Міністерство освіти і науки України

Харківський національний університет імені В.Н. Каразіна

Факультет комп'ютерних наук

Спеціальність 125 «Кібербезпека»

Освітня програма «Кібербезпека»

«Допущено до захисту»

В.о. завідувача кафедрою БІСТ

Мелкозьорова О.М.

_____

« » червня 2024р.

Пояснювальна записка

до кваліфікаційної роботи бакалавра

на тему: «Моделювання та дослідження інтелектуальних систем автентифікації з використанням згорткових нейронних мереж та нечітких екстракторів»

оцінка « »

Голова ЕК

Лемешко О.В. _____

Керівник проф. Кузнецов О.О.

(прізвище та ініціали/підпис)

Рецензент _____

(прізвище та ініціали/підпис)

Виконавець студент групи КБ-42

Кононенко М.О.

(прізвище та ініціали/підпис)

Харків – 2024

РЕФЕРАТ

Дипломна робота містить 62 сторінки, 7 рисунків, 6 графіків, 7 таблиць, 7 додатків, 66 посилань на джерела. Додатки містять додаткові матеріали та програмний код.

Метою дипломної роботи є аналіз можливостей застосування згорткових нейронних мереж та нечітких екстракторів у системах автентифікації, а також розробка моделі інтелектуальної системи автентифікації на їх основі. Робота зосереджується на дослідженні теми автентифікації та її видів, ефективності згорткових нейронних мереж у розпізнаванні біометричних патернів та можливостях нечітких екстракторів у забезпеченні надійної автентифікації.

Робота присвячена розробці концептуальної моделі застосунку, в який включена система автентифікації, яка інтегрує згорткові нейронні мережі для аналізу біометричних даних та нечіткі екстрактори для генерації та перевірки унікальних ідентифікаторів, після чого створює секретні ключі для безпечного обміну даними між користувачем та системою. Розробка моделі спрямована на підвищення рівня безпеки автентифікаційних процесів, забезпечуючи високу точність ідентифікації особи на основі її біометричних характеристик і створюючи надійні механізми захисту ідентифікаційних даних. Особлива увага приділяється аналізу та оптимізації алгоритмів згорткових нейронних мереж та методів нечіткої логіки, що дозволяє ефективно працювати з неточними або неповними даними, характерними для біометричних систем. Досліджується ефективність такої системи у різних сценаріях використання та її спроможність протистояти сучасним кіберзагрозам.

Актуальність теми дослідження виходить з необхідності розробки ефективних методів автентифікації для забезпечення безпеки інформаційних систем у сучасному цифровому світі. Розвиток технологій штучного інтелекту

та зростаюче використання біометричних даних вимагають нових підходів до автентифікації, які зможуть ефективно протистояти сучасним загрозам безпеці.

В рамках дипломної роботи буде проведене детальне дослідження сучасних методів автентифікації, включаючи традиційні парольні системи, двофакторну автентифікацію, а також інноваційні біометричні методи. Окрема увага приділяється можливостям застосування штучного інтелекту для аналізу біометричних даних, що відкриває нові перспективи для створення більш ефективних і безпечних систем автентифікації.

Ключові слова: ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ АВТЕНТИФІКАЦІЇ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, НЕЧІТКІ ЕКСТРАКТОРИ, БЕЗПЕКА ІНФОРМАЦІЇ, БІОМЕТРИЧНІ ДАНІ.

ABSTRACT

The thesis contains 62 pages, 7 figures, 6 graphs, 7 tables, 7 appendices, 66 references to sources. Appendices contain additional materials and software code.

The aim of the thesis is to analyze the possibilities of using convolutional neural networks and fuzzy extractors in authentication systems, as well as to develop a model of an intelligent authentication system based on them. The work focuses on researching the topic of authentication and its types, the effectiveness of convolutional neural networks in biometric pattern recognition, and the capabilities of fuzzy extractors in providing reliable authentication.

The work is devoted to the development of a conceptual model of the application, which includes an authentication system that integrates convolutional neural networks for the analysis of biometric data and fuzzy extractors for the generation and verification of unique identifiers, after which it creates secret keys for secure data exchange between the user and the system. The development of the model is aimed at increasing the level of security of authentication processes, ensuring high accuracy of identification of a person based on his biometric characteristics and creating reliable mechanisms for the protection of identification data. Special attention is paid to the analysis and optimization of algorithms of convolutional neural networks and methods of fuzzy logic, which allows effective work with imprecise or incomplete data, characteristic of biometric systems. The effectiveness of such a system in various usage scenarios and its ability to resist modern cyber threats are investigated.

The relevance of the research topic stems from the need to develop effective authentication methods to ensure the security of information systems in the modern digital world. The development of artificial intelligence technologies and the growing use of biometric data require new approaches to authentication that can effectively counter modern security threats.

The thesis will conduct a detailed study of modern authentication methods, including traditional password systems, two-factor authentication, as well as innovative biometric methods. Particular attention is paid to the possibilities of applying artificial intelligence to the analysis of biometric data, which opens up new perspectives for the creation of more efficient and secure authentication systems.

Keywords: INTELLIGENT AUTHENTICATION SYSTEMS, CONVOLUTIONAL NEURAL NETWORKS, Fuzzy EXTRACTORS, INFORMATION SECURITY, BIOMETRIC DATA.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

AI – Штучний інтелект;

ЗНМ – Згорткові нейронні мережі;

НЕ – Нечіткі екстрактори;

БД – Біометричні дані;

АС – Автентифікаційні системи;

БСА – Біометричні системи автентифікації;

ОФА – Однофакторна автентифікація;

2FA – Двофакторна автентифікація;

MFA – Багатофакторна автентифікація;

ЕП – Електронний підпис;

ЦС – Цифрові сертифікати;

МН – Машинне навчання;

БЗ – Безпарольні засоби;

РСА – Розподілені системи автентифікації;

БТ – Блокчейн технології;

ПС – Парольні системи;

PIN – Персональний ідентифікаційний номер;

RFID – Радіочастотна ідентифікація;

SC – Смарт-карти (Smart Cards).

CSC – Секретний номер картки (Card Security Code)

SSL – Протокол захищених сокетів;

TLS – Протокол транспортного рівня безпеки;

HTTP – Протокол передачі гіпертексту;

HTTPS – Безпечний протокол передачі гіпертексту;

API – Інтерфейс програмування додатків;

ВСТУП

У 2024 році, на фоні все більшого проникнення цифрових технологій у наше повсякденне життя, питання захисту особистих даних і цифрових систем стає дедалі актуальнішим. Збільшення кількості кібератак, зломів і випадків витоку конфіденційної інформації спонукає наукову спільноту і розробників постійно шукати нові, більш ефективні способи захисту. Одним із ключових елементів забезпечення цифрової безпеки є процес автентифікації, який дозволяє перевіряти особу на основі унікальних ідентифікаторів або характеристик.

Наукові прогреси у сфері штучного інтелекту та машинного навчання сьогодні відкривають нові можливості для вдосконалення методів автентифікації. Зокрема, використання біометричних даних та впровадження адаптивних алгоритмів, здатних до самонавчання та адаптації в змінних умовах, привертає значну увагу. Це дозволяє створювати високоефективні системи безпеки, що відрізняються високою точністю та надійністю.

Однак, з постійним розвитком кіберзагроз та ускладненням методів кібератак, розробники систем автентифікації змушені постійно вдосконалювати свої підходи, роблячи їх більш гнучкими та надійними. У цьому контексті, дослідження та розробка інноваційних систем автентифікації, які базуються на згорткових нейронних мережах та алгоритмах нечіткої логіки, набувають особливої важливості. Ці системи здатні не тільки ефективно захищати дані, але й пристосовуватися до нових умов використання, виявляючи потенційні загрози на основі аналізу поведінкових патернів користувачів [4]. Такий підхід дозволяє створювати майбутнє цифрової безпеки, де системи автентифікації стають більш інтелектуальними, надійними та зручними у використанні.

1 АВТЕНТИФІКАЦІЯ: ФУНДАМЕНТАЛЬНІ ПРИНЦИПИ, МЕТОДИ ТА ВИКЛИКИ В ЗАБЕЗПЕЧЕННІ ЦИФРОВОЇ БЕЗПЕКИ

1.1 Визначення автентифікації, цілі та завдання автентифікаційних систем, розмежування між автентифікацією, авторизацією та ідентифікацією

Автентифікація виконує важливу роль у забезпеченні безпеки системи, обмежуючи доступ до ресурсів лише для підтверджених користувачів чи пристроїв. Методи автентифікації варіюються від простих, таких як введення пароля, до складніших, які включають біометричні дані (відбитки пальців, сканування сітківки ока), а також використання цифрових сертифікатів і одноразових паролів.[7]

1.1.1 Визначення автентифікації

Автентифікація представляє процес верифікації особи чи пристрою на основі даних, які підтверджують їхню справжність. В основі цього процесу лежить ідентифікація або підтвердження ідентичності суб'єкта за допомогою унікальних характеристик або даних, які відомі лише цьому суб'єкту або про нього [9].

Залежно від сценарію застосування, методи автентифікації мають відрізнятися за рівнем складності та безпеки. Вибір конкретного методу обумовлюється необхідністю знайти баланс між зручністю користувача та вимогами безпеки. Основною метою автентифікації є забезпечення конфіденційності, цілісності та доступності даних, водночас мінімізуючи ризик несанкціонованого доступу до інформаційних ресурсів.

1.1.2 Цілі та завдання автентифікаційних систем

Автентифікаційні системи були створені для забезпечення доведення ідентичності користувачів або пристроїв перед наданням доступу до інформаційних ресурсів [11, 12].

Основні цілі включають забезпечення конфіденційності, цілісності, та доступності інформації, що зумовлює важливість ретельного підходу до вибору та реалізації методів автентифікації.

Однією з ключових завдань автентифікаційних систем є забезпечення надійного механізму верифікації, що дозволяє з упевненістю ідентифікувати користувача або пристрій. Це включає розробку процесів, які можуть ефективно протистояти спробам несанкціонованого доступу, використовуючи підроблені або вкрадені дані [13].

1.1.3 Розмежування: автентифікація

Автентифікація передбачає підтвердження заявленої ідентичності. Під час цього процесу система перевіряє надані користувачем докази ідентичності, такі як пароль, біометричні дані, або токен, з метою переконатися, що користувач є тим, за кого себе видає [14].

1.1.4 Розмежування: авторизація

Авторизація настає після успішної автентифікації та визначає, до яких ресурсів чи операцій має доступ користувач. Цей процес включає призначення прав та привілеїв користувачу на основі його ролі чи атрибутів і визначає, що користувач може робити всередині системи [12].

1.1.5 Розмежування: ідентифікація

Ідентифікація є процесом, під час якого користувач представляє або стверджує свою ідентичність системі, найчастіше за допомогою імені користувача або електронної пошти. Цей етап не включає перевірку заявленої

ідентичності; він лише дає знати системі, ким користувач стверджує себе бути [10].

