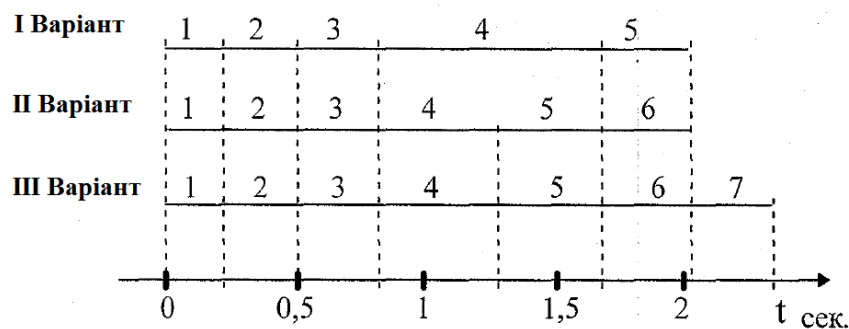


Рисунок 2.7 – Три варіанти фрагментації підпису

Ситуація з неоднозначною фрагментацією підпису можна віришити двома способами. Перший пов'язаний з формуванням декількох незалежних еталонів, при цьому для формування кожного еталона потрібно кілька підписів, що ускладнює процес навчання системи. Другий пов'язаний з приведенням всіх варіантів до деякого усередненого варіанту. Стосовно до рис. 2.7 як усереднений може використовуватися II варіант, тоді для приведення до нього I варіанта потрібно штучно вводити в нього додаткову точку розмітки підпису, розбиваючи ділянку 4 на 2 частини. Для приведення третього варіанту до другого досить відкинути останню 7-му ділянку фрагментації.

Однією з основних проблем систем ідентифікації цього класу є те, що їх параметри істотно залежать від психологічного стану людей і стабільності їх почерку. Для людей зі стабільним рукописним почерком ці системи працюють значно краще, ніж для середньостатистичної людини, але, в той же час, системи цього класу можуть бути дискредитовані людьми з нестабільним почерком. Крім того, введення біометричної авторизації та ідентифікації, сильно обмежують свободу дій користувачів в корпоративних мережах, перший час можуть спостерігатися спроби навмисної дискредитації біометричних систем через штучну симуляцію нестабільного почерку.

Підводячи підсумки, слід зазначити наступне.

1. На сьогоднішній день біометричні системи, побудовані на аналізі динаміки рукописного почерку, здатні надійно розрізняти близько 100 почерків при обліку двох координат (X, Y) і здатні розрізняти порядку 1000 почерків при врахуванні трьох координат (X, Y, Z) .

2. Вкрай важливим є те, що цей тип систем здатний зберігати в таємниці біометричний образ користувача (образ накреслення слова-пароля).

3. Смартфони мають порівняно низьку вартість, і немає пристроїв, здатних перехоплювати таємний образ у вигляді самого слова-пароля, відтвореного користувачем наосліп. Все це робить цей тип систем вкрай привабливим для використання в мобільних пристроях.

2.3. Аутентифікація за клавіатурним почерком.

Однією зі складних задач, повсякденно розв'язуваних багатьма людьми, є швидке введення текстів з клавіатури комп'ютера. Зазвичай швидкого клавіатурного введення інформації вдається досягти за рахунок використання всіх пальців обох рук, при цьому у кожної людини з'являється свій унікальний клавіатурний почерк.

Традиційно обмеження доступу до інформації здійснюється за паролями, однак якщо кількість символів в паролі збільшити, то з'являється можливість аналізувати динамічні параметри введення пароля, тобто фіксувати характерний для користувача клавіатурний почерк. Наприклад, при введенні паролі фрази біометрична система фіксує час натискання кожної клавіші і інтервал часу між натисканням чергової клавіші і відпусканням попередньої клавіші. Графік співвідношення інтервалів часів натискання і відпускання клавіш для слова «Пароль» наведено на рис. 2.8.

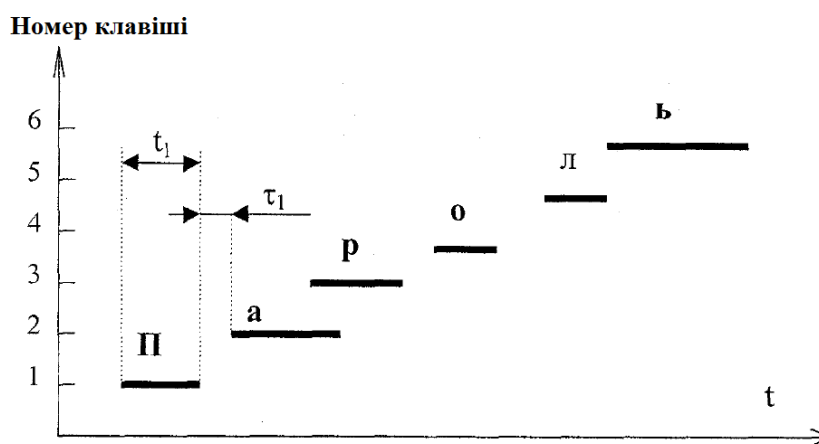


Рисунок 2.8 – Часова діаграма введення слова «Пароль»

З рис. 2.8 видно, що часи натискань клавіш $t_1, t_2, \dots, t_N$ різні і, відповідно, значення цих параметрів можуть бути використані для виявлення характерних особливостей індивідуального клавіатурного почерку користувача. Крім того, можуть бути використані, як контрольні параметри, інтервали часу між натисканням сусідніх клавіш $\tau_1, \tau_2, \dots, \tau_{N-1}$. Контрольні параметри t_k і τ_k істотно залежать від того, скільки пальців використовує при наборі користувач, від характерних для користувача поєднань рухів різних пальців руки і

від характерних рухів рук в процесі набору. Зокрема, якщо змусити користувачів працювати одним пальцем однієї руки, то клавіатурний почерк практично повністю втрачає індивідуальність. В цьому випадку часи натискання клавіш для різних людей перестають відбивати їх індивідуальність. Інтервали між натисканнями стають пропорційні відстані між клавішами, а перекриття натискань сусідніх клавіш стає неможливим (параметр τ_k завжди виявляється більше нуля). З іншого боку, по міру збільшення навичок роботи з клавіатурою і по міру переходу до сліпого набору всіма пальцями обох рук істотно зростає індивідуальність клавіатурного почерку кожного з користувачів.

У сучасних системах аутентифікації за клавіатурним почерком використовуються й інші параметри клавіатурного почерку [13]: загальний час набору паролі фрази, частота виникнення помилок при наборі, факт використання додаткових клавіш (використання числової клавіатури), особливості введення великих літер (використання клавіші Shift або CapsLock) і т.д.

Крім того, в роботі [14] автор розглядає наступні характеристики клавіатурного почерку:

- 1) кількість символів – розмір тексту без урахування символів, видалених за допомогою BackSpace чи Delete;
- 2) загальний час – розраховується від моменту натискання першої клавіші до моменту натискання останньої;
- 3) мінімальна пауза між натисканнями на клавіші;
- 4) максимальна пауза між натисканнями на клавіші;
- 5) середня пауза між натисканнями на клавіші;
- 6) середній час натискання на клавіші;
- 7) швидкість набору символів, розраховується як кількість набраних символів в хвилину;
- 8) чиста швидкість набору символів нетто, розраховується як кількість набраних символів в хвилину без урахування видалених за допомогою BackSpace чи Delete;
- 9) втрати від виправлень, розраховуються в процентному співвідно-

шенні і показують наскільки падає швидкість через час, витрачений на пошук помилок та їх виправлення;

10) ступінь аритмічності набору, характеризує ступінь нерівномірності набору в процентах, розраховується як середнє відхилення паузи між натисканнями від середнього значення;

11) кількість виправлень кількість символів, видалених за допомогою BackSpace чи Delete;

12) максимальний розмір фрагменту тексту без виправлень.

Алгоритми розпізнавання клавіатурного почерку можна розділити на три групи:

1) алгоритми, які аналізують почерк під час введення пароля;

2) алгоритми, які аналізують почерк після введення додаткового текстового фрагмента або фрази;

3) алгоритми, які постійно проводять прихований моніторинг клавіатурного почерку користувача.

Алгоритми першої групи забезпечують найбільшу швидкодію: користувачеві потрібно лише ввести свій пароль. Однак точність в цьому випадку невисока, особливо в разі короткого пароля.

Алгоритми другої групи дозволяють забезпечити більшу точність, в порівнянні з першою групою. Однак на введення додаткового фрагмента тексту потрібен час, що може викликати негативні емоції у користувача, особливо в разі, якщо йому часто доводиться проходити процедуру аутентифікації.

Алгоритми третьої групи можуть забезпечити високу точність. При цьому вони вимагають більше ресурсів. Перевагою цієї групи є можливість розпізнати злоумисника, який використовує комп'ютер, на якому раніше авторизувався користувач. В цьому випадку система може заблокувати комп'ютер для запобігання доступу до конфіденційної інформації.

Для алгоритмів другої та третьої груп аналізуються фрагменти фраз, що містять приблизно 10-40 символів. Ця кількість обумовлена «інтервалом

копіювання» користувача. Під цим терміном розуміється число символів, які можуть бути надруковані в точності після однократного перегляду тексту.

Для визначення ступеня відповідності двох зразків клавіатурного почерку існує ряд алгоритмів. Розглянемо деякі з них.

Ймовірісно-статистичний алгоритм.

У дослідженнях [15] встановлено, що час натискання клавіш (ЧНК) при наборі тексту має бімодальний розподіл. У другий нормальний розподіл елементи потрапляють у разі, якщо набір відбувався з перекриттям клавіш. Отже, необхідно розділяти ЧНК з перекриттям і без перекриттів. Таким чином, можливо отримати дві незалежні вибірки і дві величини математичного сподівання ЧНК – з перекриттями і без перекриттів. Такий підхід дозволяє зменшити кількість помилок 1-го і 2-го роду, збільшивши точність ідентифікації користувача. Кожна з двох вибірок підпорядковується нормальному закону розподілу.

Робота ймовірісно-статистичного алгоритму складається з двох етапів. Першим етапом є збір статистик характеристик клавіатурного почерку різних користувачів. Для цього проводиться ідентифікація оператора по його унікальному логіну. Далі заповнюється масив KeyEventsArr, поки не буде натиснуто достатню кількість клавіш. Структура масиву представлена на рис. 2.9.

Клавіша 1	Клавіша 2	...	Клавіша N
Подія KeyUp або KeyDown	Подія KeyUp або KeyDown	...	Подія KeyUp або KeyDown
Системний час події	Системний час події	...	Системний час події

Рисунок 2.9 – Структура масиву KeyEventsArr

Наступним етапом є підрахунок ЧНК. Для кожної події натискання клавіші знаходиться подія відпускання цієї клавіші. Величина ЧНК становить різницю між системним часом відпускання і системним часом натискання.

Це дозволяє отримати значення ЧНК в мілісекундах. Для зберігання характеристик клавіатурного почерку користувача використовується структура наступного вигляду (див. рис. 2.10).

Клавіша	ЧНК 1	ЧНК 2	Кількість 1	Кількість 2	Вибірка нормального розподілу 1	Вибірка нормального розподілу 2
А						
а						
Б						
б						
...						
я						

Рисунок 2.10 – Структура біометричного еталону

У полі «Вибірка» заносяться значення ЧНК, розділені символом «;». У полі «Кількість» зберігається кількість елементів у вибірці. При отриманні достатньої для аналізу кількості ЧНК, підраховується математичне сподівання ЧНК і заноситься у поле «ЧНК».

Наступним етапом роботи системи є аутентифікація користувача за клавіатурним почерком. У розглянутому алгоритмі порівняння почерків відбувається за принципом «один до одного», тобто завантажується еталон конкретного користувача, логін якого було введено. У разі успішної аутентифікації відбувається процес авторизації, в інакшому випадку користувач отримує відмову у доступі.

Для визначення клавіатурного почерку використовується порівняно невелика кількість клавіш – не більше 20. Для порівняння отриманого зразка клавіатурного почерку користувача з еталонним відбувається шляхом підрахунку міри Евкліда для кожної клавіші. При цьому ЧНК-1 і ЧНК-2 аналізуються як окремі елементи.

Для підрахунку міри Евкліда використовується формула:

$$P = \sqrt{\sum_{i=1}^N (A_i - B_i)^2}, \quad (2.3)$$

де P – міра несхожості; A – значення ЧНК-1 в мс; B – значення ЧНК-2 в мс; N – кількість елементів в аналізованій вибірці.

Підраховане значення міри несхожості порівнюється з порогом доступу. Якщо несхожість менше порога доступу, то вважається, що почерки збігаються і користувач пройшов аутентифікацію. Інакше аутентифікацію не пройдено.

Гістограмний метод.

Даний метод розглянуто в [16]. У гістограмному методі аутентифікація користувача відбувається шляхом аналізу клавіатурного почерку при введенні певної контрольної фрази. За результатами аналізу виноситься рішення про успішну аутентифікації або відбувається відмова в доступі.

В якості основних параметрів клавіатурного почерку використовується час натискання і час пауз між натисканнями. Нехай контрольна фраза містить q символів і час, витрачений користувачем на введення цієї фрази, становить T . Тоді необхідно проаналізувати r подій клавіатури:

$$r = q + p, \quad (2.4)$$

де q – кількість натискань клавіш; $p = q - 1$ – кількість пауз між натисканнями. Як приклад можна розглянути введення користувачем контрольної фрази з шести символів. В цьому випадку $r = 11, q = 6, p = 5$.

При введенні можливі перекриття натискань. Такий вид події клавіатури слід інтерпретувати як негативні значення тривалості пауз між натисканнями (див. рис. 2.8).

Часова розкладка процесу введення контрольної фрази індивідуальна для кожного користувача. Для кожного результату введення у відповідність ставиться r -мірний вектор клавіатурних параметрів

$$V = \{v_j\}, j = 1 \div r, \quad (2.5)$$

де кожен компонент v_j відповідає тривалості чергової події клавіатури: натисканню клавіші або паузі між натисканнями, що сталася за період часу T . Таким чином, вектор біометричних параметрів V можна інтерпретувати як зразок клавіатурного почерку користувача.

Для кожного користувача необхідно скласти навчальну вибірку зразків S -класу. Для цього використовується серія з L зразків клавіатурного почерку цього користувача.

$$\psi^{(s)} = \{v_i\}, i = 1 \div L. \quad (2.6)$$

Допускається наявність в системі множини $K = \{k_1, k_2, \dots k_M\}$ користувачів, для кожного з яких представлено власний еталон клавіатурного почерку. Даний еталон співвідноситься з одним з класів множини $S = \{s_1, s_2, \dots s_M\}$. Отже, сукупність користувачів системи $\{K\}$ однозначно відображається на множину класів $\{S\}$. Таким чином, необхідно M навчальних вибірок для того щоб сформувати еталони M легітимних користувачів:

Коли система працює в режимі аутентифікації, невідомий користувач повинен пред'явити їй зразок свого клавіатурного почерку. Почерк характеризується вектором біометричних параметрів $V^{(s_x)} = \{v_j\}, j = 1 \div r$. На основі отриманого вектора система формує еталонний опис невідомого x -класу, після чого порівнює його з еталонами всіх відомих системі $\{k_1, k_2, \dots k_M\}$ користувачів. В результаті порівняння виносяться рішення про особистість користувача. При цьому необхідно вирішити задачу класифікації вектора $V^{(s_x)}$ на $M + 1$ взаємовиключних класів. З них M класів відносяться до множини $S = \{s_1, s_2, \dots s_M\}$, тобто позначають відомих системі користувачів, а $M + 1$ клас відповідає за опис всіх інших користувачів, тобто зловмисників.

Таким чином, мета навчання системи розпізнавання клавіатурного почерку – це формування еталонних описів для класів, відповідним користувачам. Найчастіше, використовується вид вирішальних правил, запозичений з теорії статистичних рішень. В цьому випадку необхідно сформувати відно-

шення правдоподібності для умовних щільностей розподілу і порівняти з порогом C_n :

$$\frac{w_r(V|s_1)}{w_r(V|s_2)} \geq C_n, \quad (2.7)$$

де $w_r(V|s_j)$ – умовна -мірна щільність ймовірності вибірових значень $\{v_j\}$, $j = 1 \div r$ за умови їх приналежності до класу s_j .

Для апроксимації розподілу векторів можна використовувати гаусів розподіл. Однак необхідно враховувати, що на параметри клавіатурного почерку впливає психофізичний стан користувача, добові біоритми та інші фактори. Таким чином, для клавіатурного почерку характерні флуктуації параметрів, тому має сенс використовувати змішане гаусів розподіл з декількома центрами.

При цьому вирішальне правило має вигляд:

$$S = \begin{cases} S_c, & C \geq C_n; \\ S_x, & C < C_n; \end{cases} \quad (2.8)$$

де C_n – значення порога, яке вибирається з урахуванням необхідної точності розпізнавання.

Таким чином система визначає, чи відповідає пропонований зразок клавіатурного почерку користувача еталону. У разі успішної аутентифікації користувач може перейти до авторизації. В іншому випадку вважається, що користувач не аутентифікований.

Метод на основі нейронних мереж.

У роботі [17] розглядається можливість використання штучної нейронної мережі для аналізу клавіатурного почерку.

Використання нейромережевого підходу дозволяє досягти більшої точності в порівнянні з ймовірнісно-статистичними методами. Проте, необхідно враховувати дві групи принципових проблем. По-перше, до них відносяться власні проблеми штучних нейронних мереж, пов'язані з можливістю виникнення невизначено довгого процесу навчання, тупиків, стану «паралічу». По-друге, можуть виникнути проблеми, які визначаються біометричною приро-

дою дослідних образів, головна з яких навчання – на всіх можливих «чужих» користувачах.

Для аналізу характеристик клавіатурного почерку в якості первинних даних пропонується використовувати два вектора, один з яких містить значення тривалості натискань окремих клавіш, а другий описує інтервали між натисканнями.

На основі описаних часових рядів формується фазова траєкторія (крива в фазовому просторі, складена з точок, що представляють стан динамічної системи в послідовні моменти часу протягом усього часу еволюції) на площині, утворена двома осями: віссю часу натискання клавіш і віссю часових інтервалів між двома сусідніми натисканнями. Дана фазова траєкторія аналізується за допомогою нейронної мережі, що самостійно організується. Навчання нейронної мережі пропонується проводити на основі алгоритму нейронного газу.

Зазначена нейронна мережа є одношаровою, і в ній кожен з нейронів з'єднаний з кожним елементом вхідного вектора. Після початкового визначення вагових коефіцієнтів розраховується відстань за Евклідом між вхідним вектором і векторами вагових коефіцієнтів, що належать відповідним нейронам. На кожній ітерації роботи алгоритму нейрони сортуються залежно від їх відстані до вхідного вектора. Таким чином, після сортування нейрони розташовуються в порядку, відповідному збільшенню віддаленості.

$$d_0 < d_1 < d_2 < \dots < d_{n-1}, \quad (2.9)$$

де d_i – віддаленість i -го нейрона, що займає в результаті сортування m -ту позицію в послідовності, яку очолює нейрон-переможець; d_0 – віддаленість нейрона-переможця.

Переможцем визнається той нейрон, якому відповідає вектор вагових коефіцієнтів з найменшим евклідовою відстанню до вхідного вектора. Значення функції сусідства для i -го нейрона визначається відповідно до формули:

$$G(x, i) = \exp\left(-\frac{m(i)}{\lambda}\right), \quad (2.10)$$

де $G(x, i)$ – значення функції сусідства; i – номер нейрона; $m(i)$ – черговість, отримана в результаті сортування; λ – параметр, значення якого з часом зменшується.

При $\lambda = 0$ адаптації піддається тільки нейрон-переможець, але при $\lambda \neq 0$ уточненню підлягають ваги багатьох нейронів, при цьому рівень уточнення залежить від величини $G(x, i)$.

Нейрон-переможець і все нейрони, що лежать в межах його околиці, піддаються адаптації, в ході якої їх вектори вагових коефіцієнтів змінюються за правилом Кохонена.

$$w_i(k+1) = w_i(k) + \eta_i(k)[x - w_i(k)], \quad (2.11)$$

де $\eta_i(k)$ – коефіцієнт навчання.

Описаний алгоритм аналізу параметрів клавіатурного почерку дозволяє коректно обробити до 95% поданих на вхід векторів.

Підводячи підсумки, слід зазначити наступне.

Аутентифікації за допомогою клавіатурного почерку притаманний ряд переваг і недоліків. До переваг слід віднести:

- 1) стабільність клавіатурного почерку конкретного користувача, що дозволяє з ймовірністю не менш 90% ідентифікувати користувача за його клавіатурним почерком;
- 2) відносна дешевизна впровадження системи розпізнавання;
- 3) можливість контролювати як доступ до ресурсів, так і фізичний стан співробітника.

До недоліків методу можна віднести:

- 1) застосування можливо тільки для розпізнавання користувачів зі сформованим клавіатурним почерком.

2.4. Голосова аутентифікація.

Ідентифікація особистості за особливостями голосу має ряд привабливих сторін. По-перше, існує добре розвинена телефонна мережа, по-друге, звукові карти фактично стали стандартним обладнанням сучасних персональних комп'ютерів. Крім того, ідентифікація за голосом є традиційною для людей і не викликає психологічної неприязні. Для сліпих ідентифікація за голосом є основним способом впізнавання особистості.

Перші системи ідентифікації особистості за особливостями голосу будувалися, виходячи з частотних уявлень і можливостей засобів аналогової фільтрації [18, 19]. В основу їх роботи покладено різне тембральне забарвлення голосів і індивідуальна нерівномірність розподілу потужності вимовної фрази за частотним спектром. Базовими процедурами для цього класу пристроїв є вузько смугова фільтрація сигналу і відновлення його обвідної. Зокрема, подібна система фірми Texas Instruments [20] використовує гребінку з 16 вузькосмугових фільтрів з шириною смуги пропускання 220 Гц, що рівномірно покриває частотний діапазон від 300 до 3000 Гц. Структура аналогової частини системи голосової ідентифікації представлено на рис. 2.11.

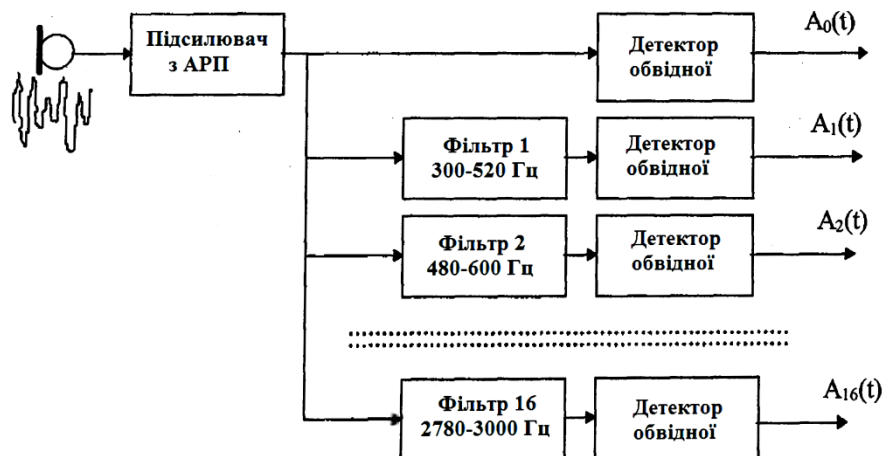


Рисунок 2.11 – Голосова ідентифікація за допомогою багатоканального аналізу розподілу енергії сигналу за частотним спектром

В процесі вимови контрольної фрази система ідентифікації масштабує сигнал за рахунок роботи АРП вхідного підсилювача. Смугові фільтри і детектори обвідної їх відгуків дозволяють отримати 16 функцій часу

$A_1(t), A_2(t), \dots, A_{16}(t)$, що характеризують розподіл енергії звукового сигналу за частотним спектром. Функція $A_0(t)$ характеризує розподіл енергії повного сигналу у всьому діапазоні звукових частот. В процесі навчання система запам'ятовує найбільш ймовірні еталонні значення функцій $A_k(t)$ для конкретної людини, а також допустимі коридори відхилень для цих функцій.

Однією з проблем, що виникають при ідентифікації особистості за голосом, є те, що частина ділянок $A_k(t)$, що відповідають шиплячим звукам, не приносять користі. Більш того, ці фрагменти функцій $A_k(t)$ повинні обов'язково бути виключені з аналізу, оскільки їх врахування може тільки погіршити якість ідентифікації.

Про надходження на вхід пристрою звукової фонемі судять за наявністю великого рівня відгуку каналу $A_0(t)$, при цьому, якщо на вході пристрою присутній шиплячий звук, то його спектр виявляється рівномірним або ж $A_k(t) \cong A_0(t)/16$. Шиплячі звуки є модульованим по амплітуді «білим шумом» і легко розпізнаються. Всі інші звуки (які не є шиплячими) обов'язково мають яскраво виражену нерівномірність спектральної характеристики, і на цю нерівномірність істотно впливають індивідуальні особливості м'язової активності мовного тракту людини.

Ймовірність присутності характерних індивідуальних особливостей голосу людини у 18 фонемах української мови приведено в табл. 2.1 [20]. Найбільш інформативними є фонemi першої частини таблиці. У другій частині таблиці поміщені фонemi, непридатні для цілей ідентифікації особистості, бо вони дозволяють ідентифікувати диктора з ймовірністю 0.5 і менше. Ці фонemi можуть відображати особливості голосу особистості тільки в поєднанні з іншими звуками.

Дуже важливою задачею, що розв'язується в процесі ідентифікації особистості за голосом, є нормування функцій $A_k(t)$ за часом. Як правило, при проголошенні парольної фрази тривалості звуків і пауз між ними можуть варіюватися в межах від 10 % до 50 %. Наразі існує два способи компенсації часової нестабільності проголошення диктором пральних фраз. Перший спо-

сіб зводиться до масштабування під еталон шляхом стиснення або розтягування ділянок, відповідних окремим звукам, засобами динамічного програмування. Другий підхід зводиться до виділення центру звукової області та ідентифікаційним вимірам в околицях центральної частини фонем, тоді абсолютні значення тривалостей фонем і пауз між ними не відіграють істотної ролі.

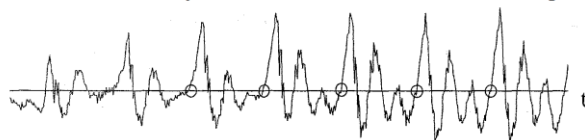
Таблиця 2.1 – Імовірність розпізнавання людини
по одній ізольованій фонемі

Фонема	Ймовірність	Фонема	Ймовірність	Фонема	Ймовірність
Е	0.9	Р	0.78	Ч	0.54
О	0.86	В	0.76	Ц	0.50
Л	0.84	Ж	0.74	Х	0.48
А	0.83	М	0.62	С	0.44
І	0.83	Г	0.61	Ш	0.37
З	0.79	У	0.60	К	0.30

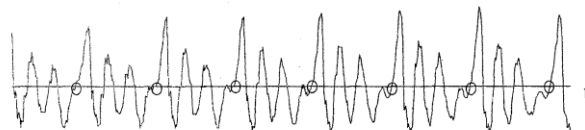
Описаний вище частотний підхід до ідентифікації особистості міг бути реалізований засобами аналогової фільтрації вже 30-40 років тому і саме з цієї причини в той час спостерігався сплеск інтересу до систем голосової ідентифікації. По мірі розвитку засобів обчислювальної техніки і методів цифрової фільтрації інтерес з частотних методів ідентифікації заміщається до систем, що використовують моделі лінійного передбачення мови.

В основу кодування мови методом лінійного передбачення покладено хвильову структуру мовного сигналу, яка особливо добре спостерігається при вимові голосних. На рис. 2.12 представлено хвильову структуру голосного звуку, де видно, що сусідні хвилі хвильової пачки досить схожі один на одну. Метод лінійного передбачення побудований на апроксимації сусідніх хвиль в звуковій пачці перехідним процесом деякого лінійного цифрового фільтру.

Початок пачки звукових хвиль після шиплючої фонемі



Центральна частина хвильової пачки



Кінець хвильової пачки голосної фонемі

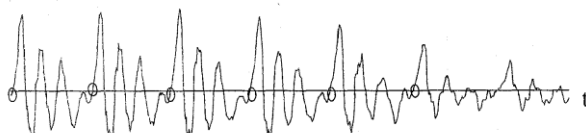


Рисунок 2.12 – Типовий приклад пачки звукових хвиль,
що утворюють голосний звук

При описі звукового сигналу методом лінійного передбачення вихідний акустичний сигнал розбивається на окремі інтервали аналізу фіксованої довжини (зазвичай довжина інтервалу аналізу становить 20 мс). Далі здійснюють визначення типу звуку всередині інтервалу аналізу (шум або тональний звук). Якщо всередині інтервалу знаходиться шумова фонема, то визначаються тільки його енергетичні параметри $A_0(t)$. Якщо всередині інтервалу аналізу присутній тональний фрагмент, то сигнал додатково описують шляхом розрахунку коефіцієнтів лінійного передбачення (лінійного цифрового фільтра) і розрахунку періоду імпульсів основного тону, що збуджують перехідні процеси на виході лінійного фільтра передбачення. Структура формування такого типу опису звукового сигналу відображена на рис. 2.13.

За своєю суттю лінійний фільтр передбачення є лінійним динамічним колом, імпульсна характеристика якого $h(t)$ задається вектором коефіцієнтів $\{a_1, a_2, a_3, \dots, a_k\}$. Коефіцієнти лінійного фільтра передбачення можуть бути знайдені різними способами. Зокрема, якщо в якості такого фільтра використовується рекурсивний цифровий фільтр, то його коефіцієнти $\{a_1, a_2, a_3, \dots, a_k\}$ можуть бути знайдені рішенням системи лінійних кореля-

ційних рівнянь. Цей підхід до визначення коефіцієнтів лінійного передбачення зручний тим, що він потребує обчислення малого числа коефіцієнтів, а також не вимагає попередньої класифікації тон/шум і знання періоду основного тону T_0 . Задачі класифікації тон/шум і пошук функції $T_0(t)$ можуть бути вирішені пізніше з використанням помилки лінійного передбачення. Основним недоліком використання лінійних фільтрів передбачення з рекурсивною структурою є потенційна нестійкість через наявність зворотного зв'язку.

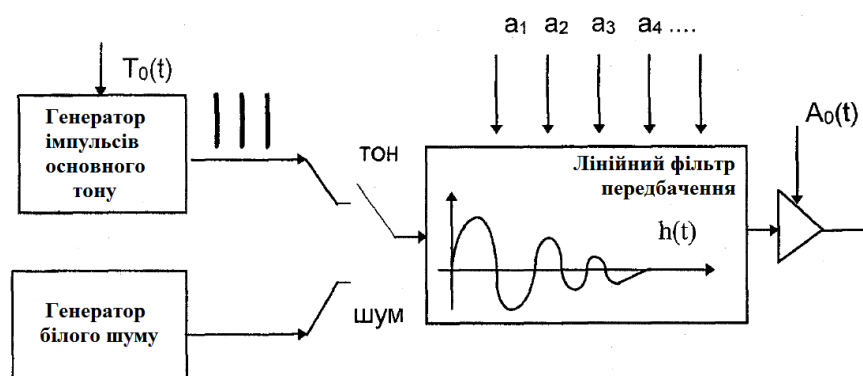


Рисунок 2.13 – Структура опису мовного сигналу методом лінійного передбачення

Ще одним підходом до вирішення задачі голосової ідентифікації є використання стійкого лінійного фільтру передбачення з нерекурсивною структурою. В цьому випадку його коефіцієнти $\{a_1, a_2, a_3, \dots, a_k\}$ повторюють імпульсну характеристику $h(t)$, але для точного опису цієї характеристики число коефіцієнтів істотно збільшується. Формально вектор коефіцієнтів фільтру може бути отриманий безпосередньо з параметрів звукової хвилі без рішення системи лінійних рівнянь, однак для цього потрібно знати період імпульсів основного тону і моменти приходу імпульсів основного тону на вхід фільтру. У першому наближенні моменти часу приходу імпульсів збудження основного тону можуть бути знайдені як лівий перетин нульового рівня максимальним по модулю імпульсом (див. рис. 2.12).

В останні роки з'явилися ефективні нейромережеві засоби визначення ділянок звук/пауза і класифікації тон/шум. Крім того, очевидно, то після нормування за амплітудою і часом ШНМ здатні з високою точністю вирішувати

ти задачу порівняння форми акустичного сигналу з еталоном. Наразі єдиною проблемою є те, висока точність ідентифікації вимагає складних і великих за структурою ШНМ, які доводиться занадто довго вчити на великій кількості біометричних зразків.

Незалежно від типу використовуваних процедур попередньої обробки, після сегментації парольної фрази на послідовні інтервали аналізу та їх обробки отримують опис особливостей голосу особистості у вигляді характерних функцій $T_0(t)$, $A_0(t)$ і системи функцій коефіцієнтів лінійного фільтру передбачення $\{a_1, a_2, a_3, \dots, a_k\}$.

До недоліків біометричних систем ідентифікації особистості за голосом можна віднести, перш за все те, що парольну фразу важко зберегти в таємниці. Сучасні засоби акустичного прослуховування дозволяють досить успішно здійснювати несанкціоноване копіювання парольної фрази. Таким чином, системи голосової ідентифікації мають більший у порівнянні з системи аутентифікації за динамікою почерку рівень ймовірності помилок другого роду – таємниця парольної фрази виявляється багато сильніше унікальності біометричних параметрів.

2.5. Аутентифікація за динамікою системи «користувач – комп'ютерна миша».

Динаміка системи «користувач – комп'ютерна миша» активно досліджується для поведінкової аутентифікації [21, 22]. З тих пір як у [23] вперше було опубліковано результати досліджень з питань можливості розрізнення користувачів за стилем їх роботи з мишею, було запропоновано кілька підходів до аутентифікації користувачів на основі динаміки системи «користувач – комп'ютерна миша» [24].

У [25] представлено схему входу в систему, засновану на рухах миші, з використанням двохетапного підходу, при якому фази реєстрації і тестування включають переміщення курсору миші між парами точок, що послідовно ві-

дображаються на екрані. В експерименті брало участь 15 учасників в контрольованому середовищі, EER склала 15 %, а час аутентифікації – 20 секунд.

У [26] описано метод входу в систему для доступу до комп'ютерних систем на основі рухів миші в неконтрольованому середовищі. В експерименті брало участь 28 користувачів. Учасникам дослідження було запропоновано зіграти в лабіринт з наперед заданим шляхом. Помилка EER склала 26.8 %, час аутентифікації невідомий.

У [27] представлено статичну систему аутентифікації для Web-додатків, засновану на рухах миші, коли користувач вводить PIN-код. У запропонованій схемі система розміщала віртуальну клавіатуру на екрані і вимагала, щоб користувачі використовували тільки мишу для введення пари логін-пароль. В експерименті брало участь 50 користувачів, помилка EER склала 6.2 % на основі 15 рухів миші.

Нещодавно у [28] було представлено систему статичної аутентифікації на основі динаміки миші «Mouse-Lock». В цій системі користувач проходив аутентифікацію за допомогою комбінованого графічного інтерфейсу у вигляді тесту CAPTCHA з іконками, розташованими по колу. В експерименті брало участь 6 користувачів, помилки FAR та FRR склали відповідно 3.5 % і 4 %.

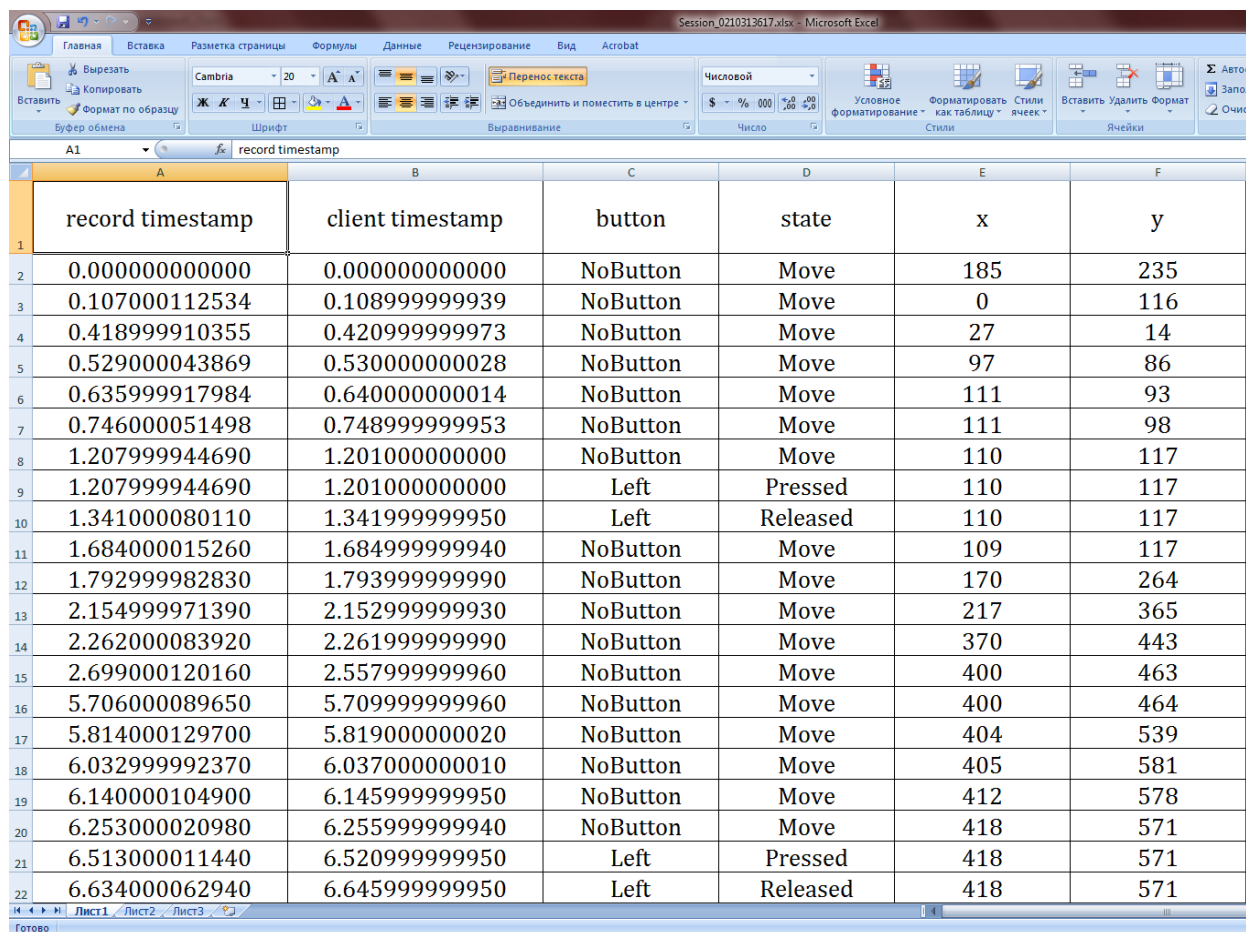
Переваги ідентифікації за динамікою комп'ютерної миші співпадають з такими для ідентифікації за клавіатурним почерком. Основним недоліком ідентифікації за динамікою курсору комп'ютерної миші є низька точність і, отже, вимога до використання цього методу тільки у мультибіометричних системах.