1.1.6 Приклади застосування ідентифікації, автентифікації та авторизації

В системі електронної пошти користувач ідентифікує себе, вводячи адресу електронної пошти. Потім він автентифікується, вводячи пароль, пов'язаний з цією адресою. Нарешті, він отримує авторизацію для доступу до своєї електронної скриньки, відправлення, та отримання повідомлень.

В корпоративній мережі співробітник спочатку ідентифікує себе за допомогою свого корпоративного логіна. Автентифікація відбувається за допомогою пароля або біометричних даних. За результатами автентифікації система виконує авторизацію, надаючи доступ до специфічних файлів та програм згідно з рівнем привілеїв співробітника [15].

1.2 Аналіз факторів автентифікації

Автентифікаційні системи класифікують фактори автентифікації на основі трьох основних категорій: знання, володіння та індивідуальності. Кожна з цих категорій має свої характеристики та застосування, спрямовані на підвищення безпеки ідентифікаційних процесів [16].

1.2.1 Визначення фактору знання

Фактор знання є найпоширенішим типом автентифікації і заснований на інформації, яку знає тільки користувач. Типовими прикладами є паролі, ПІН-коди та відповіді на секретні питання. Ці фактори легко реалізувати та використовувати, однак вони можуть бути вразливими до вгадування або перехоплення [22].

1.2.2 Визначення фактору володіння

Фактор володіння включає елементи, які фізично належать користувачеві, наприклад, смарт-картки, USB-ключі, мобільні телефони або інші токени. Автентифікація здійснюється за допомогою пред'явлення або використання цих предметів. Фактор володіння забезпечує вищий рівень безпеки порівняно з фактором знання, оскільки вимагає фізичного доступу до автентифікаційного засобу [9].

1.2.3 Визначення фактору індивідуальності

Фактор індивідуальності базується на унікальних фізичних або поведінкових характеристиках особи, таких як відбитки пальців, розпізнавання обличчя, голосу, візерунок райдужки ока або підпис. Біометричні методи автентифікації є найбільш надійними з усіх трьох типів, оскільки вони майже неможливо відтворити або вкрасти. Однак вони можуть вимагати спеціалізованого обладнання для зчитування та обробки біометричних даних.

1.3 Методології та інструментарій сучасної автентифікації

Автентифікація, залежно від кількості використовуваних факторів, класифікується на однофакторну, двофакторну, та багатофакторну. Кожен із цих методів відіграє важливу роль у забезпеченні безпеки, однак відрізняється за рівнем захисту та зручності використання.

1.3.1 Аналіз автентифікаційних методологій

Класифікація автентифікаційних методів заснована на використанні різних типів доказів або факторів, які можуть включати знання (щось, що користувач знає, наприклад, пароль), володіння (щось, що користувач має, наприклад, мобільний телефон або токен) та індивідуальність (біометричні характеристики, такі як відбитки пальців або риси обличчя). Розрізняють

однофакторну, двофакторну та багатофакторну автентифікацію, залежно від кількості та типів використовуваних доказів.

1.3.2 Аналіз однофакторної автентифікації

Однофакторна автентифікація вимагає від користувача подання лише одного доказу своєї легітимності, найчастіше у формі пароля. Цей метод є найпростішим і найшвидшим у використанні, але водночас найменш безпечним, оскільки доступ може бути отриманий через вгадування або викрадення пароля [17].

1.3.3 Аналіз двофакторної автентифікації

Двофакторна автентифікація посилює безпеку шляхом використання двох різних типів доказів, які зазвичай включають щось, що користувач знає (пароль), та щось, що користувач має (токен або мобільний телефон). Цей підхід значно ускладнює несанкціонований доступ, оскільки атакуючому потрібно буде володіти обома факторами. 2FA є стратегією безпеки, яка забезпечує подвійний рівень захисту, вимагаючи від користувачів надати два різних типи ідентифікаційних даних перед наданням доступу до ресурсу. Цей метод ефективно мінімізує ризик компрометації облікових даних, оскільки навіть у разі витоку або вгадування пароля, несанкціонований доступ до системи залишається обмеженим без другого фактора. Зазвичай, другим фактором є одноразовий пароль, що надсилається на мобільний телефон користувача, або використання фізичного токена [18].

Багатофакторна автентифікація включає використання двох і більше факторів, що можуть також включати індивідуальні особливості, такі як біометричні дані. Цей метод забезпечує найвищий рівень безпеки, вимагаючи від користувача декількох форм підтвердження. Багатофакторна автентифікація застосовується в критично важливих системах, де потрібен високий рівень захисту інформації.

MFA, з іншого боку, розширює поняття 2FA, включаючи три або більше незалежних факторів для підтвердження ідентичності користувача. Це може охоплювати комбінації знань (наприклад, пароль або CSC), володіння (наприклад, смарт-картка або мобільний телефон) та індивідуальності (наприклад, відбитки пальців або розпізнавання обличчя). MFA надає значно міцніший захист, оскільки ймовірність того, що атакуючий одночасно скомпрометує кілька різних типів кредитів, є надзвичайно низькою [23].

1.3.4 Аналіз використання біометричних даних у процесах автентифікації

Біометрична автентифікація представляє собою метод ідентифікації особи на основі одного або декількох фізіологічних або поведінкових ознак.

Основні види біометричних методів включають відбитки пальців, розпізнавання облич, сканування сітківки ока, голосову ідентифікацію, та аналіз підпису. Кожен з цих методів має унікальні особливості, що дозволяють точно ідентифікувати особу [19].

Переваги використання біометрії полягають у високому рівні безпеки та зручності. Оскільки біометричні дані унікальні для кожної особи, вони значно ускладнюють несанкціонований доступ до систем або пристроїв.

Також, використання біометричних методів зменшує необхідність запам'ятовувати паролі або носити з собою фізичні ключі.

Проте, використання біометрії також має недоліки, які включають питання приватності та потенціал для помилкових спрацьовувань. Занепокоєння з приводу приватності виникають через необхідність зберігання та обробки особистих біометричних даних. Також, існує ризик помилкового відхилення легітимного користувача або неправильного підтвердження особи несанкціонованою особою через обмеження технології або варіабельність біометричних даних [20].

1.3.5 Аналіз електронного підпису і цифрових сертифікатів

Електронний підпис являє собою набір електронних даних, що додаються до електронного документа або повідомлення та використовуються для ідентифікації підписувача та підтвердження цілісності документа. Електронний підпис має важливе правове значення, оскільки в багатьох юрисдикціях він рівносильний традиційному рукописному підпису та може використовуватися у судовому порядку для підтвердження автентичності документів [8].

Цифрові сертифікати відіграють ключову роль у процесах автентифікації, дозволяючи встановлювати довірені зв'язки між користувачами та системами. Цифровий сертифікат — це вид документа, що виданий надійною третьою стороною (наприклад, центром сертифікації) і містить інформацію про ключ власника та підтвердження його автентичності. Сертифікати використовуються для підтвердження справжності публічного ключа, асоційованого з особою, організацією або пристроєм, і забезпечують безпеку цифрових взаємодій шляхом захисту від атак "людина посередині" та інших форм шахрайства [21].

1.4 Стандарти, методи та протоколи автентифікації

Огляд основних стандартів та протоколів демонструє широкий спектр інструментів, доступних для забезпечення автентифікації та обміну даними. Серед важливих стандартів слід виділити OAuth, що забезпечує делегування доступу до ресурсів без необхідності розкриття пароля, та OIDC, який дозволяє автентифікацію користувача через довірені сервери. Протокол SAML використовується для обміну автентифікаційними та авторизаційними даними між доменами.

1.4.1 Аналіз стандарту OAuth

OAuth є відкритим стандартом для делегування доступу, що дозволяє користувачам надавати безпечний делегований доступ до своїх ресурсів на

одному сайті за допомогою ідентифікації на іншому сайті, без необхідності передавати свої логіни та паролі. Основна ідея OAuth полягає в тому, щоб надати додатку обмежений доступ до ресурсу користувача з можливістю отримання та відправлення інформації від імені користувача без використання його основних облікових даних [24].

OAuth використовує токени доступу, які служать для авторизації користувачів та додатків для взаємодії з ресурсами без необхідності відкривати пароль. Такий підхід забезпечує додатковий рівень безпеки, оскільки токени можуть бути обмежені у часі дії, в обсязі доступних даних, а також можуть бути відкликані користувачем у будь-який час.

1.4.2 Аналіз стандарту OIDC (OpenID Connect)

OIDC – це протокол ідентифікації, побудований на основі стандарту OAuth 2.0, який дозволяє клієнтським додаткам перевіряти ідентичність користувача на основі автентифікації, проведеної службою автентифікації. OpenID Connect розширює можливості OAuth, надаючи додатковий шар ідентифікації. Це робиться шляхом введення поняття «ID токена», який містить інформацію про автентифікацію сеансу та дані про користувача.

OpenID Connect використовується для створення систем, де користувач може увійти за допомогою одного облікового запису для доступу до різноманітних додатків, спрощуючи процес управління ідентичностями та доступом [26].

1.4.3 Аналіз стандарту SAML (Security Assertion Markup Language)

SAML є стандартом для обміну даними автентифікації та авторизації між різними безпековими доменами. SAML базується на XML і дозволяє провайдерам ідентифікаційних даних передавати інформацію про авторизацію провайдерам послуг. Завдяки SAML користувачі можуть виконувати вхід один раз і отримувати доступ до послуг у різних доменах без повторної автентифікації, що відомо як SSO.

SAML визначає три ролі [25]:

- принципала (зазвичай користувача);
- провайдера ідентифікаційних даних (IdP);
- провайдера послуг (SP).

IdP виступає як орган, який автентифікує ідентичність користувача і відправляє SAML твердження SP. SP, в свою чергу, надає або відмовляє доступ користувачу на основі цього твердження. Цей процес підвищує безпеку, зменшуючи втому від паролів серед користувачів та мінімізуючи шанси на фішинг.

1.5 Методи встановлення автентифікації

Методи встановлення автентифікації еволюціонували від простих до складних, відповідаючи на зростаючі вимоги до безпеки інформаційних систем. Найпростішим і найдавнішим методом є використання статичних паролів, що, незважаючи на свою поширеність, мають значні недоліки з точки зору безпеки через вразливість до вгадування, фішингу та інших видів атак.

Розвиток технологій сприяв появі двофакторної та багатофакторної автентифікації, які забезпечують підвищений рівень захисту шляхом комбінування декількох незалежних критеріїв: щось, що користувач знає (ПС), щось, що він має (токен або SC), або щось, що є невід'ємною частиною його особистості (БСА) [5].

Криптографічні методи, включаючи цифрові підписи та сертифікати, надають надійну основу для автентифікації в цифровому світі, дозволяючи забезпечити цілісність та неперервність даних, а також підтвердження справжності джерела інформації [26].

1.6 Принципи побудови автентифікаційних систем

Принципи побудови автентифікаційних систем вимагають глибокого розуміння потреб безпеки та можливих загроз. Найкращі практики проектування акцентують на необхідності розробки комплексних, але зручних

у використанні систем, що забезпечують надійну захист від несанкціонованого доступу [8].

Першочергово, система автентифікації має бути розроблена з урахуванням принципу мінімальних привілеїв, що забезпечує користувачам доступ лише до тих ресурсів та даних, які необхідні для виконання їхніх завдань. Такий підхід допомагає зменшити потенційну шкоду від можливих вразливостей або зловмисних дій.

Системи повинні бути розроблені з врахуванням потенційних загроз та включати механізми їхнього запобігання. Це включає регулярний аналіз безпеки, виявлення вразливостей, оновлення програмного забезпечення, а також реалізацію ефективних стратегій реагування на інциденти [27].

1.7 Висновки до розділу

У першому розділі дипломної роботи здійснено всеосяжний аналіз фундаментальних принципів, методів та викликів, які виникають при забезпеченні цифрової безпеки через автентифікацію. Розділ включає детальне розглядання різниці між автентифікацією, авторизацією та ідентифікацією, а також огляд різних факторів автентифікації, які включають знання, володіння та індивідуальність.

Розглянуто роль біометричних даних та цифрових сертифікатів у посиленні безпеки, аналізовано стандарти та протоколи, які визначають сучасні підходи до автентифікації, включаючи OAuth, OpenID Connect та SAML, приділено увагу аналізу сучасних методологій автентифікації, включаючи однофакторну, двофакторну та багатофакторну системи.

Висновки підкреслюють необхідність комплексного підходу до розробки і впровадження автентифікаційних технологій, що включають як технічні, так і правові аспекти, з метою забезпечення високого рівня безпеки та дотримання прав людини у цифровому світі.

2 ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ ДЛЯ БІОМЕТРИЧНОЇ АВТЕНТИФІКАЦІЇ

2.1 Принцип роботи згорткових нейронних мереж

Архітектура згорткових нейронних мереж (CNN) складається з декількох типів шарів, кожен з яких виконує специфічні функції у процесі обробки зображень та відео. Ключові елементи такої архітектури включають згорткові шари, шари пулінгу та повнозв'язні шари.

2.1.1 Згорткові нейронні мережі

CNN — це клас глибоких нейронних мереж, які широко використовуються в задачах обробки зображень та відео завдяки своїй здатності ефективно вловлювати просторові та тимчасові залежності. Цей тип мережі відрізняється від звичайних повнозв'язних нейронних мереж архітектурою, що включає згорткові шари, шари пулінгу та повнозв’язні шари [29, 52].

2.1.2 Згорткові шари

Згорткові шари є фундаментальними в CNN та призначені для виявлення візуальних особливостей на зображенні на різних рівнях складності. Ці шари використовують набори вагових фільтрів, які автоматично адаптуються під час навчання мережі. Фільтри "прокручуються" через вхідне зображення (або попередню карту особливостей), виконуючи операцію згортки та утворюючи карту активації, яка вказує, де певні особливості з'являються на зображенні [33].

Згорткові шари використовують набір навчувальних фільтрів, які автоматично виявляють важливі характеристики на різних рівнях абстракції. Вони використовують фільтри або ядра, які згортаються над вхідним зображенням для створення карт активації. Карти активації ілюструють, де

певні особливості, такі як краї або текстури, з'являються на зображенні. На рисунку 2.1 наданий перелік різних варіантів використання згроткових шарів.

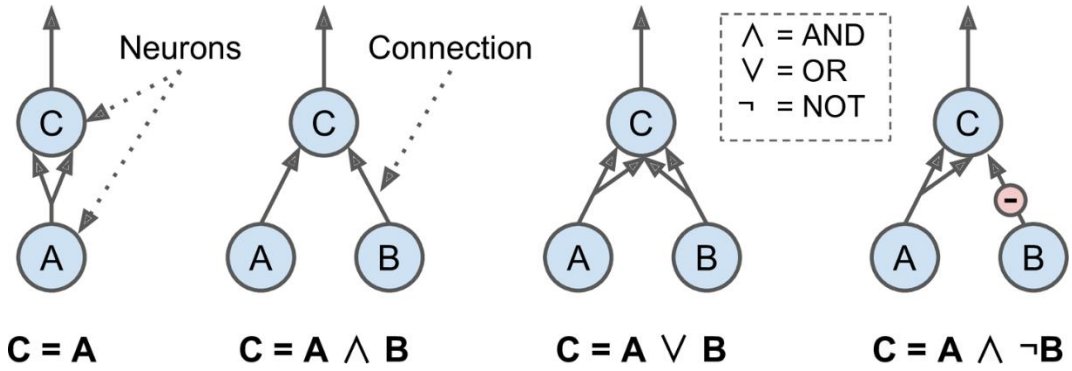


Рисунок 2.1 – Варіанти згорткових шарів.

Процес згортки дозволяє мережі автоматично вчитися визначати та виділяти ці особливості на різних рівнях складності, що сприяє глибшому розумінню зображення.Кожен фільтр у згортковому шарі активується, коли він розпізнає певний тип особливості на зображенні, такий як краї, кути або текстури. Після застосування фільтра утворюється карта особливостей, яка подається на наступний шар [33].

2.1.3 Шари пулінгу

Шари пулінгу, які зазвичай розміщуються між згортковими шарами, допомагають зменшити розмірність вхідних даних та підвищити інваріантність мережі до невеликих зміщень та перекосів у зображенні. Це досягається шляхом агрегації даних в більш великих регіонах (наприклад, вибору максимального або середнього значення). На рисунку 2.2 описаний процес роботи пулінгу з 16 клітинок перетворення у 4.

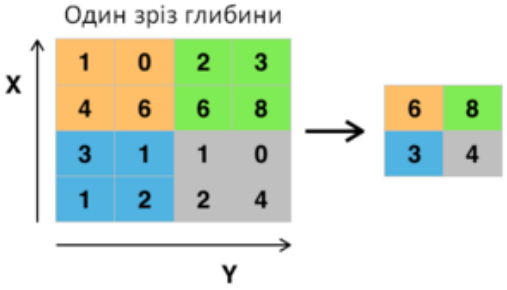


Рисунок 2.2 – Процес роботи шарів пулінгу.

Це виконується шляхом об'єднання виходів згорткового шару в менші області. Найпоширеніші типи пулінгу включають максимальний та середній пулінг, які вибирають відповідно найбільше або середнє значення з кожної підобласті [30].

2.1.4 Повнозв'язні шари

Повнозв'язні шари розташовуються в кінці архітектури CNN. Всі вхідні особливості з попередніх шарів трансформуються до вектора (зазвичай через процес званий "сплющуванням"), і кожний нейрон у повнозв'язному шарі має з'єднання з усіма активаціями попереднього шару. Ці шари інтегрують виявлені особливості для виконання класифікації або інших завдань. Вони обробляють дані, зібрані з усіх карт особливостей, для вироблення кінцевого результату або прогнозу [52]. На рисунку 2.3 наданий приклад використання нейронної мережі та її вихідне значення.

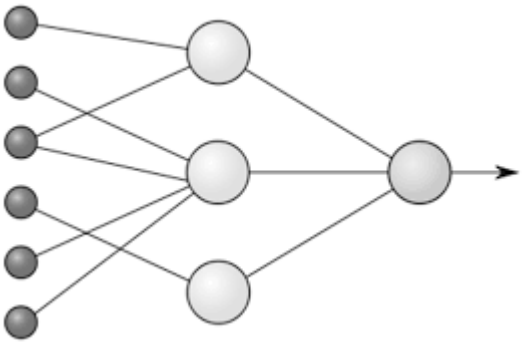


Рисунок 2.3 – Вихід нейронної мережі.

Вони отримують вхідні дані з попередніх шарів, які зазвичай бувають "сплющені" у вектор, і використовують їх для визначення, до якої категорії належить вхідне зображення. Ці шари мають з'єднання з усіма активаціями з попереднього шару, що дозволяє мережі враховувати всі навчені особливості при прийнятті рішення.

2.2 Датасети для тренування біометричних систем

Для тренування біометричних систем автентифікації використовуються спеціалізовані датасети, що містять біометричні дані, такі як відбитки пальців, обличчя, іриси очей, та голосові зразки. Наявність якісних та різноманітних датасетів є критичною для розробки ефективних і надійних систем.

2.2.1 Labeled Faces in the Wild

LFW — це база даних фотографій облич, призначена для вивчення проблеми необмеженого розпізнавання облич. Цю базу даних створили та підтримували дослідники з Університету Массачусетса, Амгерст (конкретні посилання наведено в розділі Подяки).

13 233 зображення 5 749 людей були виявлені та відцентровані детектором облич Viola Jones і зібрані з Інтернету. 1680 із зображених людей мають дві або більше різних фотографій у наборі даних.

Оригінальна база даних містить чотири різні набори зображень LFW, а також три різні типи «вирівняних» зображень. За словами дослідників, глибоко спрямовані зображення дали кращі результати для більшості алгоритмів перевірки обличчя порівняно з іншими типами зображень.

2.2.2 CASIA Iris Image Database

CASIA-IrisV4 містить загалом 54 601 зображення райдужної оболонки більш ніж 1800 справжніх і 1000 віртуальних об’єктів. Усі зображення райдужної оболонки є 8-бітними файлами JPEG із рівнем сірого, зібраними під ближнім інфрачервоним освітленням або синтезованими. Деякі статистичні дані та особливості кожної підмножини наведено в таблиці 1. Шість наборів даних були зібрані або синтезовані в різний час, і CASIA-Iris-Interval, CASIA-Iris-Lamp, CASIA-Iris-Distance, CASIA-Iris-Thousand можуть мати невелике збіг суб'єктів між підгрупами.

2.2.3 FERET Facial Recognition Technology Database

Управління програми розвитку технологій боротьби з наркотиками Міністерства оборони (DoD) фінансувало програму технології розпізнавання облич (FERET). Мета програми FERET полягала в тому, щоб розробити можливості автоматичного розпізнавання облич, які можна було б використовувати для допомоги персоналу служби безпеки, розвідки та правоохоронних органів у виконанні їхніх обов'язків.

Програма FERET розпочалася у вересні 1993 року, коли доктор П. Джонатон Філліпс, Армійська науково-дослідна лабораторія, Адельфі, штат Меріленд, виконував обов'язки технічного агента. Спочатку програма FERET складалася з трьох етапів, кожна тривалістю один рік.

2.2.4 The VoxCeleb Dataset

VoxCeleb2 містить понад 1 мільйон висловлювань 6112 знаменитостей, взятих із відео, завантажених на YouTube. Набір для розробки VoxCeleb2 не збігається з ідентифікаторами в наборах даних VoxCeleb1 або SITW. В таблиці 2.1 наданий опис кількості висловлювань на тестування та розробку нейронної мережі

Таблиця 2.1 – Варіанти датасету VCD

	Розробка	Тестування
Кількість спікерів	5994	118
Кількість відео	145569	4911
Кількість висловлювань	1092009	36237

2.2.5 NIST Biometric Databases

У зв'язку з потребою покращити національну безпеку біометрику було визначено як ключову сприятливу технологію. NIST (через свою місію та досвід роботи) підтримує зусилля уряду збільшити збір якісних біометричних

даних, забезпечити належний обмін зібраними даними з іншими агентствами, а також забезпечити точність і сумісність біометричних систем.