Підводячи підсумки можна стверджувати, що одночасне виконання вимог по безперервності й зручності використання в комп'ютерних мережах не може забезпечити ніякий спосіб ідентифікації, крім динамічного.

3 ДОСЛІДЖЕННЯ ІНФОРМАТИВНИХ ПАРАМЕТРІВ ДИНАМІКИ СИСТЕМИ «КОРИСТУВАЧ-КОМП'ЮТЕРНА МИША» ДЛЯ ЗАДАЧ БІОМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ

3.1. The Balabit Mouse Challenge Data Set.

База даних «Balabit Mouse Challenge Data Set» [29] була опублікована у серпні 2016 р. Датасет містить дві папки, перша – для навчальних файлів, друга – для тестових файлів. Дані були зібрані від 10 користувачів, які працювали за персональними комп'ютерами, підключеними через Internet до віддалених серверів. Навчальні та тестові дані представлені як сесії, проте тестові дані значно коротші, ніж навчальні. Тестова частина набору даних містить як позитивні (легальні), так і негативні (незаконні) сесії.



	A	B	C	D	E	F
	record timestamp	client timestamp	button	state	x	y
1	0.000000000000	0.000000000000	NoButton	Move	185	235
2	0.107000112534	0.108999999939	NoButton	Move	0	116
3	0.418999910355	0.420999999973	NoButton	Move	27	14
4	0.529000043869	0.530000000028	NoButton	Move	97	86
5	0.635999917984	0.640000000014	NoButton	Move	111	93
6	0.746000051498	0.748999999953	NoButton	Move	111	98
7	1.207999944690	1.201000000000	NoButton	Move	110	117
8	1.207999944690	1.201000000000	Left	Pressed	110	117
9	1.341000080110	1.341999999950	Left	Released	110	117
10	1.684000015260	1.684999999940	NoButton	Move	109	117
11	1.792999982830	1.793999999990	NoButton	Move	170	264
12	2.154999971390	2.152999999930	NoButton	Move	217	365
13	2.262000083920	2.261999999990	NoButton	Move	370	443
14	2.699000120160	2.557999999960	NoButton	Move	400	463
15	5.706000089650	5.709999999960	NoButton	Move	400	464
16	5.814000129700	5.819000000020	NoButton	Move	404	539
17	6.032999992370	6.037000000010	NoButton	Move	405	581
18	6.140000104900	6.145999999950	NoButton	Move	412	578
19	6.253000020980	6.255999999940	NoButton	Move	418	571
20	6.513000011440	6.520999999950	Left	Pressed	418	571
21	6.634000062940	6.645999999950	Left	Released	418	571

Рисунок 3.1 – База даних «Balabit Mouse Dynamics Challenge»

Параметри взаємодії кожного користувача з мишею записано в txt-файл з 6 стовпчиками: «record timestamp» – час (в секундах), що пройшов від початку моніторингу сесії, записаний на віддаленому сервері; «client timestamp» – час (в секундах), що пройшов від початку моніторингу сесії, записаний на клієнтському комп'ютері; «button» – поточний стан клавіші миші (використовується чи ні); «state» – поточний стан курсору миші (переміщується чи ні); «x» – поточна x координата курсору миші; «y» – поточна y координата курсору миші (рис. 3.1).

Автори проекту [30] розмістили у вільному доступі програмне забезпечення та базу даних з 40 параметрів динаміки системи «користувач-комп'ютерна миша», що була отримана шляхом обробки датасету «Balabit Mouse Challenge Data Set» (див. табл. 3.1).

Таблиця 3.1 – Інформативні параметри динаміки системи «користувач-комп'ютерна миша» за даними «Balabit Mouse Challenge Data Set»

Інформативний параметр (згідно [30])	Описання
type_of_action	Оскільки особливості роботи з комп'ютерною мишею описують як характеристики взаємодії з маніпулятором на рівні інтерфейсу (переміщення курсору), так і взаємодії з маніпулятором як з фізичним об'єктом (тривалість натискання клавіші миші), то усі дані в датасеті розділено на три типи дії миші: mouse movement (MM), point click (PC), drag-and-drop (DD). MM описує параметри переміщення курсору між двома точками екрану. PC описує часові параметри натискання клавіші миші. Це дуже схоже на дію MM, але завжди закінчується кліком клавіші миші. DD описує параметри дії перетягування об'єкта на екрані. Дія DD завжди починається натисканням лівої клавіші миші, далі послідовністю подій переміщення і закінчується відпусканням лівої клавіші.
travelled_distance_in pixels	Кількість взаємодій користувача з мишею в залежності від пройденої курсором відстані на екрані монітору впродовж даної дії миші
elapsed_time	Час даної дії миші
direction of movement	Один з восьми квадрантів (див. рис. 3.2), якому належить кут між віссю абсцис та відрізком, що з'єднує дві кінцеві точки траєкторії

Продовження таблиці 3.1

dist_end_to_end_line	Довжина відрізка між двома кінцевими точками траєкторії: $L = \sqrt{(x_n - x_1)^2 + (y_n - y_1)^2},$ де x_i, y_i – декартові координати курсору (див. рис. 3.3)
dist_length	Довжина траєкторії курсору мишки впродовж даної дії миші: $r = \sum_{i=2}^n \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2}.$
straightness	Прямолінійність траєкторії руху курсору миші: $Str = L/r.$
largest_deviation	Найбільша відстань від траєкторії до відрізка, що з'єднує дві кінцеві точки траєкторії
num_points	Кількість взаємодій користувача з мишею, що містяться в даній дії миші
sum_of_angles	Кількість кутів, що формують траєкторію курсору миші
mean_vx	Середнє значення горизонтальної швидкості курсору мишки впродовж даної дії миші: $V_{x_i} = \frac{x_i - x_{i-1}}{t_i - t_{i-1}}, \quad i = 2 \dots n, \quad t_1 = 0, \quad V_{x_1} = 0.$
sd_vx	Середньоквадратичне відхилення горизонтальної швидкості курсору мишки впродовж даної дії миші
max_vx	Максимальне значення горизонтальної швидкості курсору мишки впродовж даної дії миші
min_vx	Мінімальне значення горизонтальної швидкості курсору мишки впродовж даної дії миші
mean_vy	Середнє значення вертикальної швидкості курсору мишки впродовж даної дії миші: $V_{y_i} = \frac{y_i - y_{i-1}}{t_i - t_{i-1}}, \quad i = 2 \dots n, \quad t_1 = 0, \quad V_{y_1} = 0.$
sd_vy	Середньоквадратичне відхилення вертикальної швидкості курсору мишки впродовж даної дії миші
max_vy	Максимальне значення вертикальної швидкості курсору мишки впродовж даної дії миші
min_vy	Мінімальне значення вертикальної швидкості курсору мишки впродовж даної дії миші
mean_v	Середнє значення загальної швидкості курсору мишки впродовж даної дії миші: $V_i = \sqrt{V_{x_i}^2 + V_{y_i}^2}, \quad i = 2 \dots n, \quad V_1 = 0.$
sd_v	Середньоквадратичне відхилення загальної швидкості курсору мишки впродовж даної дії миші
max_v	Максимальне значення загальної швидкості курсору мишки впродовж даної дії миші
min_v	Мінімальне значення загальної швидкості курсору мишки впродовж даної дії миші

Продовження таблиці 3.1

mean_a	Середнє значення прискорення курсору мишки впродовж даної дії миші: $a_i = \frac{V_i - V_{i-1}}{t_i - t_{i-1}}, \quad i = 2 \dots n, \quad a_1 = 0.$
sd_a	Середньоквадратичне відхилення прискорення курсору мишки впродовж даної дії миші
max_a	Максимальне значення прискорення курсору мишки впродовж даної дії миші
min_a	Мінімальне значення прискорення курсору мишки впродовж даної дії миші
mean_jerk	Середнє значення швидкості зміни прискорення (ривку) курсору мишки впродовж даної дії миші: $j_i = \frac{a_i - a_{i-1}}{t_i - t_{i-1}}, \quad i = 2 \dots n, \quad j_1 = 0.$
sd_jerk	Середньоквадратичне відхилення швидкості зміни прискорення (ривку) курсору мишки впродовж даної дії миші
max_jerk	Максимальне значення швидкості зміни прискорення (ривку) курсору мишки впродовж даної дії миші
min_jerk	Мінімальне значення швидкості зміни прискорення (ривку) курсору мишки впродовж даної дії миші
mean_omega	Середнє значення кутової швидкості курсору мишки впродовж даної дії миші: $\omega_i = \frac{\Theta_i - \Theta_{i-1}}{t_i - t_{i-1}}, \quad i = 2 \dots n, \quad \omega_1 = 0,$ $\Theta_i = \arctg2\left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}}\right), \quad i = 2 \dots n; \quad \Theta_0 = 0,$ функція arctg2 – варіант тригонометричної функції arctg, що видає результати в діапазоні $[-\pi, \pi]$
sd_omega	Середньоквадратичне відхилення кутової швидкості курсору мишки впродовж даної дії миші
max_omega	Максимальне значення кутової швидкості курсору мишки впродовж даної дії миші
min_omega	Мінімальне значення кутової швидкості курсору мишки впродовж даної дії миші
mean_curv	Середнє значення кривизни траєкторії курсору мишки впродовж даної дії миші: $c_i = \frac{\Theta_i - \Theta_{i-1}}{s_i - s_{i-1}}, \quad i = 2 \dots n, \quad c_0 = 0, \quad s_1 = 0,$ де s_i – довжина траєкторії від початкової точки дії до i-ї точки $s_i = \sum_{j=1}^i \sqrt{(x_j - x_{j-1})^2 + (y_j - y_{j-1})^2}$

Продовження таблиці 3.1

sd_curv	Середньоквадратичне відхилення кривизни траєкторії курсору мишки впродовж даної дії миші
max_curv	Максимальне значення кривизни траєкторії курсору мишки впродовж даної дії миші
min_curv	Мінімальне значення кривизни траєкторії курсору мишки впродовж даної дії миші
num_critical_points	Кількість критичних точок на траєкторії руху курсору мишки впродовж даної дії миші
a_beg_time	Прискорення курсору мишки впродовж від початку руху до першої точки часу, де прискорення стає від'ємним

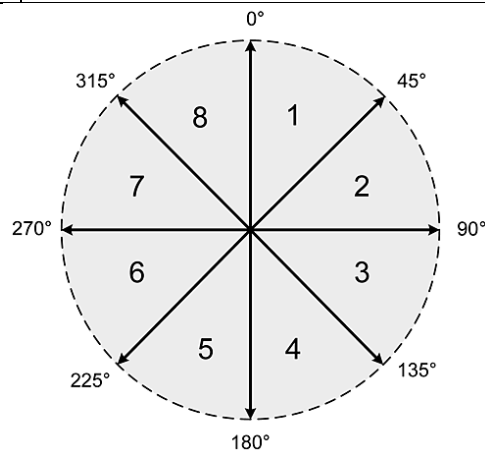


Рисунок 3.2 – Кутовий простір напрямків руху курсору

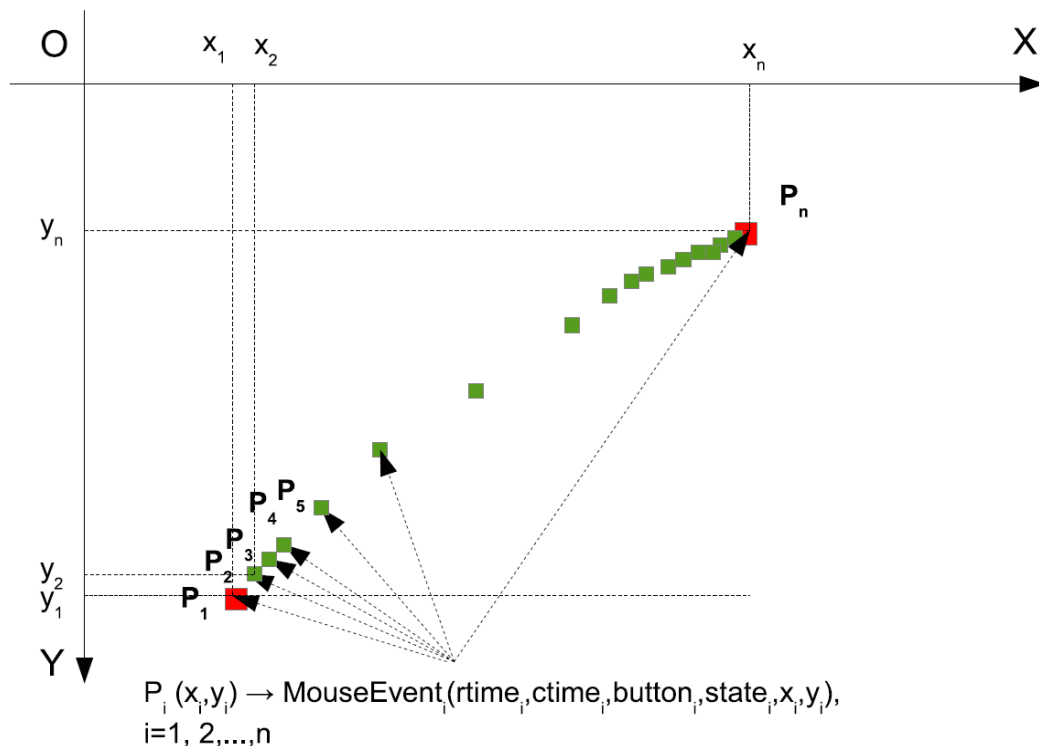
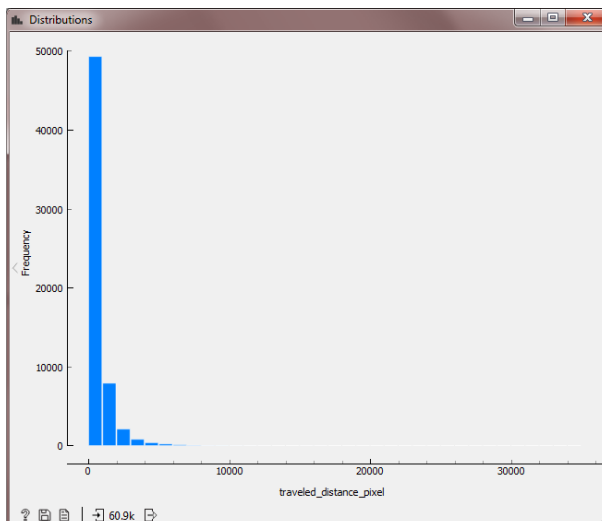


Рисунок 3.3 – Динаміка курсору комп'ютерної миші,

$\{P_1, P_2, \dots, P_n\}$ – індивідуальні події миші

Розподіл інформативних ознак в навчальній і тестовій частинах бази даних наведено на рис. 3.4. Як можна побачити, розподіл ознак у навчальній та тестовій частинах досить подібний. Гістограма пройденої відстані показує частоту дій в різних діапазонах відстаней. Зазвичай користувачі здійснювали дії з короткими пройденими відстанями частіше, ніж з довгими. На гістограмі прямолінійності траєкторії руху курсору видно, що існує велика кількість відносно простих дій як в тестовій, так і в навчальній частинах набору даних. Це також підтверджується гістограмою кривизни траєкторії курсору дій. Гістограма швидкостей показує, що користувачі в основному використовують рухи з низькою швидкістю руху курсору. Гістограми кутової швидкості ілюструють тенденцію користувачів використовувати в основному лінійні рухи курсору миші. Гістограма відстані між двома кінцевими точками аналогічна гістограмі пройденої відстані. Користувачі здійснюють руху миші, маючи невелику кількість критичних точок. Це відповідає гістограмам прямолінійності та кривизни.

Balabit_features_training



Balabit_features_test

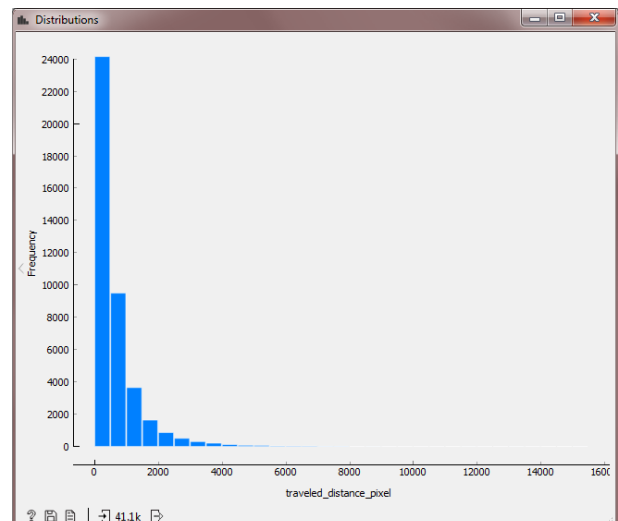
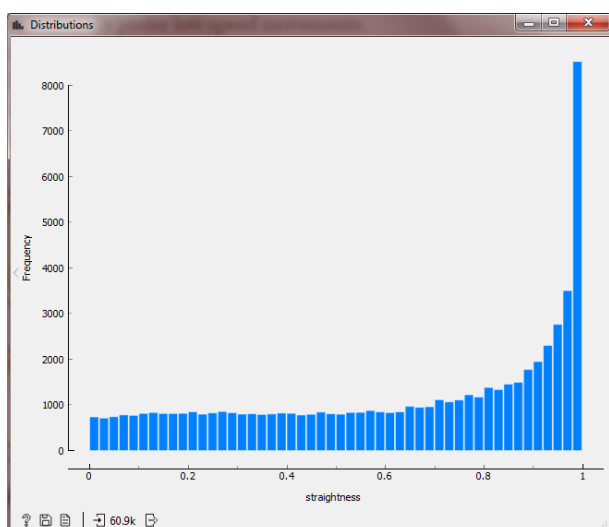
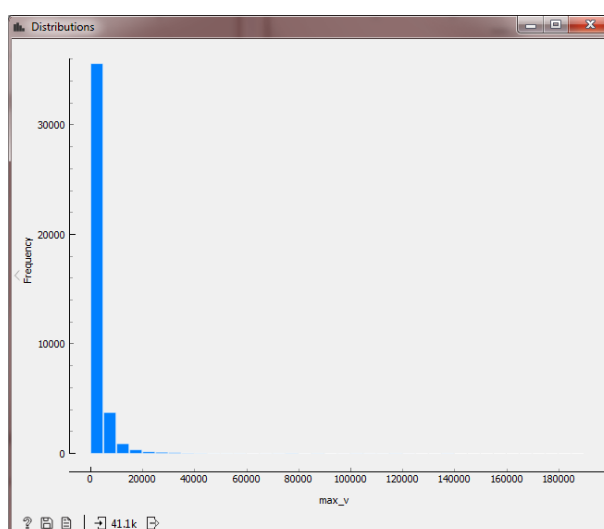
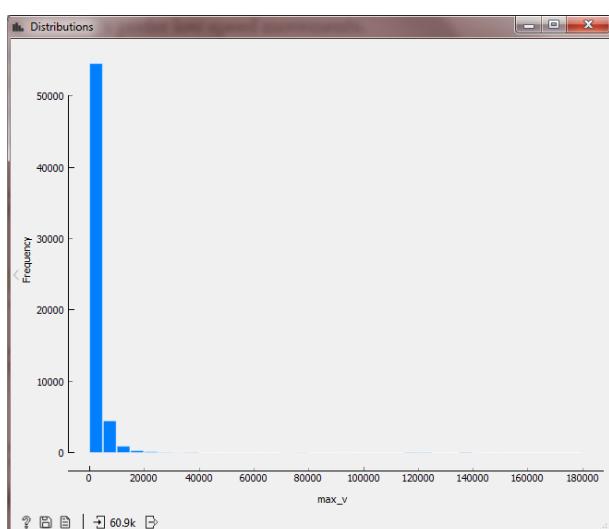
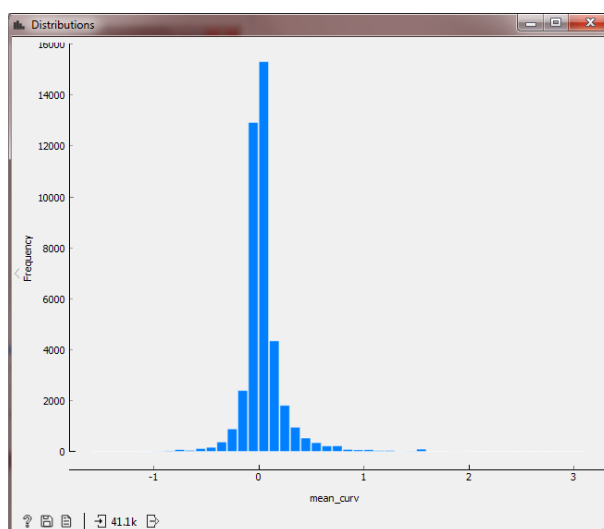
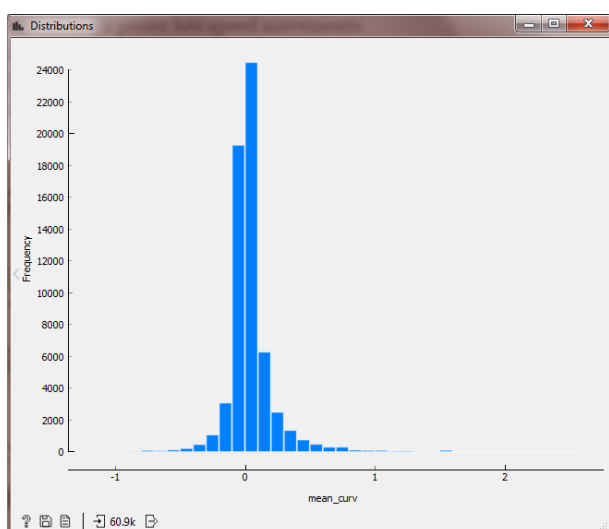
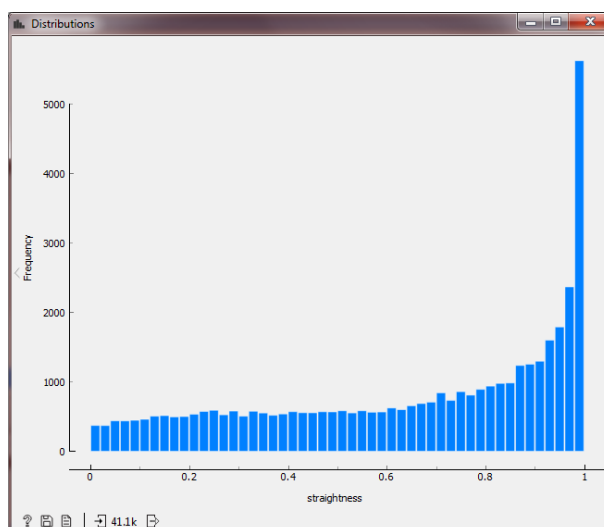