Біометрична діяльність NIST включає:

- дослідження різних біометричних модальностей: відбитків пальців, обличчя, райдужної оболонки ока, голосу, ДНК та мультимодальних;
- розробка стандартів на національному та міжнародному рівнях;
- а також тестування та оцінка технологій, що призводить до інновацій.

2.3 Критерії вибору даних для ефективного навчання мереж

Ефективне навчання згорткових нейронних мереж значною мірою залежить від якості та відповідності використовуваних даних. Вибір правильних даних вимагає розгляду кількох критеріїв, що забезпечують високу точність та надійність отриманих результатів.

2.3.1 Репрезентативність

Датасет має відображати реальні умови застосування системи. Це означає, що в ньому повинні бути представлені дані з різноманітних контекстів та від різних груп користувачів, щоб мережа могла ефективно адаптуватися до різних ситуацій та особливостей застосування [35].

2.3.2 Баланс класів

Для уникнення упередженості моделі важливо мати збалансоване представлення класів у датасеті. Нерівномірне представлення класів може призвести до перекосу в навчанні та зниження загальної точності системи.

2.3.3 Якість даних

Висока роздільна здатність та чистота зразків є критично важливими, особливо для біометричних даних, де деталі можуть вплинути на точність ідентифікації. Шуми, спотворення та інші артефакти у даних повинні бути мінімізовані [57].

2.3.4 Різноманітність

Інклюзивність даних, що охоплює різні вікові групи, етнічність, гендерні та інші біологічні відмінності, є важливою для створення універсальних систем автентифікації. Варіативність умов освітлення, позиціонування та інших експлуатаційних факторів також слід враховувати [53].

2.3.5 Обсяг даних

Для глибокого навчання потрібні великі обсяги даних. Чим більший і більш різноманітний датасет, тим краще мережа зможе узагальнювати нові дані під час реального застосування.

2.3.6 Валідаційний та тестовий набори

Наявність окремих валідаційних та тестових наборів даних дозволяє оцінювати якість моделі об'єктивно та забезпечити її здатність до узагальнення на нових даних, що є ключовим для виходу системи в продуктив [34].

2.4 Процес та проблеми навчання нейронних мереж

Процес навчання нейронних мереж починається з ретельної підготовки та попередньої обробки даних, що є критично важливими для забезпечення якісного тренування моделі.

2.4.1 Проблеми збалансованості даних та їхнє вплив на точність моделей

Незбалансованість даних виникає, коли деякі класи або категорії представлені в датасеті значно краще або гірше за інші. Це може призвести до того, що модель вчиться віддавати перевагу частіше представленим класам і відповідно гірше працює з менш представленими класами.

Коли датасет має нерівномірне представлення класів, навчена на таких даних модель може розвинути упередженість до більш частих категорій. У випадку біометричної автентифікації, де точність ідентифікації критично важлива, така упередженість може знижувати загальну ефективність системи. Наприклад, якщо в датасеті переважають дані про людей певної вікової групи або етнічної приналежності, то система буде краще працювати з цими групами і гірше з іншими, що не є представлені достатньо [39].

Одним із способів є ресемплінг даних, який включає збільшення кількості зразків для менш представлених класів або зменшення кількості зразків для перепредставлених класів.

Іншим варіантом може бути застосування штучних технік збільшення даних, таких як аугментація даних або генеративно-змагальні мережі, щоб створити додаткові зразки для менш представлених категорій.

2.4.2 Процес навчання нейронних мереж

Перший крок полягає в зборі відповідного датасету, який відповідає вимогам проекту. Дані можуть включати зображення, звуки, текстові дані тощо, в залежності від завдання, для якого навчається нейронна мережа. На рисунку 2.4 наданий опис навчання нейронної мережі за допомогою шарів пулінгу.

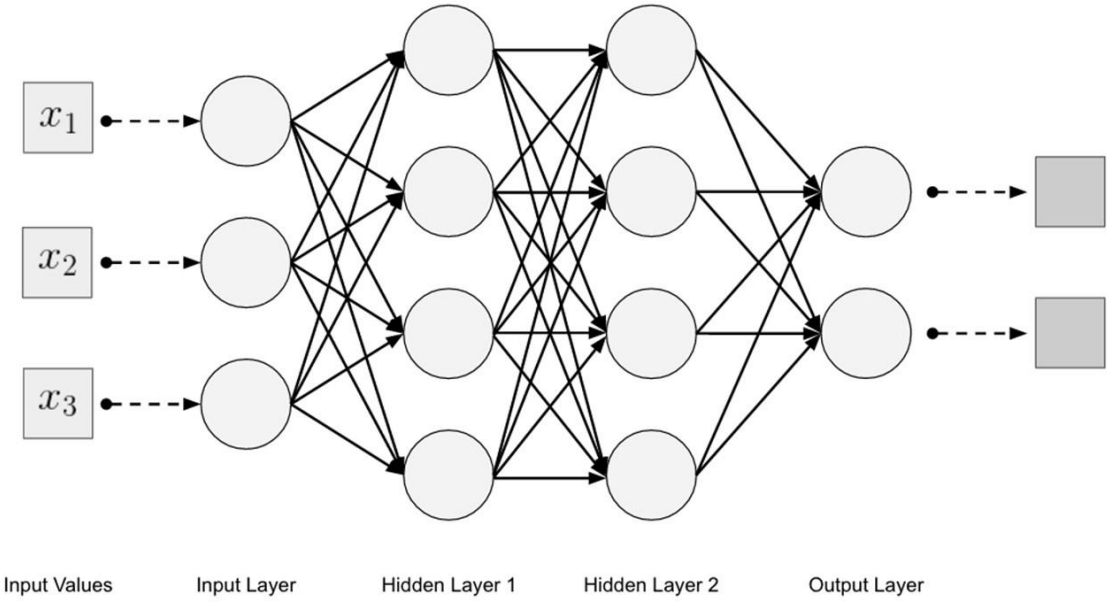


Рисунок 2.4 – Навчання нейронних мереж.

На наступному етапі дані очищаються від аномалій та виправляються помилки. Нормалізація включає масштабування числових значень до певного діапазону, зазвичай від 0 до 1, щоб забезпечити що нейронна мережа ефективно навчалася без перекосів у вагах через надмірно великі або малі вхідні значення [56].

Датасет зазвичай розділяють на три частини: тренувальний набір, валідаційний набір та тестовий набір. Тренувальний набір використовується для непосереднього навчання мережі, валідаційний – для налаштування параметрів та перевірки перенавчання, а тестовий набір служить для оцінки кінцевої ефективності моделі.

Після цього проводяться кроки проти перенавчання та для обробки нечислових даних. Аугментація може включати зміни в зображеннях (поворот, масштабування, зміна освітленості) або в текстових даних (синонімізація, зміна порядку слів). Для обробки нечислових даних, таких як текст або зображення, використовують векторизацію, перетворюючи вхідні дані в числові вектори, які можна подавати на вхід нейронної мережі [59].

2.4.3 Техніки оптимізації

Зворотне поширення помилки є фундаментальним методом у машинному навчанні, який використовується для ефективного навчання глибоких нейронних мереж. Метод полягає у вирахуванні градієнта функції втрат по відношенню до кожного ваги в мережі, починаючи з виходу та рухаючись назад до входів, що дозволяє точно визначити, як зміна кожного ваги впливає на кінцеву помилку, і відповідно оновити ваги для зменшення цієї помилки.

2.4.4 Зворотне поширення помилки

Техніки оптимізації використовують розраховані градієнти для оновлення вагів мережі з метою зменшення функції втрат. Найпоширеніші методи оптимізації включають стохастичний градієнтний спуск (SGD), Adam (Adaptive Moment Estimation), і RMSprop, кожен з яких має свої переваги та налаштування для різних типів задач та архітектур мережі.

Adam часто вважається ефективним варіантом, оскільки він адаптується до масштабу параметрів і використовує оцінки моментів першого (середнє) та другого (невідцентрована дисперсія) порядків, що дозволяє досягати більш стабільної та швидкої збіжності [52].

2.4.5 Оцінка точності моделей та методи уникнення перенавчання

Оцінка точності допомагає зрозуміти, наскільки ефективно модель виконує задане завдання, в той час як методи уникнення перенавчання гарантують, що модель буде ефективно працювати не тільки на тренувальних даних, але й у реальних умовах.

Оцінка точності зазвичай включає використання метрик, таких як точність, відгук, F1-скор та область під кривою (AUC) ROC. Ці метрики дозволяють оцінити, як часто модель правильно ідентифікує позитивні та негативні класи, а також баланс між помилково позитивними та помилково негативними класифікаціями.

Використання перехресної перевірки, зокрема k-кратної перехресної перевірки, також є ефективним способом оцінити здатність моделі узагальнювати результати на нових даних, не бачених під час тренування [55].

2.4.6 Методи уникнення перенавчання

Методи уникнення перенавчання включають техніки як рання зупинка, регуляризація (наприклад, L1 та L2 регуляризація), та Dropout. Рання зупинка використовується для припинення тренування моделі, коли помічається, що помилка на валідаційному наборі даних починає зростати, що є ознакою перенавчання.

Регуляризація додає штраф до функції втрат, що допомагає уникнути надмірної ваги вагів моделі, тим самим зменшуючи ризик перенавчання.

Dropout — це техніка, що випадково "вимикає" частину нейронів під час тренування, що сприяє створенню більш стійкої мережі, яка не надто залежить від будь-якого одного нейрону [54-59].

2.5 Розпізнавання обличчя за допомогою CNN

Однією з найпопулярніших моделей для розпізнавання облич є DeepFace, розроблена Facebook, яка використовує багатошарову архітектуру CNN для створення майже людської точності у визначенні осіб. Інша значуща модель, FaceNet від Google, використовує тривимірні вектори для представлення обличчя в просторі особливостей, де відстані між векторами прямо відображають ступінь схожості між різними обличчями.

Повнозв'язні шари в кінці архітектури використовують особливості для класифікації та ідентифікації конкретного обличчя.

Однак, розпізнавання обличчя за допомогою CNN може зіштовхнутися з кількома проблемами, зокрема, зі змінами у висвітленні, позиції обличчя та наявністю певних аксесуарів (окуляри, шапки).

Для покращення точності та адаптивності моделей використовуються техніки посилення даних та введення додаткових регуляризаційних шарів для управління перенавчанням.

Підходи на основі глибокого навчання також включають використання нейронних мереж з глибоким навчанням для вирішення задач, пов'язаних з розпізнаванням обличчя в реальному часі, що відкриває можливості для розробки більш складних та ефективних систем безпеки.