Рисунок 3.4 – Розподіл деяких інформативних параметрів курсорного почерку у навчальній та тестовій частинах бази даних «Balabit Mouse Challenge Data Set»

Balabit_features_training

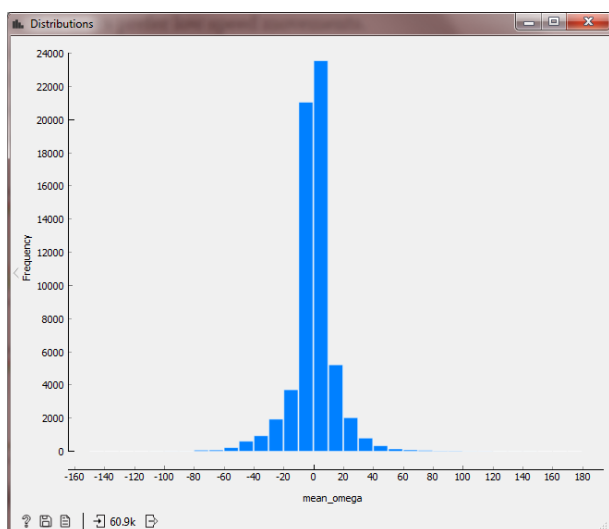


Balabit_features_test

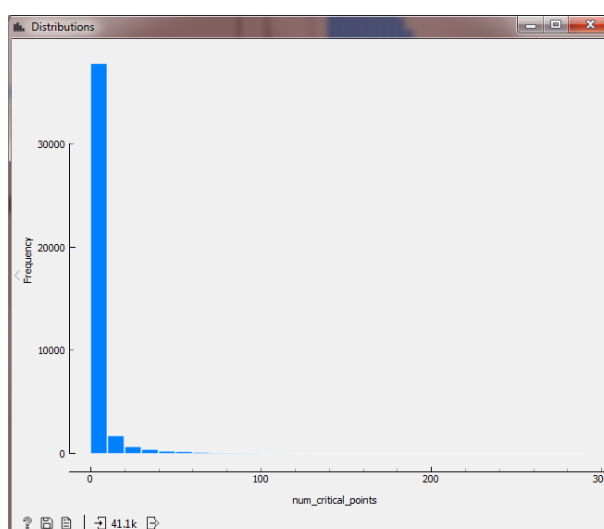
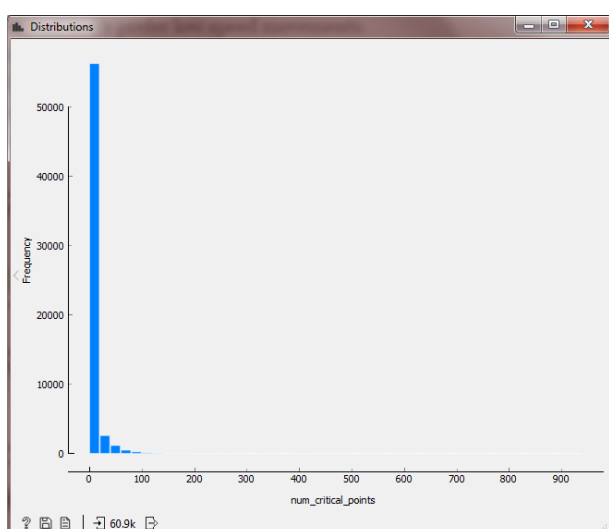
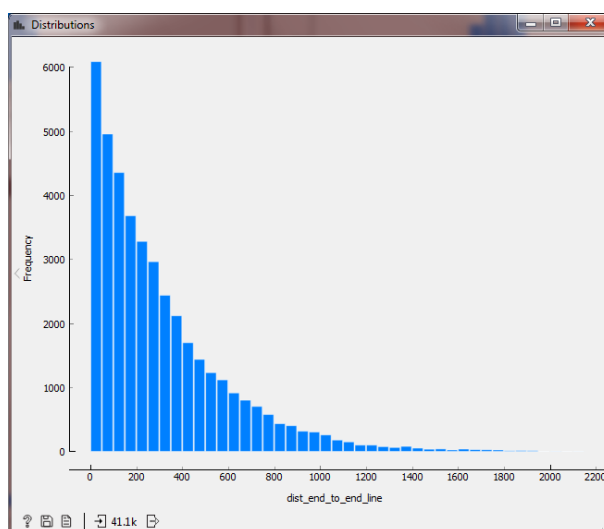
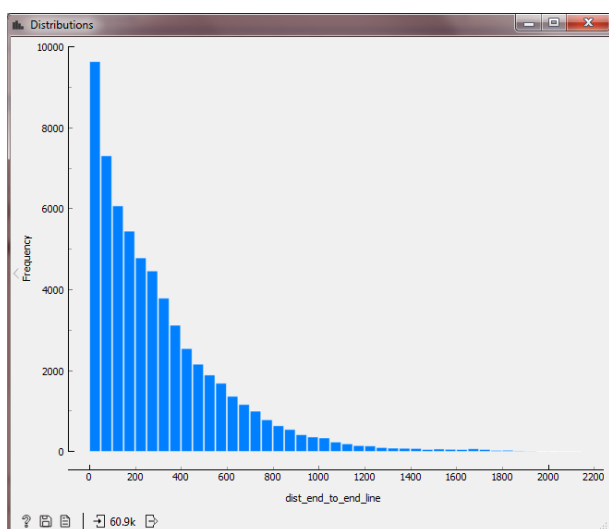
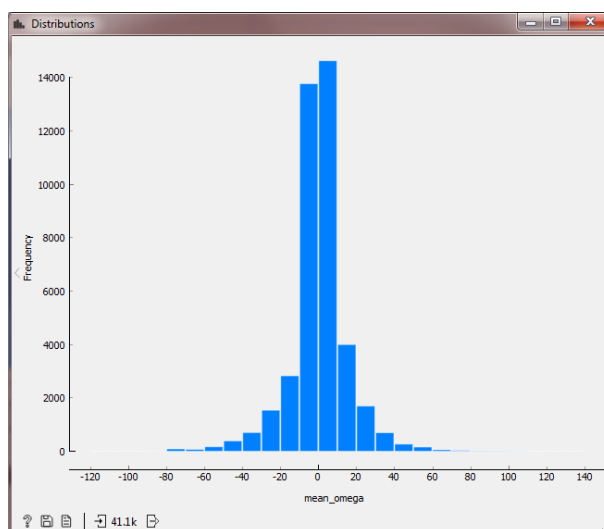


Продовження рисунку 3.4

Balabit_features_training



Balabit_features_test



Продовження рисунку 3.4

3.2. Схема експерименту та результати проведених досліджень.

Класифікація користувачів за параметрами динаміки системи «користувач – комп'ютерна миша» проводилась у програмному середовищі Orange.

Розробником даної програми є найстаріший і найбільший в Словенії Університет Любляни. Перша версія програми була представлена в 1996 році. На даний момент продукт має версію 3.26.0 (від 01 вересня 2020 року). «Orange» реалізується як система всебічної візуалізації даних, яка допомагає виявляти приховані шаблони, забезпечує інтуїтивне розуміння процедур аналізу інформації або підтримує зв'язок між підсистемами в предметній області. Візуалізації включають в себе такі методи, як графік розсіювання, блочну діаграму і гістограму, а також візуалізації для конкретної моделі, такі як дендрограма, силуетна графіка і деревоподібна візуалізація тощо. Багато інших візуалізацій доступні у вигляді доповнень і включають візуалізації мереж, хмар слів, географічних карт і багато іншого. Візуалізація Orange є інтерактивною – можна вибирати точки даних на точковій діаграмі, вузли в дереві, гілки в дендрограмі.

Структурно Orange складається з інтерфейсу Canvas, на який користувач поміщає віджети і створює робочий процес аналізу даних. Віджети – це елементи графічного інтерфейсу користувача, що мають стандартний зовнішній вигляд і виконують стандартні дії. Такі дії варіюються від простої візуалізації даних, вибору підмножин і попередньої обробки до емпіричної оцінки алгоритмів навчання і прогнозного моделювання. Візуальне програмування реалізується через інтерфейс, в якому робочі процеси створюються шляхом зв'язування віджетів (плавними лініями на моніторі всередині робочого столу програми), попередньо визначених або розроблених їх користувачем (баз даних і систем їх обробки).

Схему експерименту у Orange наведено на рис. 3.5. У якості алгоритмів класифікації користувачів використовувався метод Random forest (див. рис. 3.6. Точність класифікації перевірялась за вбудованим у віджет «Test and Score» алгоритмом 10-fold cross-validation, у відповідності до якого дослідний

набір даних розбивається на 10 однакових за розміром блоків. З 10 блоків один залишається для тестування моделі, а інші 9 блоків використовуються як тренувальний набір. Процес повторюється 10 разів, і кожен з блоків використовується один раз як тестовий набір. Наприкінці аналізу отримують 10 результатів, по одному на кожен блок, вони усереднюються і дають одну оцінку. Перевага такого способу в тому, що всі спостереження використовуються і для тренування, і для тестування моделі, і кожне спостереження використовується для тестування в точності один раз.

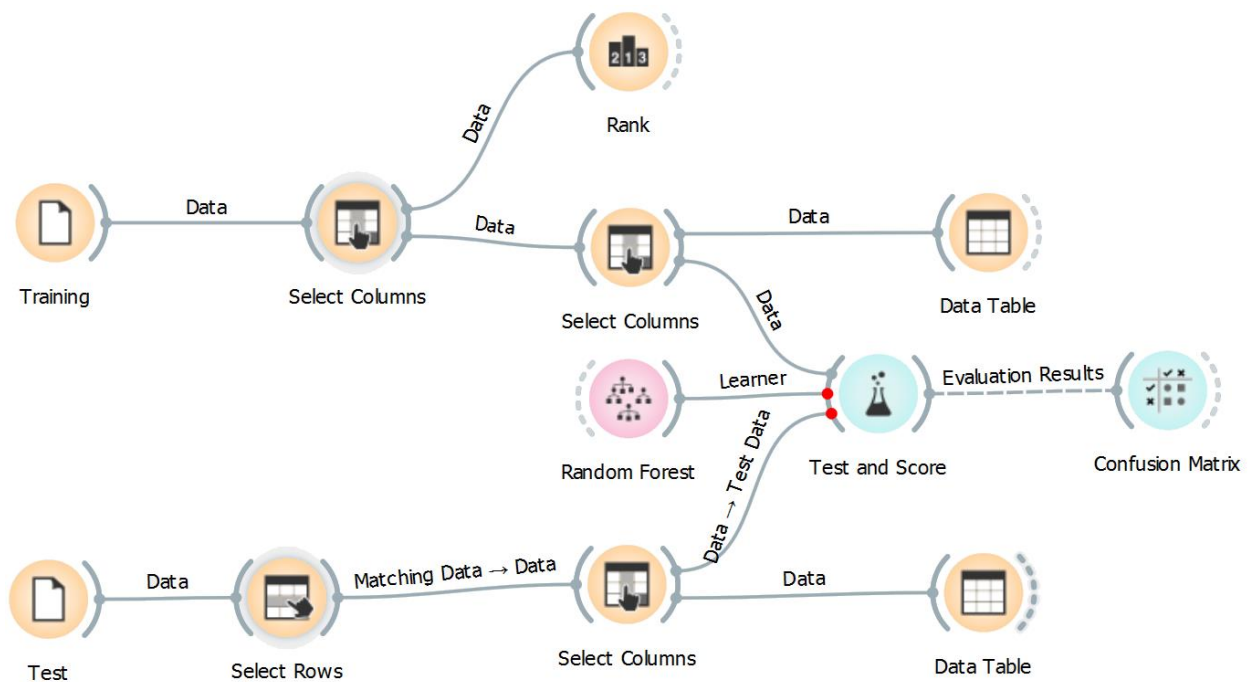


Рисунок 3.5 – Схема експерименту у Orange

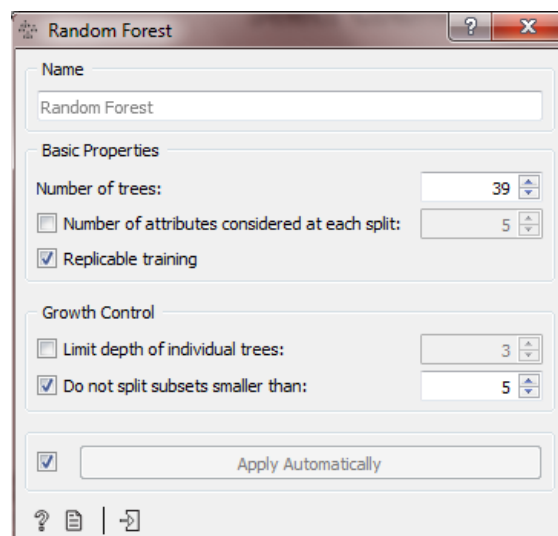


Рисунок 3.6 – Параметри використаного в дослідженнях методу класифікації

В якості параметрів достовірності отриманої математичної моделі ідентифікації (класифікації) використовувалась точність класифікації (Classification accuracy, CA):

$$CA = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3.1)$$

де TP (True Positive) – зареєстрованого користувача ідентифікований вірно і дозволено доступ; TN (True Negative) – зловмисника правильно не ідентифіковано і в доступі відмовлено; FP (False Positive) – зловмисника ідентифіковано як зареєстрованого користувача та дозволено отримати доступ; FN (False Negative) – зареєстрованого користувача ідентифіковано як зловмисника і в доступі відмовлено.

Результати мультикласової класифікації користувачів з тренувальної частини бази даних «Balabit Mouse Challenge Data Set» наведено на рис. 3.7 та рис. 3.8. Як можна побачити, інтегральна (усереднена за всіма користувачами) точність класифікації становить $CA = 47.7\%$, що є неприпустимим результатом. Проте, відповідно до матриці помилок класифікації (рис. 3.8), параметр TP (True Positive) для користувачів 7 та 9 більше 90% . Отже, перший висновок: аутентифікацію за динамікою системи «користувач – комп'ютерна миша» слід використовувати для користувачів з добре сформованим «курсорним почерком».

Результати двійкової класифікації користувачів з тренувальної частини бази даних «Balabit Mouse Challenge Data Set» наведено на рис. 3.9 (у цьому випадку цільовому користувачу було присвоєно клас 1, тобто «зареєстрований», а усім іншим користувачам – клас 2, тобто «зловмисник»). Як відомо з [31 – 33], для отримання найкращих результатів двійкової класифікації навчальні дані повинні бути збалансовані, тобто кількість позитивних і негативних навчальних даних має бути однаковою, тому для кожного користувача відповідну базу даних «зловмисників» було зменшено до 20% від максимального об'єму.

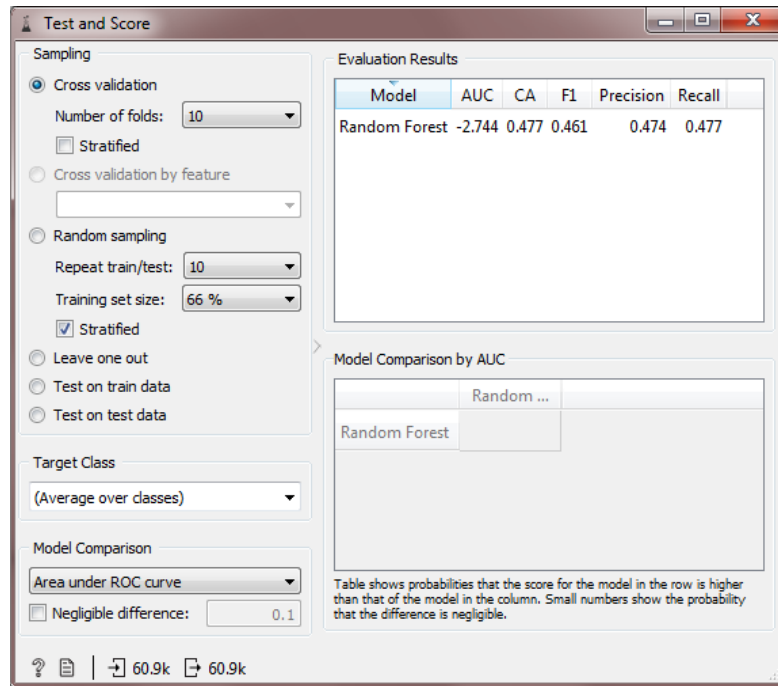


Рисунок 3.7 – Результати мультикласової класифікації користувачів з тренувальної частини бази даних «Balabit Mouse Challenge Data Set»

Confusion Matrix

Show:

Proportion of actual

		Predicted										Σ
		7	9	12	15	16	20	21	23	29	35	
Actual	7	93.2 %	2.2 %	0.1 %	0.0 %	0.1 %	3.8 %	0.3 %	0.0 %	0.3 %	0.1 %	5574
	9	4.9 %	93.8 %	0.0 %	0.0 %	0.0 %	1.1 %	0.2 %	0.0 %	0.0 %	0.0 %	3249
	12	0.2 %	0.0 %	42.3 %	9.8 %	23.7 %	0.3 %	6.4 %	4.6 %	11.0 %	1.7 %	7655
	15	0.5 %	0.0 %	10.4 %	42.3 %	25.4 %	0.3 %	6.8 %	8.9 %	3.7 %	1.8 %	7152
	16	0.2 %	0.0 %	11.8 %	10.0 %	55.0 %	0.2 %	6.5 %	4.6 %	9.9 %	1.9 %	10767
	20	28.3 %	4.3 %	6.6 %	7.2 %	15.3 %	23.7 %	4.9 %	6.1 %	1.5 %	1.9 %	3809
	21	0.7 %	0.0 %	12.4 %	10.8 %	29.4 %	0.2 %	28.6 %	6.6 %	7.5 %	3.7 %	5767
	23	0.3 %	0.0 %	12.1 %	15.0 %	24.8 %	0.2 %	8.5 %	26.1 %	8.5 %	4.6 %	5440
	29	0.3 %	0.0 %	6.0 %	4.8 %	19.7 %	0.2 %	4.3 %	3.8 %	59.5 %	1.4 %	7267
	35	0.4 %	0.1 %	15.2 %	11.1 %	30.7 %	0.3 %	11.1 %	14.8 %	8.4 %	7.9 %	4225
Σ	6588	3343	7960	7381	15918	1267	4782	4415	7816	1435	60905	

Select Correct

Select Misclassified

Clear Selection

?

Рисунок 3.8 – Матриця помилок мультикласової класифікації користувачів з тренувальної частини бази даних «Balabit Mouse Challenge Data Set»

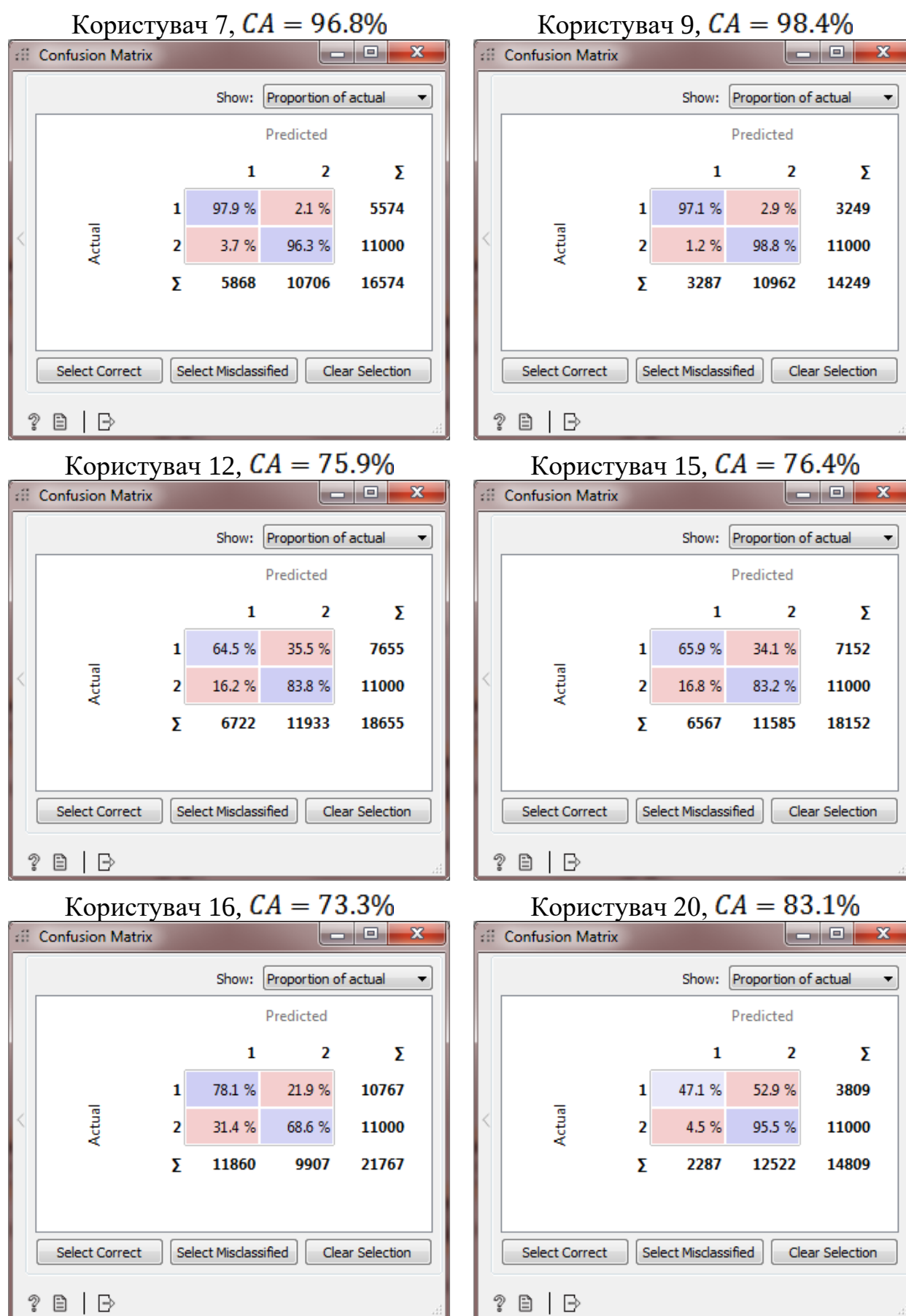


Рисунок 3.9 – Матриця помилок двійкової класифікації користувачів з тренувальної частини бази даних «Balabit Mouse Challenge Data Set»



Продовження рисунку 3.9

Як можна побачити, точність класифікації помітно зросла – мінімальне значення CA становить **73.3%** (для користувача 16); максимальне значення становить **98.4%** (для користувача 9); середнє значення становить **81.1%** проти **47.7%** для мультикласової класифікації. Також значно зросли значення параметру TP – мінімальне становить **24.7%** (для користувача 35) проти **7.9%** для мультикласової класифікації; максимальне значення становить **97.9%** (для користувача 7) проти **93.8%** для мультикласової класифікації; середнє становить **64.1%** проти **47.24%** для мультикласової класифікації. Отже, другий висновок: аутентифікацію за динамікою системи «користувач –

комп'ютерна миша» слід використовувати для сценарію двійкової класифікації, тобто для випадків, коли на комп'ютері зареєстровано тільки одного користувача.

Підтвердженням даного висновку є результати двійкової класифікації (рис. 3.10 та табл. 3.2) користувачів з тестової частини бази даних «Balabit Mouse Challenge Data Set» на основі класифікатора, побудованого за даними навчальної частини бази даних «Balabit Mouse Challenge Data Set». Очікувано точність класифікації дещо зменшилась – мінімальне значення **CA** становить **36.7%** (для користувача 21); максимальне значення становить **96.2%** (для користувача 9); середнє значення становить **60.8%**.

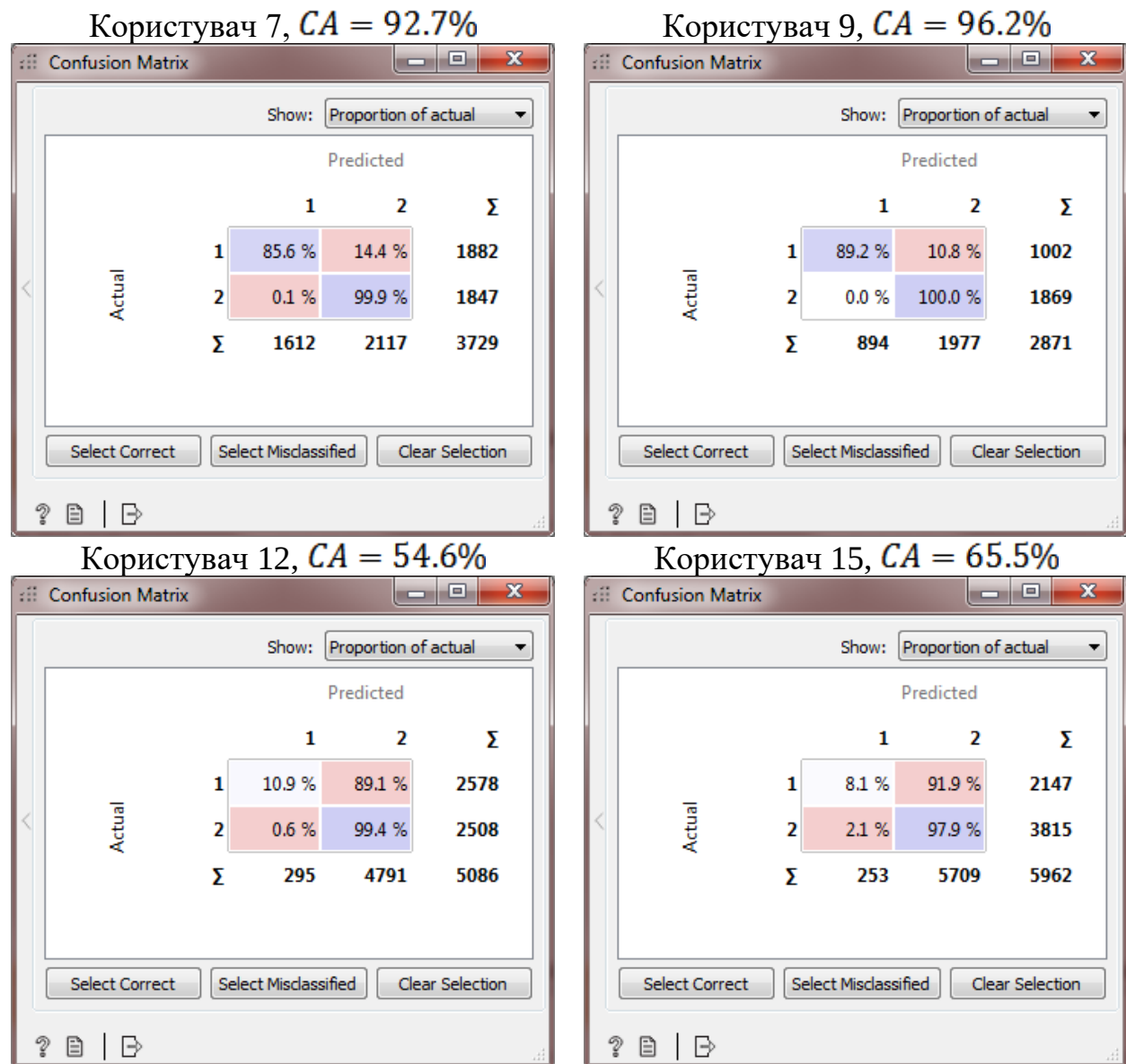
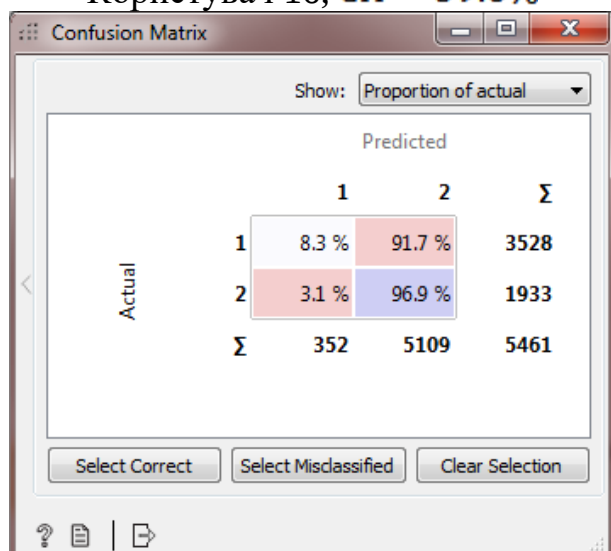
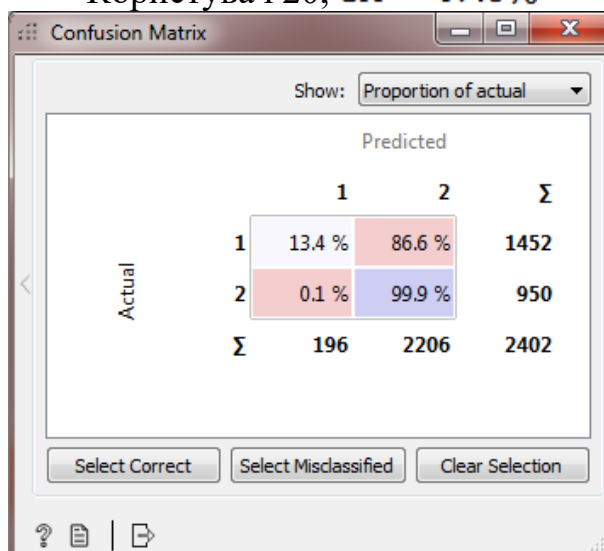
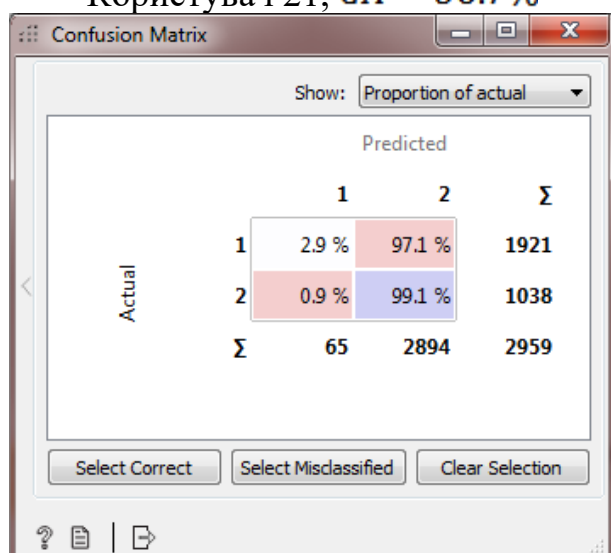
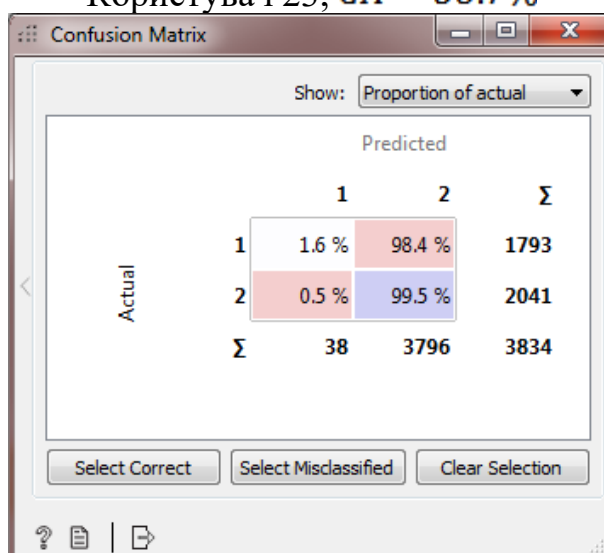
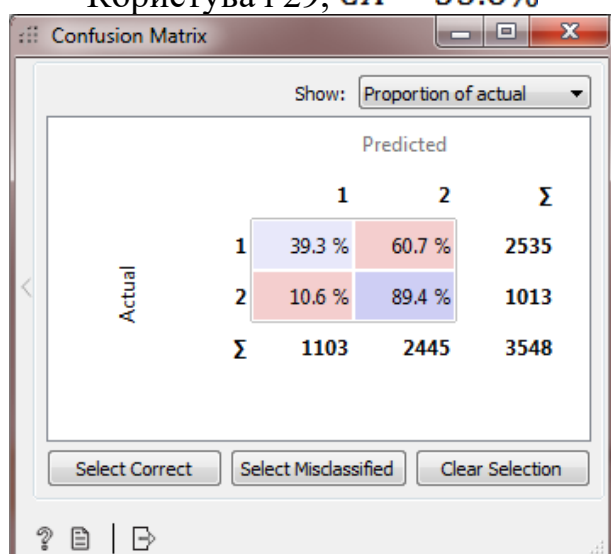
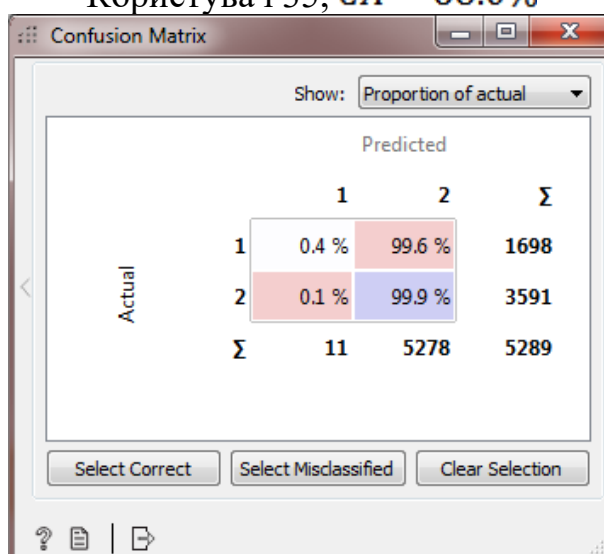


Рисунок 3.10 – Матриця помилок двійкової класифікації користувачів з тестової частини бази даних «Balabit Mouse Challenge Data Set»

Користувач 16, $CA = 39.6\%$ Користувач 20, $CA = 47.6\%$ Користувач 21, $CA = 36.7\%$ Користувач 23, $CA = 53.7\%$ Користувач 29, $CA = 53.6\%$ Користувач 35, $CA = 68.0\%$ 

Продовження рисунку 3.10

Таблиця 3.2 – Результати двійкової класифікації користувачів з тестової частини бази даних «Balabit Mouse Challenge Data Set» на основі класифікатора, побудованого за даними навчальної частини

Користувач	Точність класифікації, <i>CA</i>	Відсоток True Positive	Відсоток True Negative
7	92.7	85.6	99.9
9	96.2	89.2	100
12	54.6	10.9	99.4
15	65.5	8.1	97.9
16	39.6	8.3	96.9
20	47.6	13.4	99.9
21	36.7	2.9	99.1
23	53.7	1.6	99.5
29	53.6	39.3	89.4
35	68.0	0.4	99.9
<i>Середнє значення</i>	60.8	26.0	98.2

Також зменшилися значення параметру *TP* – мінімальне становить **0.4%** (для користувача 35); максимальне значення становить **89.2%** (для користувача 9); середнє становить **26%**. Проте значення параметру *TN* (True Negative) знаходяться в межах **89.4% – 100%** з середнім значенням **98.2%**. *Отже, третій висновок: аутентифікацію за динамікою системи «користувач – комп'ютерна миша» слід використовувати для сценарію двійкової класифікації, причому класифікатор повинен перевіряти користувача на приналежність до класу «зловмисники». В такому сценарії середній відсоток випадків, коли зловмисника правильно не ідентифіковано як зареєстрованого користувача і в доступі відмовлено, становить **98.2%**.*

Для пошуку найінформативніших параметрів динаміки системи «користувач – комп'ютерна миша» було проведено ряд експериментів.