Такі системи можуть ідентифікувати особи, навіть коли вони частково закриті або коли зображення має низьку якість, чого не можуть досягти більшість класичних методів.

2.5.1 Класичні методи

Включають в себе використання таких алгоритмів як Haar Cascade, який застосовується для виявлення обличчя у зображеннях за допомогою прикмет, визначених вручну. Інший приклад - метод власних обличчя (eigenfaces), що використовує аналіз головних компонент для зведення обличчя до набору ключових візуальних ознак. Хоча ці методи можуть бути ефективними у контрольованих умовах, вони часто страждають від обмеженої гнучкості та масштабованості.

2.5.2 Глибоке навчання

На противагу класичним методам, підходи на основі глибокого навчання використовують нейронні мережі для аналізу обличчя. Зокрема, згорткові нейронні мережі (CNN) забезпечують вирішення завдань розпізнавання обличчя на значно вищому рівні точності. Вони автоматично виявляють необхідні риси без потреби в їх явному визначенні, що дозволяє ефективно адаптуватися до різних умов освітленості, позицій обличчя та інших варіативних факторів [38].

2.5.3 Основні міри оцінки

В таблиці 2.2 наданий опис точності, повноти, оцінки F1 та ROC кривої.

Таблиця 2.2 – Точність, повнота, F1-Score, ROC-крива

Назва параметра	Опис
Accuracy	Вимірює частку правильно ідентифікованих обличчя у відношенні до всіх спроб ідентифікації. Точність важлива для оцінки загальної ефективності системи.
Precision	Precision вказує на частку правильно позитивних результатів з усіх позитивних результатів, які надала система.
Recall	Recall відображає частку правильно позитивних результатів з усіх реальних позитивних випадків у датасеті
F1 Score	Гармонійне середнє точності та повноти, F1-бал є корисним показником для випадків, де потрібно знайти баланс між точністю та повнотою, важливо в системах, де обидві помилки — як помилково позитивні, так і помилково негативні — мають велике значення.
ROC	ROC-крива демонструє відношення між часткою правильно позитивних результатів та часткою помилково позитивних результатів на різних порогах дискримінації.
AUC	AUC дозволяє оцінити, наскільки добре модель може відрізняти між класами. Вищий AUC вказує на кращу здатність моделі до класифікації.

Застосування таких технологій вимагає ретельного підходу до вибору датасетів для навчання, де важливо забезпечити репрезентативність всіх біометричних варіацій.

Крім того, забезпечення приватності та безпеки оброблюваних біометричних даних є ключовим елементом, що впливає на довіру до таких систем і їх широке впровадження [38].

Методи аналізу зазвичай включають використання CNN, які ефективно обробляють візуальні дані, такі як зображення обличчя та інших видимих частин тіла. Ці мережі здатні виявляти тонкі нюанси в структурі шкіри та колірні особливості, що робить їх ідеальними для розпізнавання особи за біометричними ознаками.

На рисунку 2.5 надана оцінка продуктивності системи розпізнавання обличчя з використанням різних класифікаторів відстані.

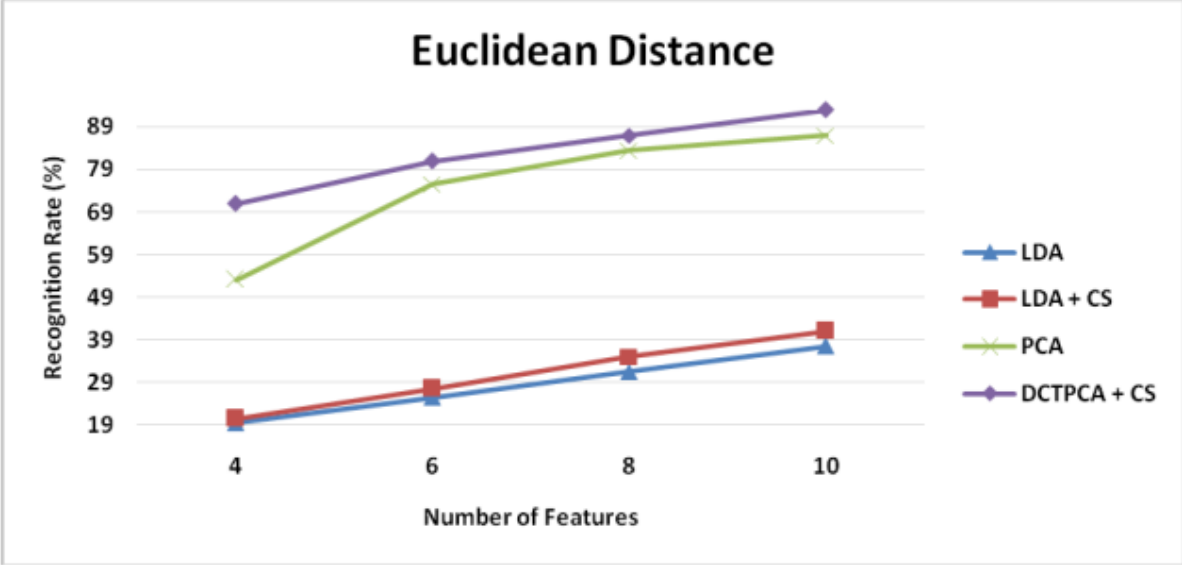


Рисунок 2.5 – Оцінка продуктивності системи розпізнавання обличчя з використанням різних класифікаторів відстані.

Класифікація даних відбувається за допомогою алгоритмів глибокого навчання, які забезпечують високу точність відділення одних біометричних характеристик від інших. Це дозволяє системі точно класифікувати, наприклад, колір очей серед широкого спектру можливих варіацій та особливостей, що представлені в різних етнічних групах [55].

2.5.4 Створення комплексних біометричних профілів

CNN забезпечують потужний інструмент для аналізу та класифікації біометричних даних. Вони особливо ефективні у створенні комплексних біометричних профілів, які інтегрують різноманітні дані, такі як обличчя, відбитки пальців, іриси ока та інші фізичні характеристики.

Використовуючи глибокі шари навчання, CNN здатні вилучати корисні особливості з великих обсягів даних, що дозволяє створювати детальні та точні профілі для ідентифікації та автентифікації особи.

Приклад коду у Додатку А.

У цьому прикладі CNN має три згорткові шари, які слугують для ефективного видобування характеристик з зображення обличчя. Після згорткових шарів використовуються шари пулінгу для зменшення розмірності вхідних даних та підвищення стійкості мережі до невеликих змін у вхідних образах. Використання повнозв’язного шару (Dense) та шару Dropout допомагає зменшити перенавчання та покращує загальну здатність мережі до класифікації.

2.6 Висновки до розділу

У другому розділі дипломної роботи детально розглянуто використання CNN для біометричної автентифікації, зокрема для розпізнавання обличчя та інших біометричних параметрів.

Розділ включає глибокий аналіз архітектури та ключових елементів CNN, а також описує основні принципи їхньої роботи, що дозволяє глибше зрозуміти механізми та переваги застосування цих технологій у сучасних системах безпеки.

Детально розглянуто датасети, що використовуються для тренування нейронних мереж, визначено критерії вибору даних і описано проблеми, пов'язані із збалансованістю даних. Це дає можливість зрозуміти, як неправильний вибір тренувального набору може вплинути на точність та ефективність моделей.

Процес навчання нейронних мереж включає підготовку даних, застосування методів оптимізації та оцінку точності моделей, включаючи стратегії уникнення перенавчання. Розглянуто різні підходи до тренування, включаючи зворотне поширення помилки та методи оцінки ефективності, що є критично важливими для розробки надійних систем.

3 РОЗРОБКА ТА ІМПЛЕМЕНТАЦІЯ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ КЛАСИФІКАЦІЇ КОЛЬОРУ ОЧЕЙ НА ОСНОВІ ФОТОГРАФІЙ

В цьому розділі буде розроблена архітектура мережі, вибір необхідних шарів та методи регуляризації, дані будуть розроблені та зібрані для збору датасету, після чого для обробки та аугментації даних, почнемо процес навчання, налаштуємо параметри навчання та відвалідуємо модель.

3.1 Архітектура мережі

Основу архітектури мережі складають згорткові шари. У моделі з додатку А використовуються два згорткові шари, де перший шар має 32 фільтри з ядром розміром 3x3, а другий — 64 фільтри того ж розміру. Згорткові шари виконують основну роль у виявленні візуальних особливостей з вхідних фотографій, таких як форма і колір очей [54].

Функція активації 'ReLU' використовується для введення нелінійності в обробку даних, що дозволяє мережі ефективно навчатися на складних даних із зображень. На рисунку 3.1 надана методологія розробленої мережі.

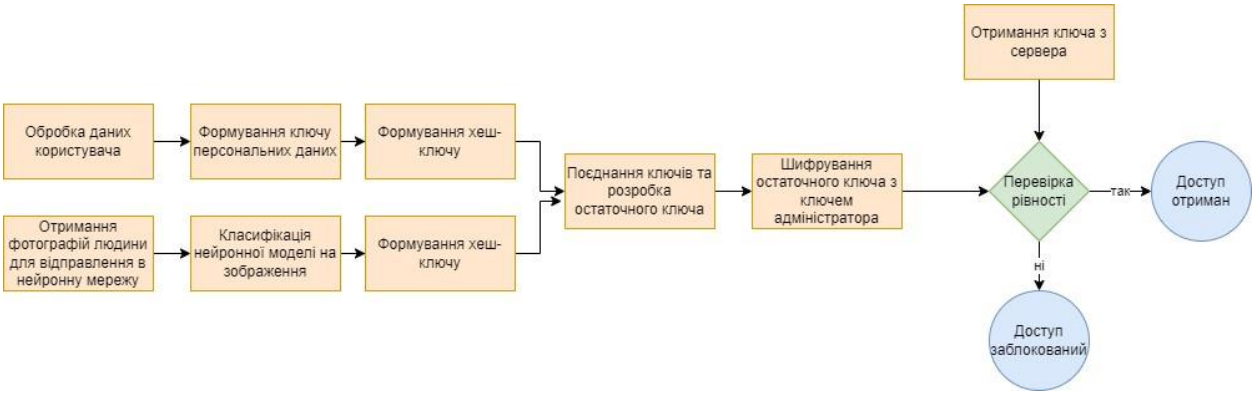


Рисунок 3.1 – Методологія нейронної мережі.

Після кожного згорткового шару слідує шар пулінгу, який виконує операцію максимального пулінгу з розміром 2x2.

Шари знижують просторові розміри обробленого зображення, зменшуючи кількість параметрів і обчислень у мережі, що допомагає уникнути перенавчання та зберігати тільки найважливіші особливості [40].

На завершальному етапі архітектури розташовані повнозв'язні шари, які використовують лінійну комбінацію особливостей, виявлених згортковими та пулінг шарами, для класифікації кольору очей.

Модель має один повнозв'язний шар з 64 нейронами та вихідний шар з п'ятьма нейронами, що відповідають п'яти класам кольору очей. На рисунку 3.2 надана архітектура моделі з 64 нейронами а вихідним шаром.

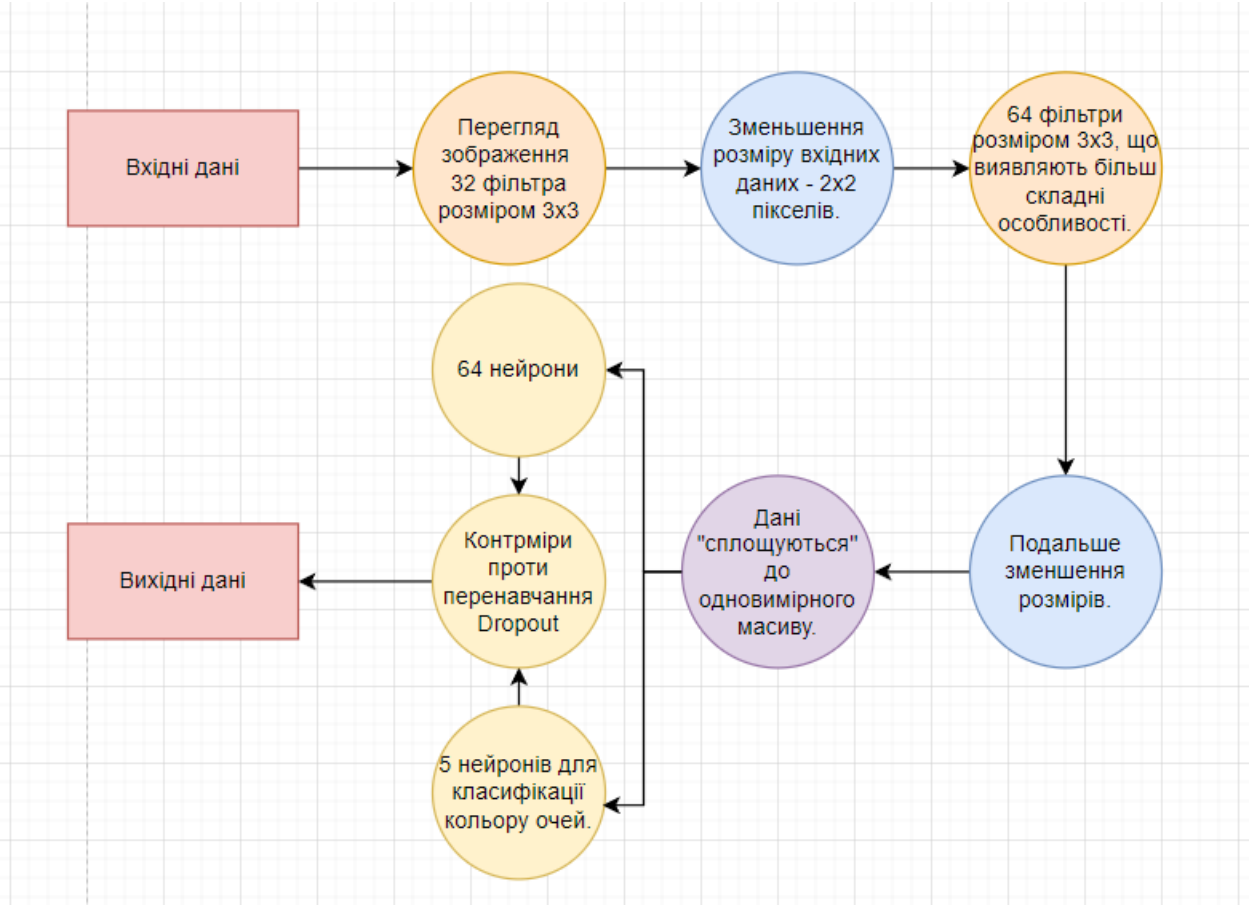


Рисунок 3.2 – Архітектура мережі з додатку А.

В цій нейронній мережі використані методи регуляризації за допомогою Dropout, регуляризація є ключовим елементом у проектуванні нейронних мереж для запобігання явищу перенавчання, коли модель надмірно

адаптується до тренувальних даних і втрачає здатність узагальнювати на нових даних.

Шар Dropout включений перед вихідним шаром для зниження перенавчання, випадково відключаючи деякі нейрони під час тренування, що допомагає моделі краще узагальнювати на невидимих даних [45].

Dropout діє шляхом випадкового відключення (ігнорування) деяких вузлів (нейронів) мережі під час тренування. Це означає, що кожен нейрон має певну ймовірність бути тимчасово виключеним з мережі на кожному кроці тренування, що змушує модель не покладатися надмірно на будь-який один нейрон та розвивати більш робастні ознаки, які можуть бути корисними взагалі, а не лише для конкретного прикладу або партії даних [39].

В нейронній мережі застосування Dropout допомагає знизити перенавчання, оскільки модель стає менш чутливою до маленьких флуктуацій у вхідних даних, а також сприяє тому, що різні нейрони навчаються виявляти різні аспекти та особливості входів.

В результаті, коли Dropout включений під час тренування, отримана модель може краще узагальнювати на невидимих даних, що є важливим для практичного застосування моделі у реальних задачах.

Dropout застосовується до повнозв'язних шарів, що є більш схильними до перенавчання через велику кількість параметрів і зв'язків. В мережі, розробленій для класифікації кольору очей, Dropout застосовується перед вихідним шаром, що сприяє розрідженості активацій та змушує модель розробляти більш розподілене представлення характеристик.

3.2 Підготовка даних

Підготовка даних для тренування нейронної мережі вимагає ретельного підходу до збору і валідації даних, що включає вибір, анотацію і перевірку якості фотографій. Важливим аспектом є забезпечення репрезентативності, балансу та різноманітності даних у тренувальному і тестувальному датасетах. [45].

Збір може відбувався за допомогою різних канали, включаючи публічно доступні бази даних та використанням власних зборів інформації. Для розробки системи розпізнавання кольору очей використали спеціалізовані біометричні датасети, які допомогли значно підвищити точність і надійність результатів. В таблиці 3.1 наданий кількість переліку зображень використаних з різних ресурсів.

Таблиця 3.1 – Підготовка даних для нейронної мережі.

Колір очей	Кількість з інтернету	Кількість з датасету LFW	Кількість з датасету CASIA
Сині	443	2000	500
Карі	1207	20000	4000
Зелені	597	1500	300
Сірі	517	1000	200
Лазурні	146	500	100
Гетерохромія	42	200	50
Буршитнові	351	1600	400
Чорні	492	3000	600
Світло-карі	139	2200	450
Темно-карі	293	2800	550

Перший крок в обробці даних — це ресайзинг зображень, що дозволяє привести всі вхідні дані до єдиного розміру, забезпечуючи консистентність та зниження обчислювальної складності. Нормалізація даних, зазвичай виконувана шляхом масштабування інтенсивностей пікселів до діапазону [0,1] або за допомогою стандартної нормалізації, забезпечує більш стабільне та швидке навчання, оскільки такі дані сприятливіше впливають на процес оптимізації.

Аугментація даних виконується з метою збільшення розмаїття тренувального датасету, що сприяє підвищенню узагальнюючих здібностей

моделі та її стійкості до перенавчання. Різноманітні методи аугментації включають повороти, масштабування, зміщення, зернистість, зміну яскравості та контрасту зображень. Ці техніки дозволяють моделі вчитися на більш широкому спектрі можливих вхідних умов, імітуючи потенційні зміни у вхідних даних, з якими модель може стикатися під час реального використання.

Процес навчання нейронної мережі починається з підходу до налаштування параметрів, щоб забезпечити оптимальну ефективність моделі.

Оптимізатор — це алгоритм, який відповідає за оновлення ваг моделі в процесі навчання з метою мінімізації функції втрат. Вибір оптимізатора може значно впливати на швидкість навчання та збіжність нейронної мережі. Наприклад, популярні оптимізатори, такі як Adam, SGD (стохастичний градієнтний спуск), та RMSprop, відрізняються механізмами коригування швидкості навчання та мають різні переваги в залежності від конкретної задачі та структури даних.

У Додатку В надані приклади зміні нейронної мережі для використання оптимізаторів.

- Adam автоматично налаштовує швидкість навчання і працює добре на широкому спектрі задач.
- SGD налаштований з моментом для покращення конвергенції.
- RMSprop автоматизує процес налаштування швидкості навчання і використовується, коли потрібно швидке збіження, що актуально в цьому прикладі, через використання рекурентної нейронної мережі.

Adam – це адаптивний метод оцінки моментів, який використовується для оновлення мережевих ваг на основі обчислення першого (середнє) і другого (невизначене) моментів градієнтів. Adam є одним із найпопулярніших оптимізаторів завдяки його високій ефективності в багатьох типах моделей.

SGD (Stochastic Gradient Descent) – це більш традиційний підхід, який оновлює параметри моделі на основі кожного навчального прикладу. SGD

часто використовується з розміром міні-пакета для зменшення варіативності оцінок градієнтів, що забезпечує стабільніше навчання, але може потребувати більше часу для конвергенції.

RMSprop – це оптимізатор, що розділяє швидкість навчання для кожного параметра за допомогою поділу градієнтів на їхні середньоквадратичні значення. RMSprop забезпечує швидке сходження у початковій фазі навчання і знижує темп навчання у міру наближення до оптимальних значень, що робить його ефективним при роботі з нестаціонарними даними та великими наборами даних.

В таблиці 3.2 наданий опис параметрів оптимізаторів використаний у нейронній мережі.

Таблиця 3.2 – Використання оптимізатора для покращення мережі.

Назва оптимізатора	Learning rate	optimizer	momentum	loss
Adam	-	adam	-	categorical_crossentropy
SGD	0.01	SGD	0.9	categorical_crossentropy
RMSprop	0.001	Rmsprop	-	categorical_crossentropy

В нейронній мережі метрики успішності використовуються для оцінки ефективності моделі, метрики визначають, наскільки точно модель виконує задану задачу. Для задач класифікації стандартні метрики включають точність (accuracy), F1-оцінку, влучність (precision) та відгук (recall).

3.3 Оцінка моделі

Після підготовки усіх даних перейшли до оцінки моделі, що є завершальним етапом у процесі розробки нейронних мереж. Після тренування та валідації модель піддалась тестуванню на окремо відокремленому тестовому наборі даних, який не використовувався під час навчання.

Спочатку зібрався тестовий датасет, який містить репрезентативні зразки, на яких модель не навчалася. Ці дані повинні точно відображати умови,

в яких буде використовуватися модель. В таблиці 3.3 описана кількість зображень відносно до кольору очей для тестування мережі.

Таблиця 3.3 – Кількість обраних зображень для тестування мережі

Колір очей	Кількість зображень
Сині	241
Карі	873
Зелені	499
Сірі	318
Лазурні	142
Гетерохромія	35
Буршитнові	72
Чорні	257
Світло-карі	100
Темно-карі	199

.