Результати двійкової класифікації користувачів з тестової частини «Balabit Mouse Challenge Data Set» на основі класифікатора, побудованого за

даними навчальної частини бази даних «Balabit Mouse Challenge Data Set» за трьома типами дії миші (mouse movement – MM, point click – PC, drag-and-drop – DD) наведено в табл. 3.3. Як можна побачити, з трьох типів дій миші найвищу точність для задач аутентифікації за динамікою системи «користувач – комп'ютерна миша» забезпечують «drag-and-drop» дії миші. Проте отримані значення точності класифікації $CA = 61.27\%$ (MM), $CA = 60.41\%$ (PC) та $CA = 63.01\%$ (DD) не можна віднести навіть до задовільних. Підтверджують незадовільність роботи класифікатора і отримані значення параметру True Positive: $TP = 23.33\%$ (MM), $TP = 26.03\%$ (PC), $TP = 30.81\%$ (DD). Водночас, якщо за сценарієм аутентифікації класифікатор перевірятиме користувача на приналежність до класу «зловмисник» (параметр True Negative), то найінформативнішими є «mouse movement» дії: мінімальне значення $TN = 92.2\%$, максимальне значення $TN = 100\%$, середнє значення $TN = 98.5\%$.

Таблиця 3.3 – Результати двійкової класифікації користувачів за трьома типами дії миші з тестової частини бази даних «Balabit Mouse Challenge Data Set» на основі класифікатора, побудованого за даними навчальної частини

Користувач	Тип дії миші	Точність класифікації, CA	Відсоток True Positive	Відсоток True Negative
7	MM	92.2	82.5	100
9	MM	96.3	88.8	100
12	MM	55.7	9.8	98.8
15	MM	75.2	11.5	97.9
16	MM	44.3	7.9	97.1
20	MM	48.1	4.3	100
21	MM	29.2	4.9	99.0
23	MM	62.7	0.4	100
29	MM	45.8	23.2	92.2
35	MM	63.2	0.0	100
Середнє значення		61.27	23.33	98.5

Продовження таблиці 3.3

Користувач	Тип дії миши	Точність класифікації, <i>СА</i>	Відсоток True Positive	Відсоток True Negative
7	PC	92.2	85.1	99.8
9	PC	97.0	91.7	100
12	PC	53.8	8.9	99.2
15	PC	61.9	7.9	97.6
16	PC	37.8	7.6	97.4
20	PC	46.0	14.6	100
21	PC	40.5	1.3	99.7
23	PC	51.6	3.3	99.0
29	PC	53.4	39.9	89.0
35	PC	69.9	0.0	100
Середнє значення		60.41	26.03	98.17
7	DD	87.2	77.7	98.2
9	DD	87.9	60.5	100
12	DD	60.1	32.3	99.0
15	DD	74.8	26.4	95.5
16	DD	49.2	26.9	89.0
20	DD	56.1	19.3	99.0
21	DD	30.7	8.2	90.5
23	DD	55.8	1.4	100
29	DD	64.6	52.9	92.9
35	DD	63.7	2.5	97.9
Середнє значення		63.01	30.81	96.2

Отже, четвертий висновок: найінформативнішим параметром з трьох типів дій миші для задач аутентифікації за динамікою системи «користувач – комп'ютерна миша» є «mouse movement» дії миші (для сценарію перевірки класифікатором користувача на приналежність до класу «зловмисник»).

Інформативність параметрів курсорного почерку також оцінювалася за допомогою інструмента «Rank» з набору віджетів «Data» програмного середовища Orange. Цей інструмент розраховує коефіцієнт впливу кожної ознаки на формування дослідного класу. Більші коефіцієнти впливу вказують на бі-

льшу інформативність тієї чи іншої ознаки. В Orange використовується 6 критеріїв ранжування.

Gain Ratio – критерій, що враховує кількість інформації, необхідної для розділення класу за поточними ознаками (ентропійний критерій).

Gini – метрика якості, яка часто використовується при оцінці моделей передбачення в задачах бінарної класифікації в умовах сильної незбалансованості класів цільової змінної.

ANOVA – критерій дисперсійного аналізу, застосовується для дослідження впливу однієї або декількох якісних ознак (факторів) на одну залежну кількісну змінну (клас).

χ^2 – непараметричний критерій дисперсійного аналізу з повторюваними вимірами.

ReliefF – цей метод випадковим чином вибирає з датасета зразки і оновлює значимість кожної ознаки на основі різниці між обраним екземпляром і двома найближчими до нього об'єктами того ж і протилежного класів. Якщо розрахована різниця в значеннях ознаки для двох найближчих сусідів одного класу, її важливість знижується, а якщо, навпаки, спостерігається відмінність між значеннями ознаки для об'єктів різних класів, важливість, відповідно, підвищується.

FCBF – метод працює швидше і ефективніше, ніж ReliefF, і тому частіше використовується для вхідних даних високої розмірності. В рамках методу FCBF спочатку для всіх пар ознак обчислюються взаємна інформація, поділена на суму їх ентропій, потім ознаки сортуються за цим критерієм і потім видаляються надлишкові ознаки.

Ранжування ознак для користувача 9 (такого, який має унікальний і добре сформований курсорний почерк) з навчальної частини бази даних «Balabit Mouse Challenge Data Set» наведено на рис. 3.11 – 3.16.

Rank

Scoring Methods

- ☐ Information Gain
- ☒ Information Gain Ratio
- ☒ Gini Decrease
- ☒ ANOVA
- ☒ χ^2
- ☒ ReliefF
- ☒ FCBF

Select Attributes

- ☐ None
- ☐ All
- ☐ Manual
- ☒ Best ranked: 10

☒ Send Automatically

	#	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
a_beq_time		0.039	0.001	133.990	410.797	0.005	0.073
mean_a		0.025	0.006	10.962	94.855	0.005	0.000
max_jerk		0.021	0.005	11.011	88.427	0.011	0.000
sd_v		0.020	0.005	97.728	75.768	0.025	0.000
sd_jerk		0.017	0.004	5.878	92.073	0.005	0.000
traveled_distance_pixel		0.017	0.005	67.461	219.694	0.017	0.000
sd_vx		0.017	0.004	66.167	65.708	0.020	0.030
dist_end_to_end_line		0.016	0.005	132.082	79.809	0.065	0.000
max_v		0.016	0.004	78.306	214.658	0.023	0.000
straightness		0.015	0.004	5.675	119.642	0.045	0.000
min_vx		0.014	0.003	49.876	158.104	0.017	0.000
elapsed_time		0.014	0.004	40.847	165.632	0.005	0.000

6467 6467

Рисунок 3.11 – Ранжування параметрів курсорного почерку користувача 9 за ранговим критерієм Gain Ratio

Rank

Scoring Methods

- ☐ Information Gain
- ☒ Information Gain Ratio
- ☒ Gini Decrease
- ☒ ANOVA
- ☒ χ^2
- ☒ ReliefF
- ☒ FCBF

Select Attributes

- ☐ None
- ☐ All
- ☐ Manual
- ☒ Best ranked: 10

☒ Send Automatically

	#	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
a_beq_time		0.039	0.001	133.990	410.797	0.005	0.073
mean_a		0.025	0.006	10.962	94.855	0.005	0.000
max_jerk		0.021	0.005	11.011	88.427	0.011	0.000
sd_v		0.020	0.005	97.728	75.768	0.025	0.000
traveled_distance_pixel		0.017	0.005	67.461	219.694	0.017	0.000
dist_end_to_end_line		0.016	0.005	132.082	79.809	0.065	0.000
sd_vx		0.017	0.004	66.167	65.708	0.020	0.030
sd_jerk		0.017	0.004	5.878	92.073	0.005	0.000
max_v		0.016	0.004	78.306	214.658	0.023	0.000
elapsed_time		0.014	0.004	40.847	165.632	0.005	0.000
min_omega		0.013	0.004	0.014	9.569	0.065	0.000
mean_vx		0.013	0.004	13.378	46.418	0.014	0.000

6467 6467

Рисунок 3.12 – Ранжування параметрів курсорного почерку користувача 9 за ранговим критерієм Gini

Rank

Scoring Methods

- ☐ Information Gain
- ☒ Information Gain Ratio
- ☒ Gini Decrease
- ☒ ANOVA
- ☒ χ^2
- ☒ ReliefF
- ☒ FCBF

Select Attributes

- ☐ None
- ☐ All
- ☐ Manual
- ☒ Best ranked: 10

☒ Send Automatically

	#	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
num_critical_points		0.001	0.000	280.407	11.720	0.013	0.000
num_points		0.013	0.003	19.995	29.016	0.018	0.000
a_beq_time		0.039	0.001	133.990	410.797	0.005	0.073
dist_end_to_end_line		0.016	0.005	132.082	79.809	0.065	0.000
sum_of_angles		0.009	0.002	199.287	87.551	0.017	0.016
sd_v		0.020	0.005	97.728	75.768	0.025	0.000
largest_deviation		0.008	0.002	93.164	89.289	0.019	0.000
mean_v		0.011	0.003	84.373	150.789	0.023	0.000
direction... movement		0.009	0.002	79.393	31.597	0.017	0.000
max_v		0.016	0.004	78.306	214.658	0.023	0.000
traveled_distance_pixel		0.017	0.005	67.461	219.694	0.017	0.000
sd_vx		0.017	0.004	66.167	65.708	0.020	0.030

6467 6467

Рисунок 3.13 – Ранжування параметрів курсорного почерку користувача 9 за ранговим критерієм ANOVA

	#	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
a_beq_time		0.039	0.001	133.990	410.787	0.005	0.000
sd_v		0.020	0.005	97.726	259.768	0.025	0.000
traveled_distance_pixel		0.017	0.005	67.461	19.694	0.017	0.000
max_v		0.016	0.004	78.306	14.658	0.023	0.000
dist_end_to_end_line		0.016	0.005	132.082	309.809	0.069	0.000
elapsed_time		0.014	0.004	40.347	165.632	0.005	0.000
min_vx		0.014	0.003	49.876	158.104	0.017	0.000
mean_v		0.011	0.003	84.373	150.789	0.023	0.000
straightness		0.015	0.004	5.675	119.642	0.045	0.000
mean_a		0.025	0.006	10.962	94.855	0.009	0.000
sd_jerk		0.017	0.004	5.878	92.073	0.005	0.000
largest_deviation		0.006	0.002	93.164	89.289	0.018	0.000

Рисунок 3.14 – Ранжування параметрів курсорного почерку користувача 9 за ранговим критерієм χ^2

	#	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
direction...movement		0.039	0.002	79.303	31.597	0.113	0.000
dist_end_to_end_line		0.016	0.005	132.082	309.809	0.069	0.000
min_omega		0.013	0.004	0.014	9.569	0.063	0.000
sd_omega		0.005	0.001	4.009	1.441	0.058	0.000
max_omega		0.004	0.001	6.751	46.931	0.051	0.000
max_curv		0.010	0.002	3.619	1.831	0.045	0.000
straightness		0.015	0.004	5.675	119.642	0.045	0.000
mean_omega		0.005	0.001	3.044	3.696	0.033	0.000
min_curv		0.012	0.002	63.353	71.296	0.026	0.000
sd_v		0.020	0.005	97.726	259.768	0.025	0.000
sd_curv		0.008	0.002	47.753	83.097	0.024	0.000
max_v		0.016	0.004	78.306	214.658	0.023	0.000

Рисунок 3.15 – Ранжування параметрів курсорного почерку користувача 9 за ранговим критерієм ReliefF

	#	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
a_beq_time		0.039	0.001	133.990	410.787	0.005	0.000
sd_vx		0.017	0.004	66.167	85.708	0.020	0.000
sum_of_angles		0.009	0.002	109.287	87.551	0.017	0.000
mean_a		0.025	0.006	10.962	94.855	0.009	0.000
max_jerk		0.021	0.005	11.011	88.427	0.011	0.000
sd_v		0.020	0.005	97.726	259.768	0.025	0.000
sd_jerk		0.017	0.004	5.878	92.073	0.005	0.000
traveled_distance_pixel		0.017	0.005	67.461	19.694	0.017	0.000
dist_end_to_end_line		0.016	0.005	132.082	309.809	0.069	0.000
max_v		0.016	0.004	78.306	14.658	0.023	0.000
straightness		0.015	0.004	5.675	119.642	0.045	0.000
min_vx		0.014	0.003	49.876	158.104	0.017	0.000

Рисунок 3.16 – Ранжування параметрів курсорного почерку користувача 9 за ранговим критерієм FCBF

Як можна побачити з рис. 3.17, дванадцять ознак курсорного почерку зустрічаються мінімум у трьох з шести критеріїв: `a_beg_time`, `max_curv`, `max_jerk`, `max_omega`, `mean_a`, `mean_jerk`, `min_curv`, `min_jerk`, `min_omega`, `num_critical_points`, `num_points`, `sd_jerk`. Отже, особливості руху курсору, пов'язані з ривками та формою траєкторії курсору, є більш інформативними, ніж особливості, пов'язані зі швидкістю курсору.

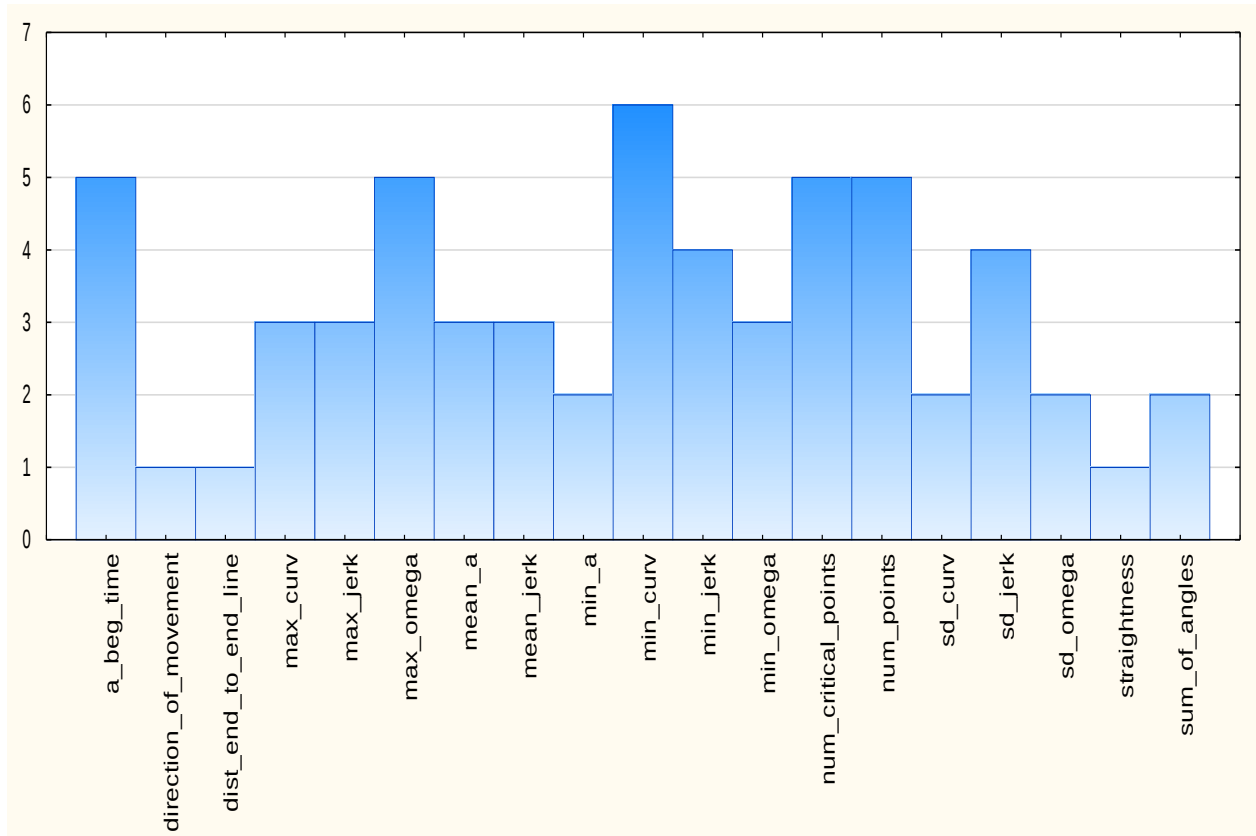


Рисунок 3.17 – Результати ранжування параметрів курсорного почерку користувача 9 за шістьма ранговими критеріями

Результати ранжування інформативних параметрів курсорного почерку для інших користувачів з навчальної частини бази даних «Balabit Mouse Challenge Data Set» наведено на рис. 3.18 – 3.26.

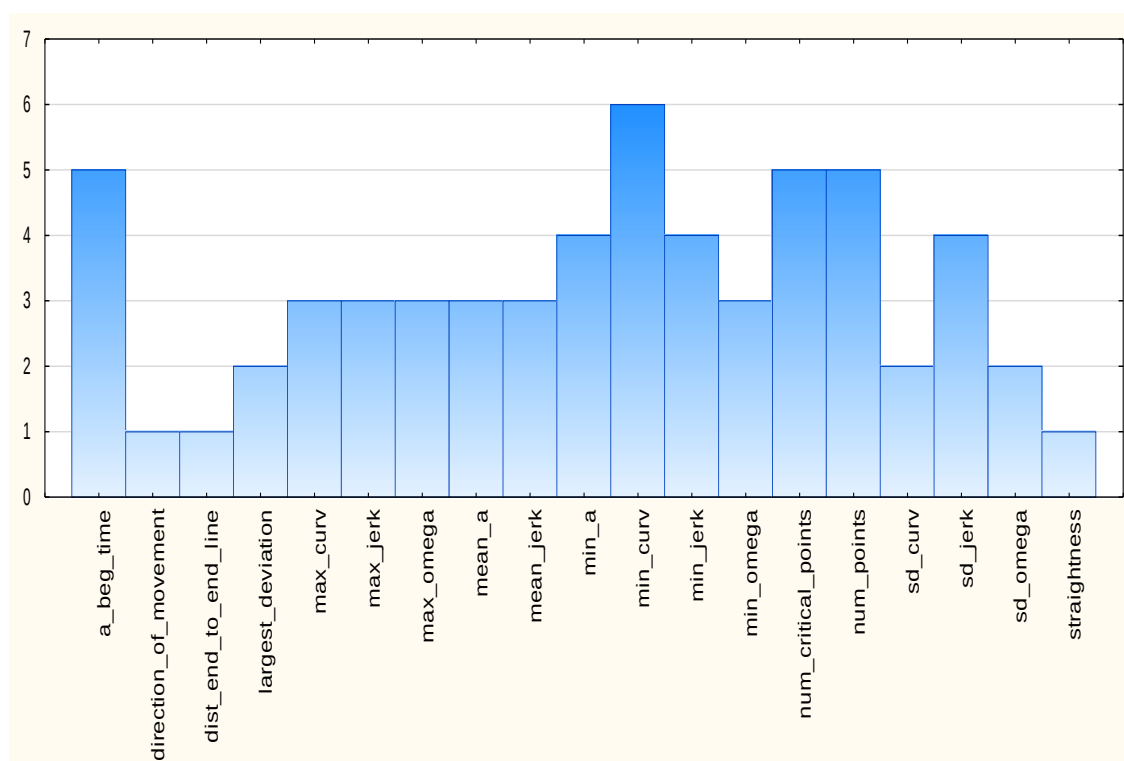


Рисунок 3.18 – Результати ранжування параметрів курсорного почерку користувача 7 за шістьма ранговими критеріями