Під час тестування модель обробила тестовий датасет, і був створений файл для результатів, порівнюються з фактичними мітками даних, що допомогло виміряти, наскільки точно модель може ідентифікувати або класифікувати нові дані.

Для оцінки ефективності моделі були використани кількісні показники: точність, повнота, F1-міра.

Щоб уникнути перенавчання були використані додаткові техніки, оскільки перенавчання може призвести до створення моделі, яка ідеально працює на тренувальних даних, але погано справляється з новими, раніше невідомими даними.

В цій роботі була використана техніка ранньої зупинки, її інструкцій надана в таблиці 3.4, оскільки вона не тільки допомагає уникнути перенавчання, але й оптимізує витрати часу та обчислювальних ресурсів, не

витрачаючи їх на тренування, яке не призводить до покращення узагальнюючої здатності моделі.

Таблиця 3.4 – Техніка ранньої зупинки

Принцип дії	Рання зупинка полягає у припиненні тренування моделі, коли показники на валідаційному наборі даних починають погіршуватися, незважаючи на подальше покращення показників на тренувальному наборі, свідчить про те, що модель починає "запам'ятовувати" тренувальні дані, а не узагальнювати з них.
Реалізація	У практичній реалізації ранньої зупинки тренування моделі відстежуються показники, такі як втрати (loss) чи точність (accuracy), на валідаційному наборі даних під час кожної епохи. Якщо ці показники перестають покращуватися або навіть погіршуються протягом певної кількості епох, тренування припиняється.
Налаштування параметрів	Для ефективної реалізації ранньої зупинки були встановлені кількість епох, протягом яких відсутнє покращення та мінімальна зміна показників, яка вважається значущою.

3.4 Тестування моделі

Тестування проводиться за допомогою тестового датасету, що складається з набору зображень, які модель раніше не "бачила".

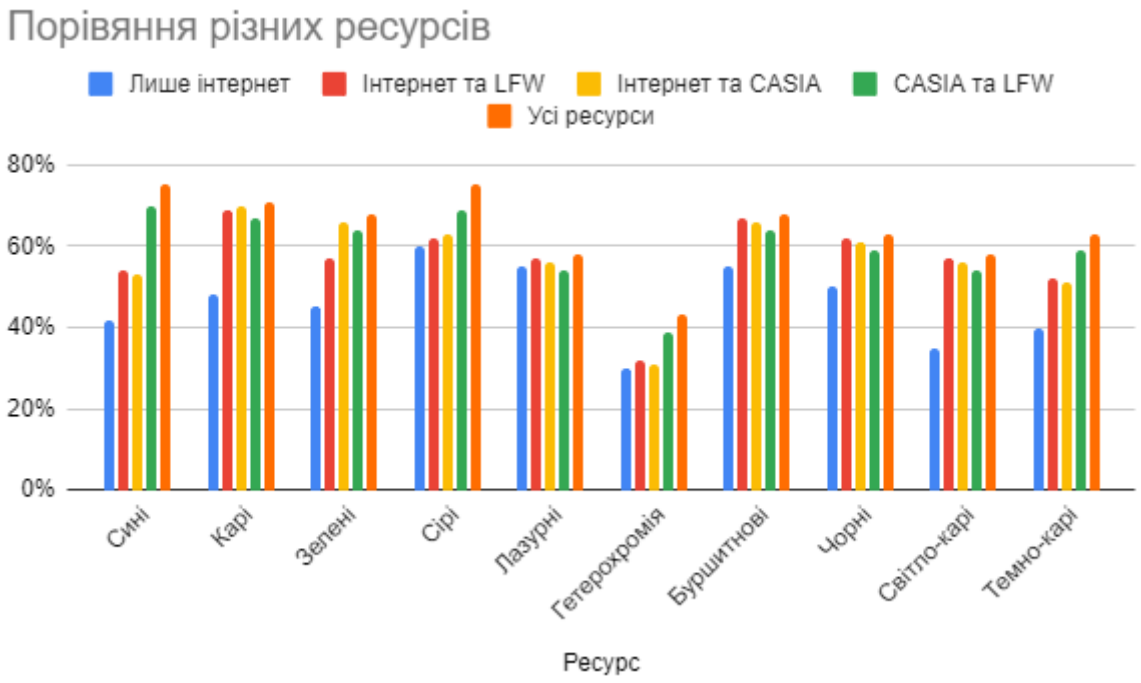
Перед початком тестування моделі необхідно було забезпечено, що всі дані були належним чином оброблені та нормалізовані для забезпечення консистентності вводу [48].

Тестування здійснюється шляхом подачі зображень з тестового набору до моделі, після чого результати порівнюються з відомими мітками класів. Аналіз помилок, які здійснила модель, може виявити потенційні проблеми з

загальністю моделі або вказати на необхідність подальшого регулювання параметрів тренування.

Окрім кількісних метрик, також було проведено тестування візуальної перевірки результатів класифікації, що дозволила визначити, як модель впоралася з візуально різними випадками.

Були проведення тестування на порівняння результатів роботи нейронної мережі після навчання на різних ресурсах на графіку 3.1.



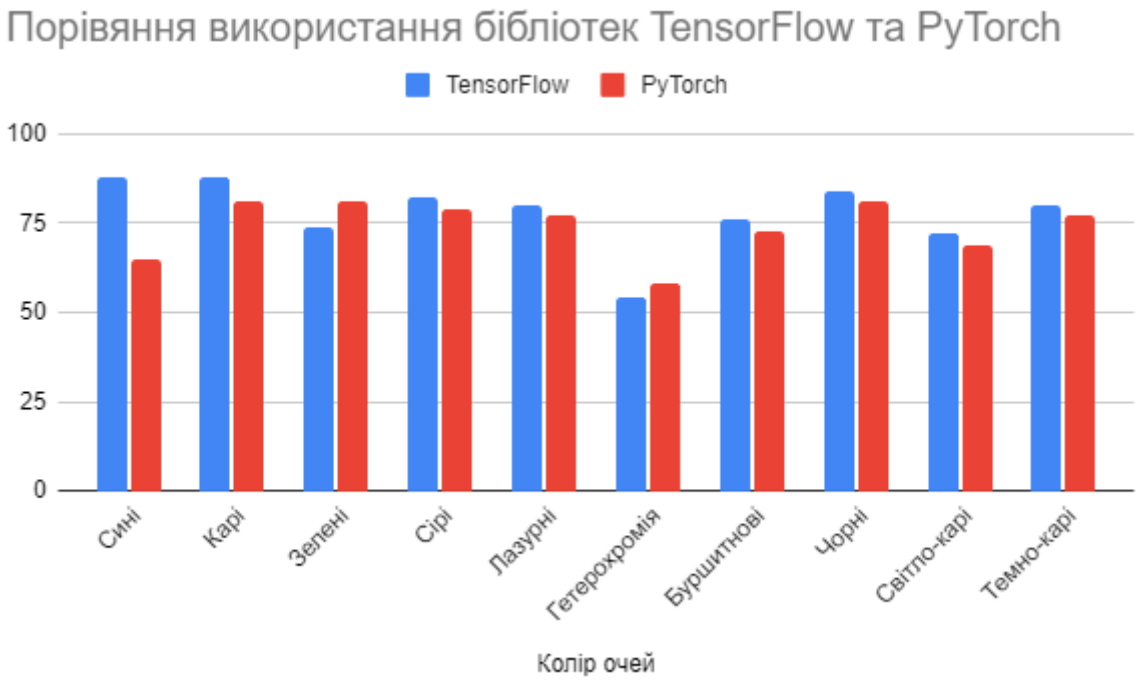
Графік 3.1 – Результати навчання нейронної мережі на різних ресурсах

З таблиці видно, що поєднання даних з різних джерел (інтернет, LFW, CASIA) зазвичай приводить до покращення точності моделей. Наприклад, коли використовуються всі ресурси, точність для всіх категорій кольору очей є найвищою. Також можна відмітити, що деякі датасети можуть бути більш ефективними для певних кольорів очей. Наприклад, поєднання CASIA та LFW демонструє високу точність для сірих та темно-карих очей, свідчить про те, що в цих датасетах є хороше представлення цих кольорів, або вони містять зразки, які допомагають мережі краще вчитися.

Однак, точність розпізнавання різних кольорів очей може істотно варіюватися, що вказує на потенційні складнощі в класифікації менш поширених або більш складних кольорів, як наприклад гетерохромія. Низькі показники точності для рідкісних кольорів очей, таких як гетерохромія, можуть вказувати на недостатнє представлення цих класів в датасетах або на вищу складність в класифікації цих унікальних варіацій [49].

Загалом, комбінація декількох джерел даних забезпечує кращу узагальнюючу здатність моделей, що допомагає у підвищенні точності розпізнавання через більшу варіативність і кількість навчальних зразків.

Після вдалого тестування нейронної мережі на основі TensorFlow, перешли до тестування використання двох різних фреймворків, результати тестування представлені на графіку 3.2.



Графік 3.2 – Результати навчання різних нейронних мереж на основі використання різних бібліотек.

Обидві мережі показують високу точність для синіх та карих очей, де TensorFlow має однаковий показник для обох кольорів – 88%, в той час як

PyTorch показує трохи меншу точність для синіх (65%) та високу для карих (81%).

Для зелених очей PyTorch демонструє вищу точність (81%) порівняно з TensorFlow (74%), що може свідчити про кращу здатність PyTorch вловлювати особливості цього кольору очей. Сірі очі мають порівняно високу точність в обох системах.

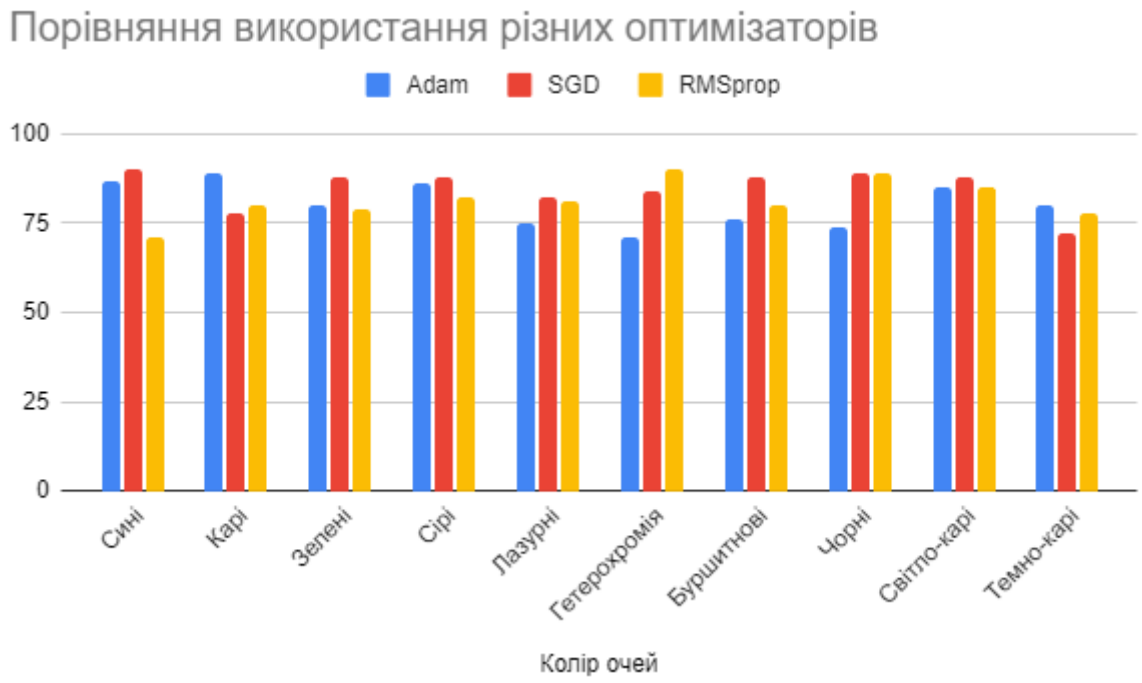
Для менш поширених кольорів, таких як гетерохромія, обидві мережі показують нижчу точність, з легкою перевагою у PyTorch (58% проти 54% у TensorFlow).