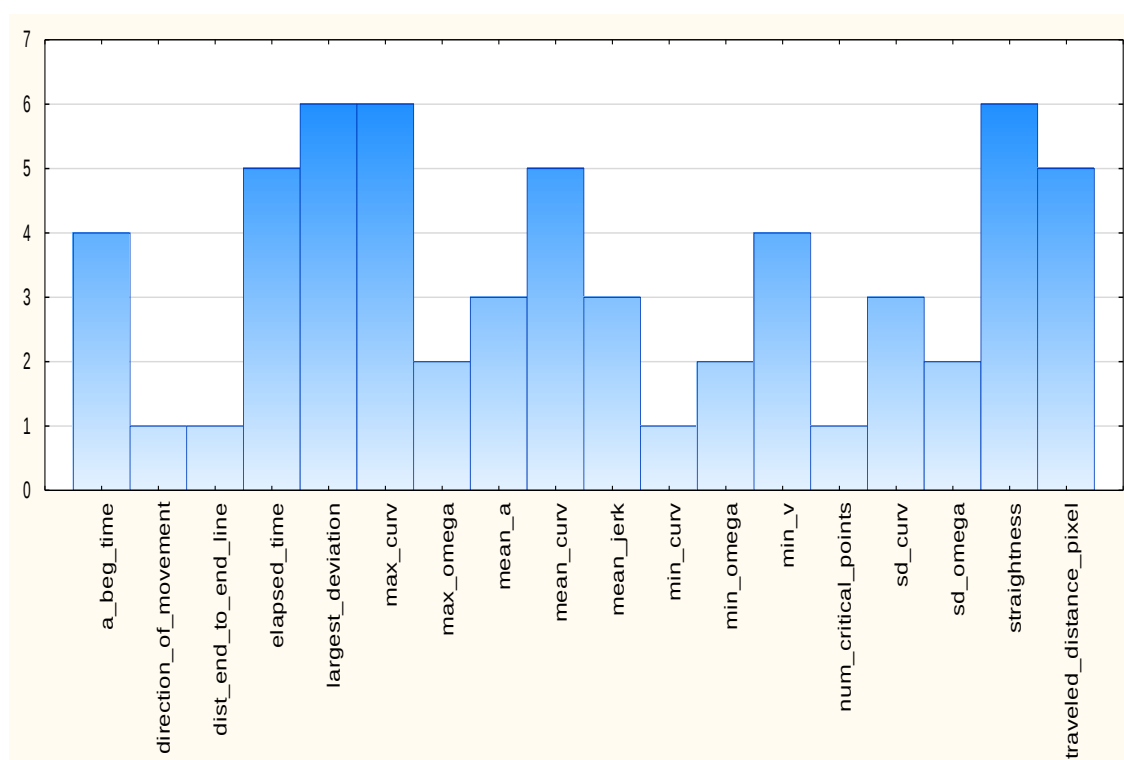


Рисунок 3.19 – Результати ранжування параметрів курсорного почерку користувача 12 за шістьма ранговими критеріями

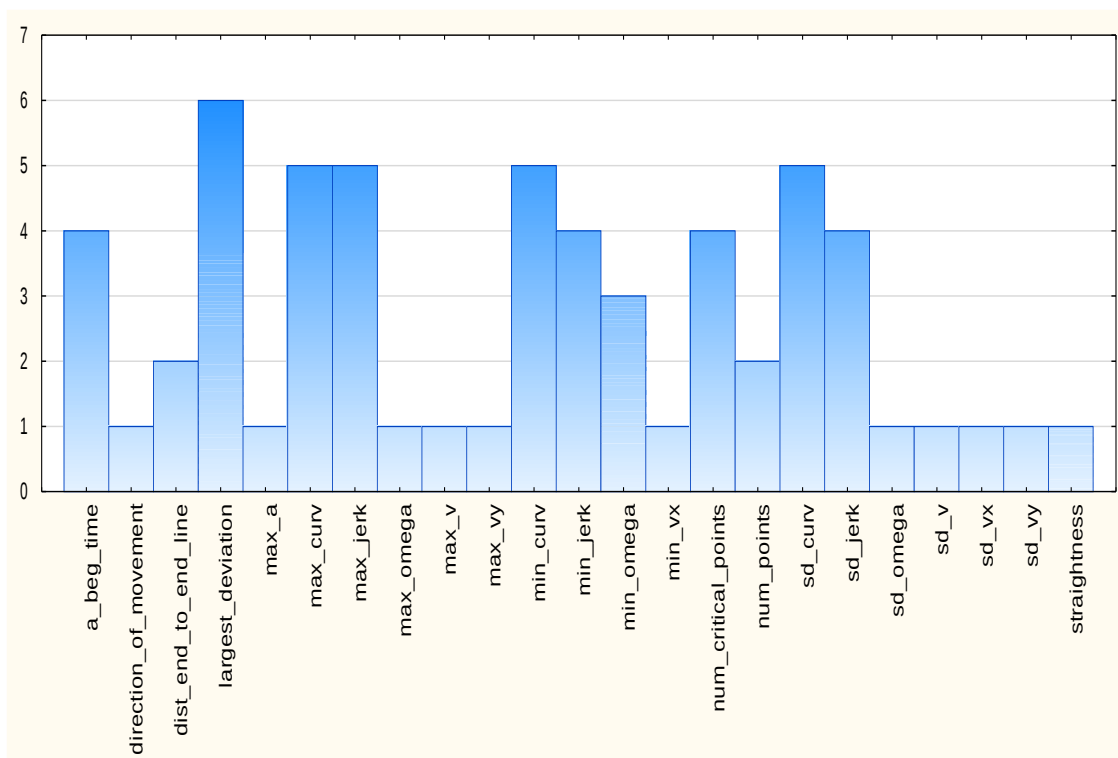


Рисунок 3.20 – Результати ранжування параметрів курсорного почерку користувача 15 за шістьма ранговими критеріями

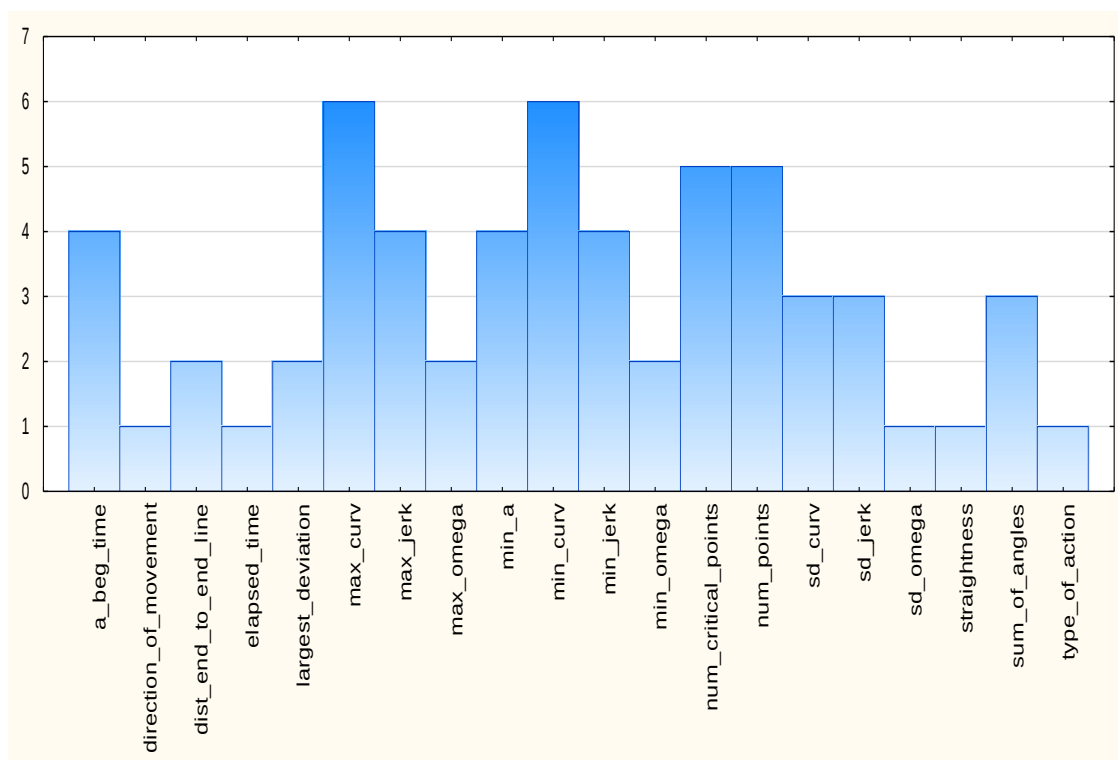


Рисунок 3.21 – Результати ранжування параметрів курсорного почерку користувача 16 за шістьма ранговими критеріями

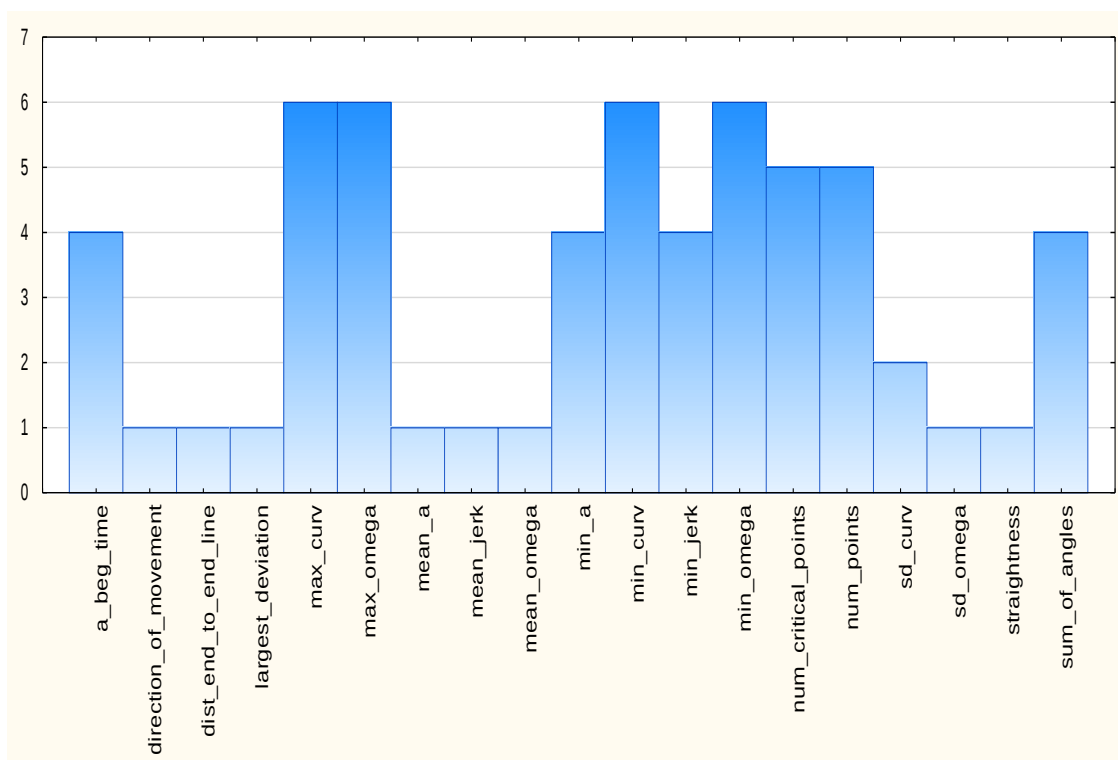


Рисунок 3.22 – Результати ранжування параметрів курсорного почерку користувача 20 за шістьма ранговими критеріями

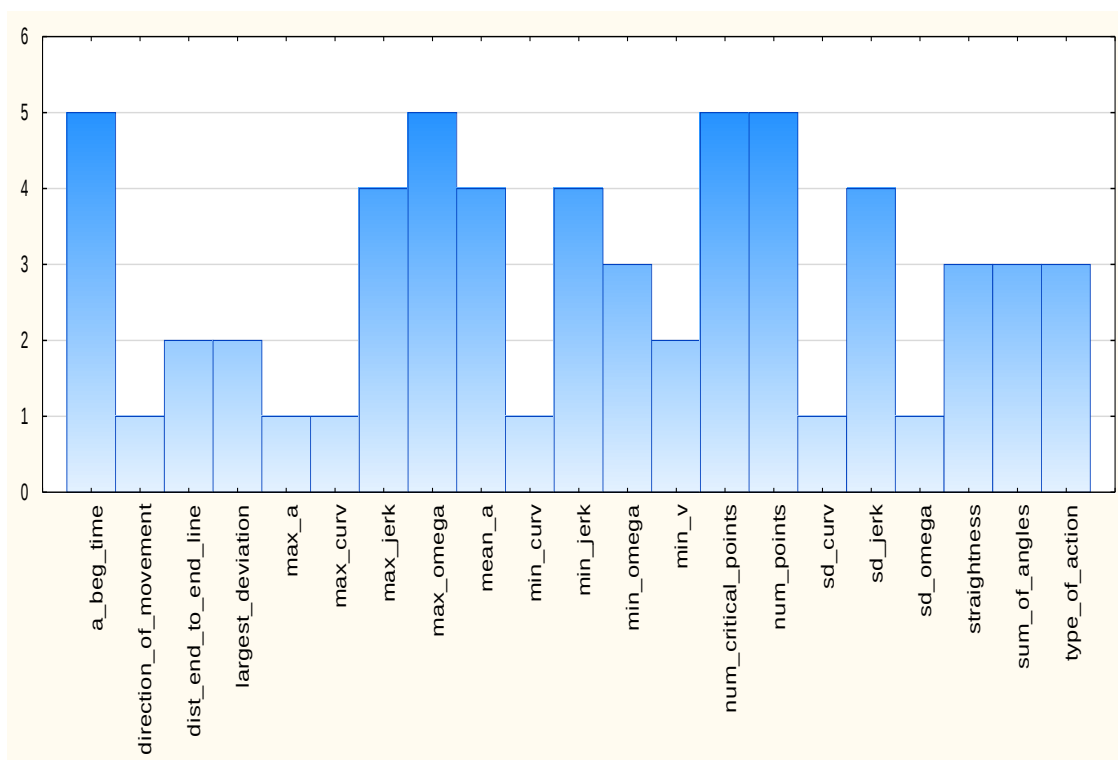


Рисунок 3.23 – Результати ранжування параметрів курсорного почерку користувача 21 за шістьма ранговими критеріями

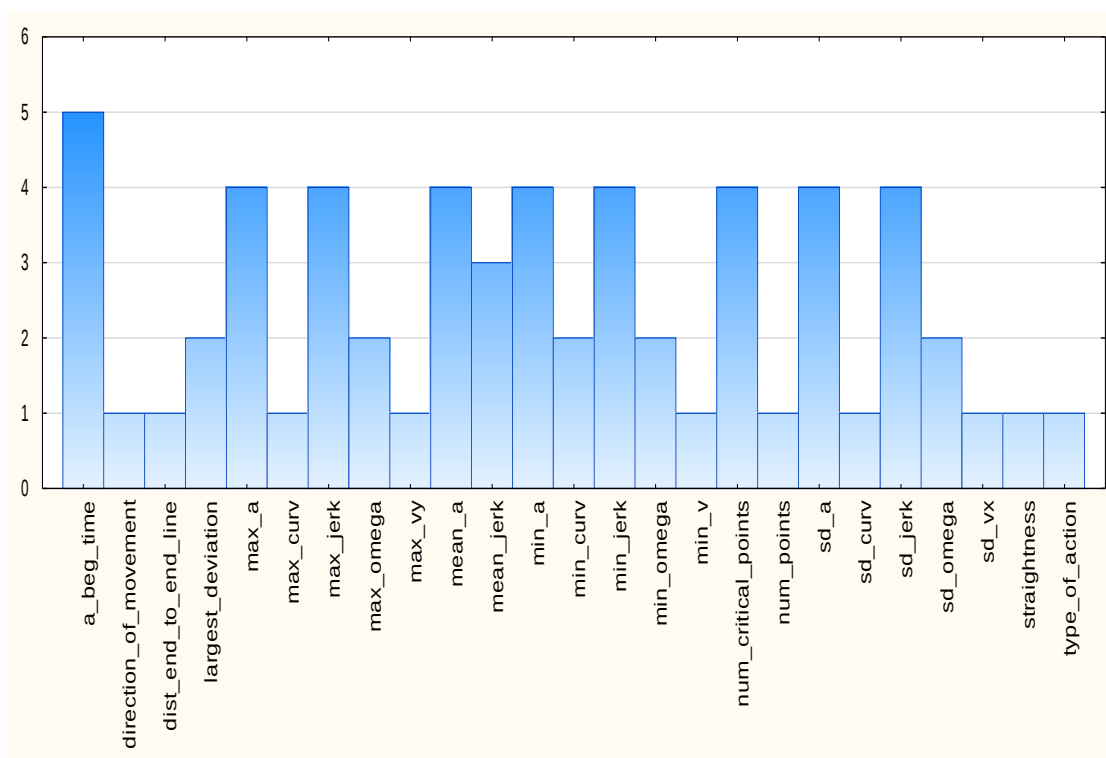


Рисунок 3.24 – Результати ранжування параметрів курсорного почерку користувача 23 за шістьма ранговими критеріями

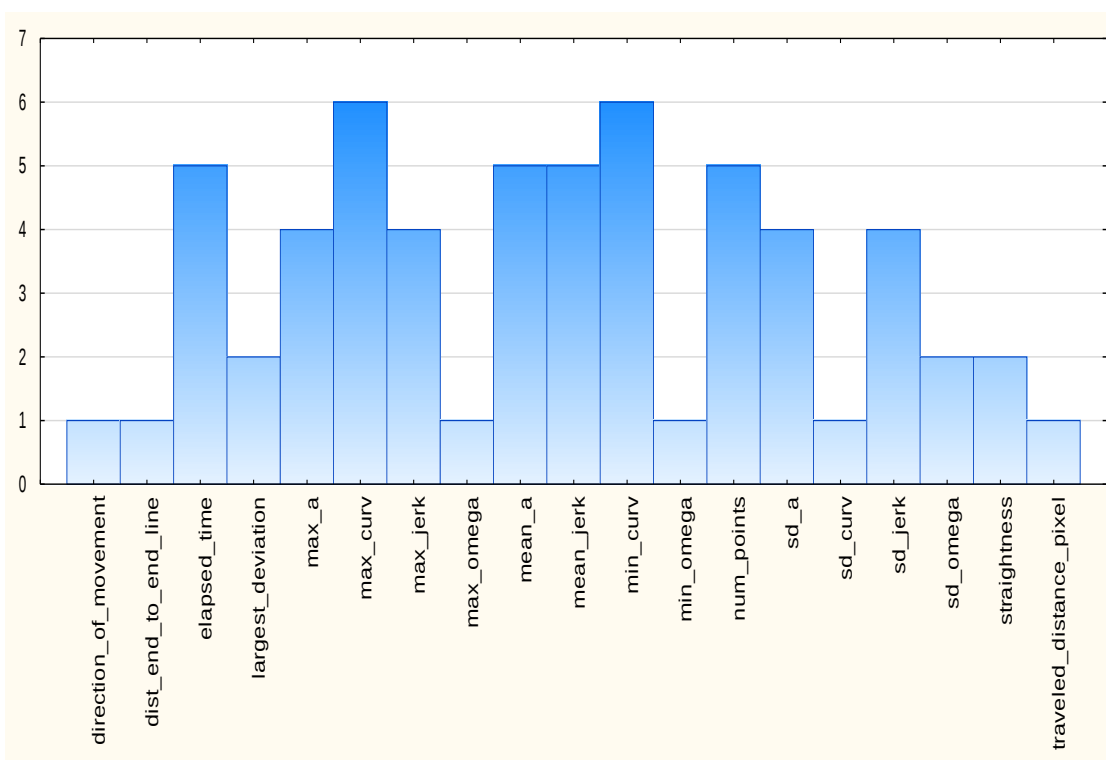


Рисунок 3.25 – Результати ранжування параметрів курсорного почерку користувача 29 за шістьма ранговими критеріями

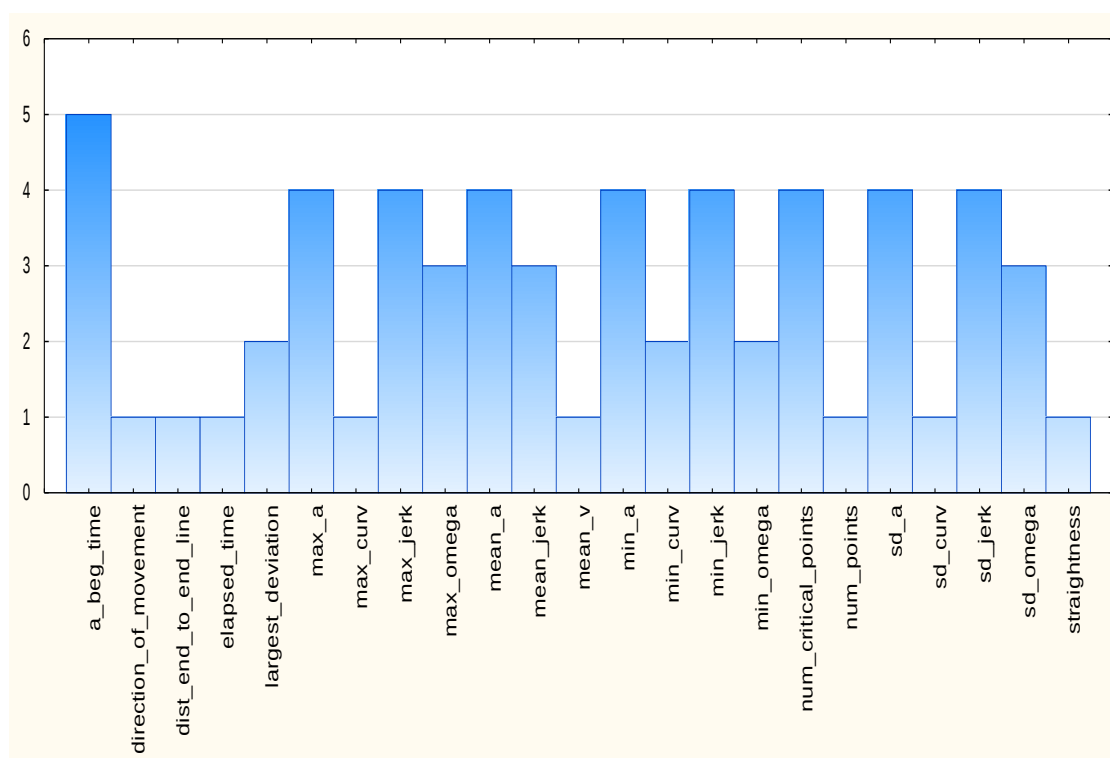


Рисунок 3.26 – Результати ранжування параметрів курсорного почерку користувача 35 за шістьма ранговими критеріями

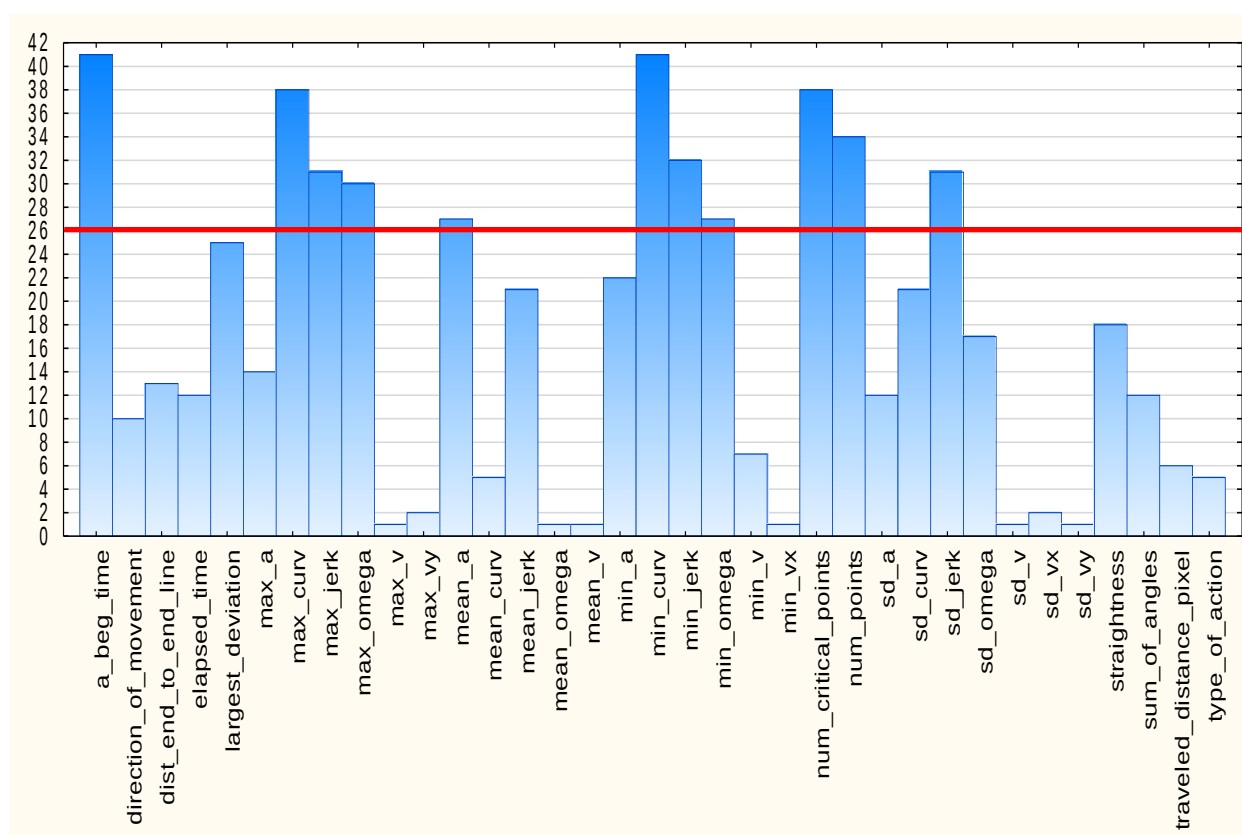


Рисунок 3.27 – Інтегральні результати ранжування параметрів курсорного почерку усіх користувачів за шістьма ранговими критеріями

Як можна побачити з рис. 3.17 – 3.26, гістограми рангів інформативних параметрів курсорного почерку помітно відрізняються одна від одної, проте відповідно до рис. 3.27, найінформативнішими параметрами курсорного почерку за результатами врахування усіх користувачів є: a_beg_time , max_curv , max_jerk , max_omega , $mean_a$, min_curv , min_jerk , min_omega , $num_critical_points$, num_points , sd_jerk . Усі ці одинадцять параметрів (за винятком $mean_jerk$) є найінформативнішими для користувача 9 (див. рис. 3.17). *Отже, п'ятий висновок: найінформативнішими параметрами курсорного почерку є a_beg_time , max_curv , max_jerk , max_omega , $mean_a$, min_curv , min_jerk , min_omega , $num_critical_points$, num_points , sd_jerk , що пов'язані з ривками та формою траєкторії курсору миші.*

Для того, щоб кількісно підтвердити обґрунтованість висновку 4 та висновку 5 було проведено двійкову класифікацію користувачів за дванадцятьма інформативними параметрами курсорного почерку (a_beg_time , max_curv , max_jerk , max_omega , $mean_a$, min_curv , min_jerk , min_omega , $num_critical_points$, num_points , sd_jerk , $type_of_action$ – mouse movement) з тестової частини «Balabit Mouse Challenge Data Set» на основі класифікатора, побудованого за даними з навчальної частини бази даних «Balabit Mouse Challenge Data Set». Результати досліджень наведено в табл. 3.4.

Як можна побачити з табл. 3.4, інтегральна точність класифікації фактично не змінилась – $CA_{12} = 60.4\%$ проти $CA_{41} = 60.8\%$. Якщо ж проаналізувати отриману точність класифікації по кожному з користувачів, то можна відмітити наступне:

1) для користувачів 7 та 9 з найунікальнішими курсорними почерками точність класифікації також можна вважати не змінилась – $CA_{12}^{(7)} = 91.2\%$ проти $CA_{41}^{(7)} = 92.7\%$ та $CA_{12}^{(9)} = 95.0\%$ проти $CA_{41}^{(9)} = 96.2\%$;

2) для користувачів 15 та 23 можна спостерігати значне підвищення точності класифікації – $CA_{12}^{(15)} = 74.1\%$ проти $CA_{41}^{(15)} = 65.5\%$ та $CA_{12}^{(23)} = 62.5\%$ проти $CA_{41}^{(23)} = 63.7\%$. Отже, можна вважати, що для корис-

тувачів 15 та 23 дії миші «point click» та «drag-and-drop» є «шумовим», тобто такими, що зменшують унікальність курсорного почерку.

Таблиця 3.4 – Результати двійкової класифікації користувачів за 12-ма та 40-ка інформативними параметрами з тестової частини бази даних «Balabit Mouse Challenge Data Set» на основі класифікатора, побудованого за даними навчальної частини

Користувач	Точність класифікації, <i>CA</i>	Відсоток True Negative	Точність класифікації, <i>CA</i>	Відсоток True Negative
	12 інформативних параметрів		41 інформативний параметр	
7	91.2	99.8	92.7	99.9
9	95.0	100	96.2	100
12	54.8	99.4	54.6	99.4
15	74.1	99.4	65.5	97.9
16	42.0	97.8	39.6	96.9
20	48.3	100	47.6	99.9
21	28.6	99.0	36.7	99.1
23	62.5	100	53.7	99.5
29	44.6	94.8	53.6	89.4
35	63.1	99.9	68.0	99.9
<i>Середнє значення</i>	60.4	99.0	60.8	98.2

Середній відсоток правильно ідентифікованих зломисників, які видавали себе зареєстрованими користувачами становить $TN_{12} = 99.0\%$, що перевищує відповідний відсоток для повної бази $TN_{41} = 98.2\%$. Таким чином, для сценарію аутентифікації, коли класифікатор перевіряє користувача на приналежність до класу «зломисник», можна отримати дуже високі показники точності. Якщо ж проаналізувати отриманий відсоток True Negative по

кожному з користувачів, то можна відмітити наступне:

1) тільки для двох користувачів (7 та 21) відсоток True Negative для повної бази був більший ніж для скороченої бази, причому різниця в значеннях параметру *TN* для обох користувачів склала **0.1%**;

2) для восьми інших користувачів відсоток True Negative для повної бази був менший ніж для скороченої бази, причому максимальна різниця в значеннях параметру *TN* складає **5.4%** (для користувача 29).

Підводячи підсумки проведених експериментів, можна відмітити наступні висновки.

1. Аутентифікацію за динамікою системи «користувач – комп’ютерна миша» слід використовувати для користувачів з добре сформованим «курсорним почерком».

2. Аутентифікацію за динамікою системи «користувач – комп’ютерна миша» слід використовувати для сценарію двійкової класифікації, тобто для випадків, коли на комп’ютері зареєстровано тільки одного користувача.

3. Аутентифікацію за динамікою системи «користувач – комп’ютерна миша» слід використовувати для сценарію двійкової класифікації, причому класифікатор повинен перевіряти користувача на приналежність до класу «зловмисники».

4. Найінформативнішим параметром з трьох типів дій миші для задач аутентифікації за динамікою системи «користувач – комп’ютерна миша» є «mouse movement» дії миші (для сценарію перевірки класифікатором користувача на приналежність до класу «зловмисник»).

5. Найінформативнішими параметрами курсорного почерку є *a_beg_time*, *max_curv*, *max_jerk*, *max_omega*, *mean_a*, *min_curv*, *min_jerk*, *min_omega*, *num_critical_points*, *num_points*, *sd_jerk*, що пов'язані з ривками та формою траєкторії курсору миші.

ВИСНОВКИ

1. Одночасне виконання вимог по безперервності й зручності використання в комп'ютерних мережах не може забезпечити ніякий спосіб ідентифікації, крім динамічного. Крім того, динамічна біометрія дешевше статичної, так як може використовувати стандартні пристрої введення інформації.

2. Аутентифікація користувачів за курсорним почерком є досить перспективним напрямком досліджень і може широко застосовуватись для забезпечення безпеки як домашніх комп'ютерів, так і комп'ютерів найбільших корпорацій, а також для запобігання несанкціонованому доступу зловмисників до web-ресурсів.

3. У роботі проаналізовано 41 інформативну ознаку курсорного почерку десяти користувачів з датасету «Balabit Mouse Challenge Data Set».

4. Аутентифікацію за курсорним почерком слід використовувати для користувачів з добре сформованим «курсорним почерком».

5. Аутентифікацію за курсорним почерком слід використовувати для сценарію двійкової класифікації, тобто для випадків, коли на комп'ютері зареєстровано тільки одного користувача.

6. Аутентифікацію за курсорним почерком слід використовувати для сценарію двійкової класифікації, причому класифікатор повинен перевіряти користувача на приналежність до класу «зловмисники».

7. Найінформативнішим параметром з трьох типів дій миші для задач аутентифікації за курсорним почерком є «mouse movement» дії миші (для сценарію перевірки класифікатором користувача на приналежність до класу «зловмисник»).

9. Найінформативнішими параметрами курсорного почерку є `a_beg_time`, `max_curv`, `max_jerk`, `max_omega`, `mean_a`, `min_curv`, `min_jerk`, `min_omega`, `num_critical_points`, `num_points`, `sd_jerk`, що пов'язані з ривками та формою траєкторії курсору миші.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Болл Р.М., Коннел Дж.Х., Панканти Ш., Ратха Н.К., Сеньор Э.У. Руководство по биометрии / Пер. с англ. Н.Е. Агаповой. – М.: Техносфера, 2007. – 368 с.
2. Obaidat, Mohammad S., Issa Traore, and Isaac Woungang. Biometric-Based Physical and Cybersecurity Systems. Springer International Publishing, 2019.
3. Иванов А.И. Биометрическая идентификация личности по динамике подсознательных движений. – Пенза: Изд-во Пенз. гос. ун-та, 2000. – 188 с.
4. Иванов А.И. Биометрическая идентификация личности по динамике быстрых движений. //Специальная техника средств связи. Серия Системы, сети и технические средства конфиденциальной связи. Пенза: Изд. ПНИЭИ.- 1997- Вып. 2.- С. 88-93.
5. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. //М.: Мир – 1992 – 240 с.
6. Анатомия человека. /Э.И.Борзяк, В.Я.Бочаров, Л.И.Волкова и др.; под редакцией М.П.Сапина. //М.: Медицина – 1987 – Т.1 – 388с.
7. M. A. Ferrer, M. Diaz, C. Carmona-Duarte, and A. Morales. A behavioral handwriting model for static and dynamic signature synthesis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6):1041–1053, 2017
8. Pirlo G, Cuccovillo V, Diaz-Cabrera M, Impedovo D, Mignone P (2015) Multidomain verification of dynamic signatures using local stability analysis. IEEE Trans Hum Mach Syst 45(6): 805–810.
9. Иванов А.И. Разработка и исследование систем идентификации личности по динамике подписи. /Иванов А.И., Сорокин И.А, Клюхов Ю.С. //Отчет по аванпроекту. НЖИРЭТ- 1993-150 с. инв. № 748.
10. Иванов А.И. Автоматическая система идентификации личности по динамике подписи. / А.И.Иванов, И.А.Сорокин //Новые промышленные тех-

нологии. – 1993-№ 6 – С. 56-63.

11. Plomondon R., Lorette G. Automatic signature verification and writer identification – the state of the art. //Pattern Recognition 1989-22, N2, p. 107-131.

12. Сорокин И.А. Анализ траектории ключевого слова в задаче биометрической аутентификации. /И.А.Сорокин, А.И.Иванов //Специальная техника средств связи. Серия. Системы, сети и технические средства конфиденциальной связи. Пенза: ПНЮИ – 1997 – ВыпускХо2.-С.108-113.

13. Шарипов Р.Р. Разработка полигауссового алгоритма аутентификации пользователей в телекоммуникационных системах и сетях по клавиатурному почерку: дис. канд. техн. наук: 05.12.13 – «Системы, сети и устройства телекоммуникаций». – Казань, 2006. – 135 с.

14. Тушканов Е.В. Разработка методов и алгоритмов повышения защищенности информации на основе анализа клавиатурного почерка: автореф. дис. канд. техн. наук: 05.13.19 – «Методы и системы защиты информации, информационная безопасность». – Санкт-Петербург, 2013. – 128 с.

15. Савинов А.Н. Методы, модели и алгоритмы распознавания клавиатурного почерка в ключевых системах: дис. канд. техн. наук: 05.13.19 – «Методы и системы защиты информации, информационная безопасность». – Санкт-Петербург, 2013. – 128 с.

16. Брюхомицкий Ю.А. Гистограммный метод распознавания клавиатурного почерка: статья / Ю.А. Брюхомицкий. – Известия Южного федерального университета. Технические науки, №11, т.112, 2010. – 8 с.

17. Костюченко Е.Ю. Идентификация непрерывных биометрических сигналов на основе нейронных сетей: дис. канд. техн. наук: 05.13.01 – «Системный анализ, управление и обработка информации (отрасль: промышленность)». – Томск, 2010. – 171 с.

18. Doddington G.R., Flanagan G.L., Lummis R . С Automatic speaker verification by non-linear time alignment of acoustic parameters. Патент США №3700815, 1972.

19. Limmis R.C. Speaker verification by computer using speech intensity for

temporal registration. //IEEE Trans. 1973, v. AU-21, N2.

20. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь – 1981 – 224 с.

21. Chao Shen et. al User Authentication Through Mouse Dynamics // IEEE Transactions on Information Forensics and Security, Volume: 8, Issue: 1, Jan. 2013, pp. 16–30.

22. E. Ahmed and I. Traore. A new biometric technology based on mouse dynamics // IEEE Trans. Depend. Secure Comput., vol. 4, no. 3, pp. 165–179, Jul.–Sep. 2007.

23. R. Everitt and P.W. McOwan. Java-Based Internet Biometric Authentication System // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9): 1166-1172.

24. R.V. Yampolskiy and V. Govindaraju. Behavioral biometric: a survey and classification. International Journal of Biometrics, 2008, 1(1): 81-113.

25. S. Hashia, C. Pollett, and M. Stamp. On using mouse movements as a biometric // Proceeding of International Conference on Computer Science and its Applications, volume 1, 2005.

26. P. Bours and C. J. Fullu. A login system using mouse dynamics // Fifth International Conference of Intelligent Information Hiding and Multimedia Signal Processing, 2009:1072-1077.

27. H. Gamboa, A.L.N. Fred, and A.K.Jain. Web biometrics: user verification via web interaction // Biometrics Symposium, 2007:1-6.

28. K. Revett, H. Jahankhani, S. T. de Magalhes, and H. M. D. Santos. A survey of user authentication based on mouse dynamics // Proceedings of 4th International Conference on Global E-Security, pages 210-219, London, June 2008.

29. Balabit Mouse Dynamics Challenge Data Set [Электронный ресурс] – Режим доступа до ресурсу: <https://github.com/balabit/Mouse-Dynamics-Challenge>. Дата звернення: 01.09.2020.

30. Mouse Dynamics – Measurements on the Balabit Data Set [Электронный ресурс] – Режим доступа до ресурсу: <https://ms.sapientia.ro/>

[~manyi/mousedynamics](https://github.com/~manyi/mousedynamics). Дата звернення: 01.09.2020.

31. Харрисон М. Машинное обучение: карманный справочник. Краткое руководство по методам структурированного машинного обучения на Python.: Пер. с англ. – СПб.: ООО "Диалектика", 2020 - 320 с.

32. Бринк Х. Машинное обучение : пер. с англ. / Х. Бринк, Д. Ричардс, М. Феверолф. – Санкт-Петербург [и др.]: Питер, 2018.

33. Барсегян А. А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. – 3-е изд., перераб. доп. – СПб.: БХВ-Петербург, 2009. – 512 с.