Чорні, світло-карі та темно-карі очі мають порівняно високі показники в обох мережах, з невеликим варіюванням між двома платформами.

Узагальнюючи, TensorFlow має трохи вищі піки точності для деяких кольорів, але PyTorch виявляється більш стабільним у широкому спектрі кольорів очей, що може бути перевагою в застосуваннях, де потрібна рівномірна ефективність для різноманітних біометричних характеристик.

Існуючі результати точності різних моделей показали, що хоча вони забезпечують задовільну точність, може існувати потенціал для їх покращення, особливо в контексті різноманітності кольорів очей та біометричних параметрів [43].

Щоб подолати ці обмеження та підвищити загальну точність і адаптивність нейронних мереж, вирішили використовувати різні оптимізатори, такі як Adam, SGD та RMSprop, результати тестування представлені на графіку 3.3 та їх код у додатку В.



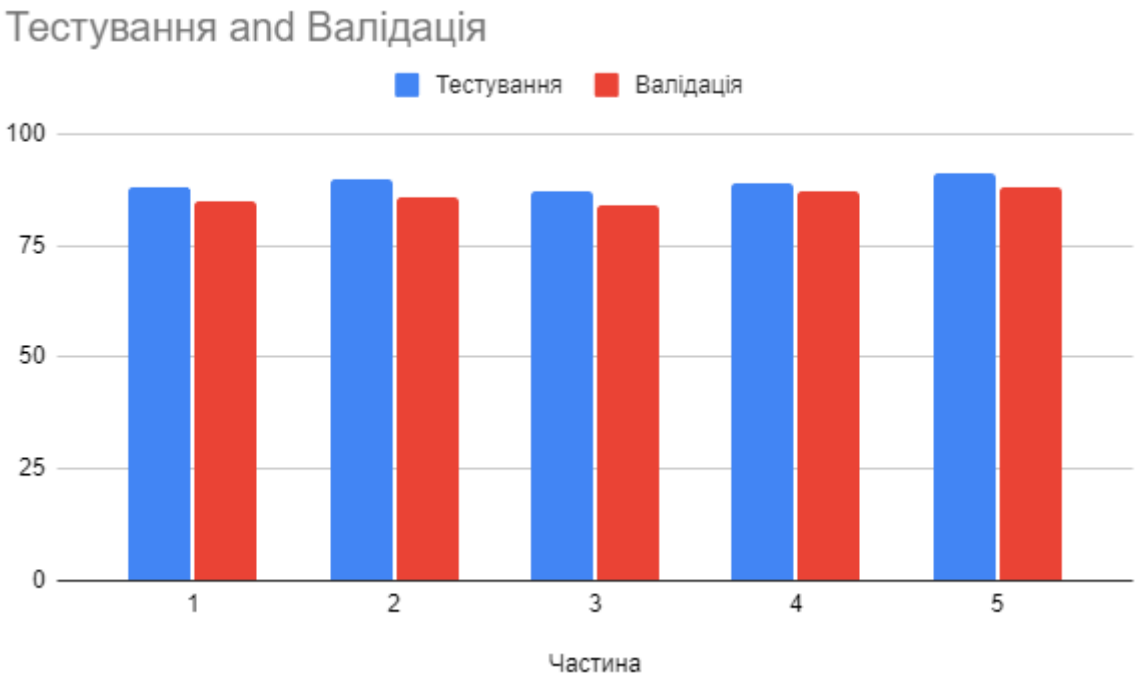
Графік 3.3 – Порівняльний аналіз використання різних параметрів оптимізацій.

Adam показав стабільно високі результати для карих та зелених очей, але був менш ефективний для незвичайних кольорів, таких як гетерохромія.

SGD виявився найкращим у більшості випадків, особливо для рідкісних кольорів очей, показуючи високу точність для гетерохромії.

RMSprop виявився досить балансованим, забезпечуючи хороші результати у широкому спектрі кольорів, але мав випадки, де його ефективність була нижчою порівняно з іншими оптимізаторами.

Для забезпечення більш об'єктивної оцінки узагальнюючої здатності розробленої моделі з розпізнавання кольору очей, було вирішено застосувати метод перехресної валідації, зокрема п'ятикратну перехресну валідацію [50]. Результати тестування представлені на графіку 3.4.



Графік 3.4 – Тестування при розділі на декілька частин.

Кожна з частин використовує різні частини датасету як тренувальну та валідаційну вибірки. Показники точності на тренувальних даних та на валідаційних даних вказують на високу узагальнюючу здатність моделі.

Точність на тренувальній вибірці знаходиться у діапазоні від 87% до 91%, в той час як точність на валідаційній вибірці знаходиться між 84% і 88%, що свідчить про добре збалансованість моделі і мінімізацію явища перенавчання.

Результати підтверджують, що модель достатньо стабільна та може ефективно класифікувати біометричні дані у різних умовах, що робить її придатною для використання в реальних сценаріях.

Для оцінки впливу різних параметрів вхідних даних на результати класифікації, було проведено аналіз чутливості. Аналіз дозволив ідентифікувати ключові характеристики, які найбільше впливають на точність моделі.

В таблиці 3.5 представлені результати тестування нейронної мережі.

Таблиця 3.5 – Аналіз чутливості нейронної мережі

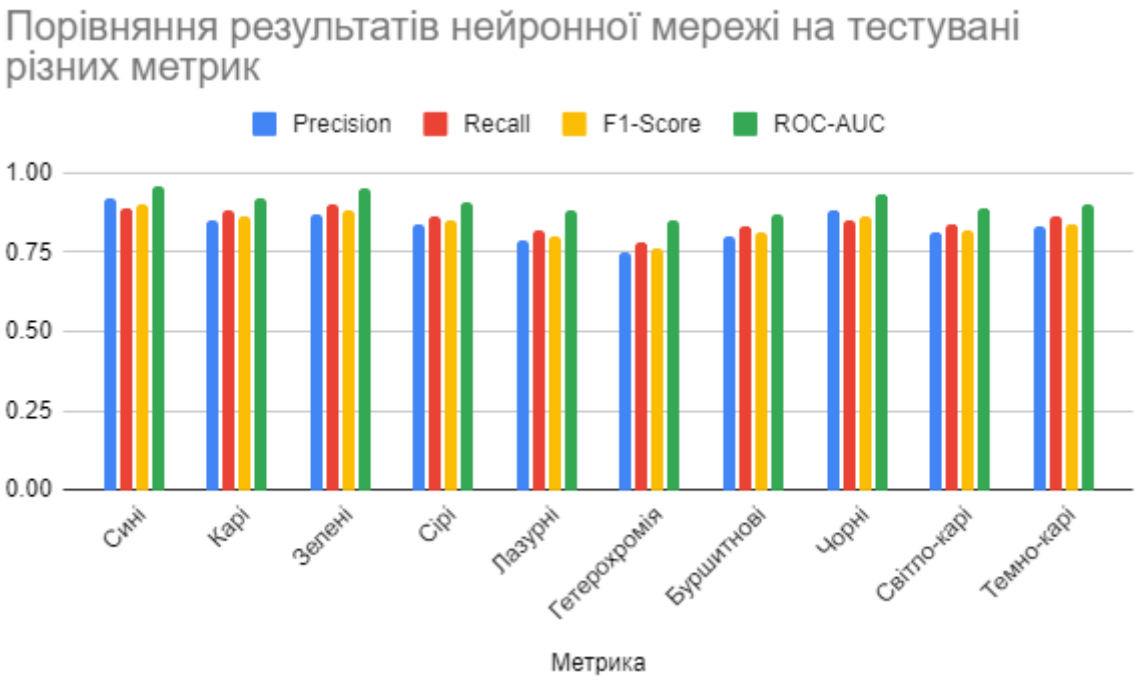
Параметр	Опис змін
Яскравість	Зміна яскравості зображень мала відносно малий вплив на точність, змінивши її лише на 2%.
Контраст	Зміна контрасту мала більший вплив, змінивши точність на 5%.
Шум	Додавання шуму до зображень вплинуло на точність, знизивши її на 3%.
Кількість тренувальних даних	Збільшення кількості даних для тренування показало найбільший вплив, підвищивши точність на 10%.

Аналіз помилок у класифікації допоміг глибше зрозуміти умови, за яких модель видає невірні прогнози. Дослідження зосередилось на виявленні кольорів очей, для яких точність класифікації є низькою.

Наприклад, модель помилялася при ідентифікації рідкісних кольорів очей, таких як лазурні або гетерохромія, через недостатність представлення цих категорій у навчальному датасеті.

Також помилки можуть з'явилися внаслідок недоліків у процесі попередньої обробки зображень, як-от неправильне вирівнювання або недостатня якість зображень, що ускладнює розпізнавання патернів.

Проведене тестування за допомогою різних метрик, таких як Precision, Recall, F1-Score, та ROC-AUC, дозволяє глибше оцінити якість роботи нейронної мережі, особливо у випадках, коли точність класифікації критично важлива, результати тестування на графіку 3.5. Основна мета цих тестів полягає в тому, щоб визначити, наскільки добре модель впорається з розпізнаванням різних кольорів очей у різноманітних умовах [55].



Графік 3.5 – Оцінка роботи моделі за допомогою різних метрик

Значення Precision відображає, яка частка ідентифікованих як певний колір очей випадків дійсно відповідає цьому кольору, що вказує на здатність моделі мінімізувати помилково позитивні результати. Recall показує, яку частину випадків фактично цього кольору модель успішно ідентифікує, що є критично важливим для ситуацій, де пропуск класифікації може призвести до небажаних наслідків.

F1-Score є гармонійним середнім між Precision та Recall, забезпечуючи баланс між чутливістю та специфічністю, що є важливим для оцінювання загальної точності моделі в умовах, де обидві характеристики мають значення [45].

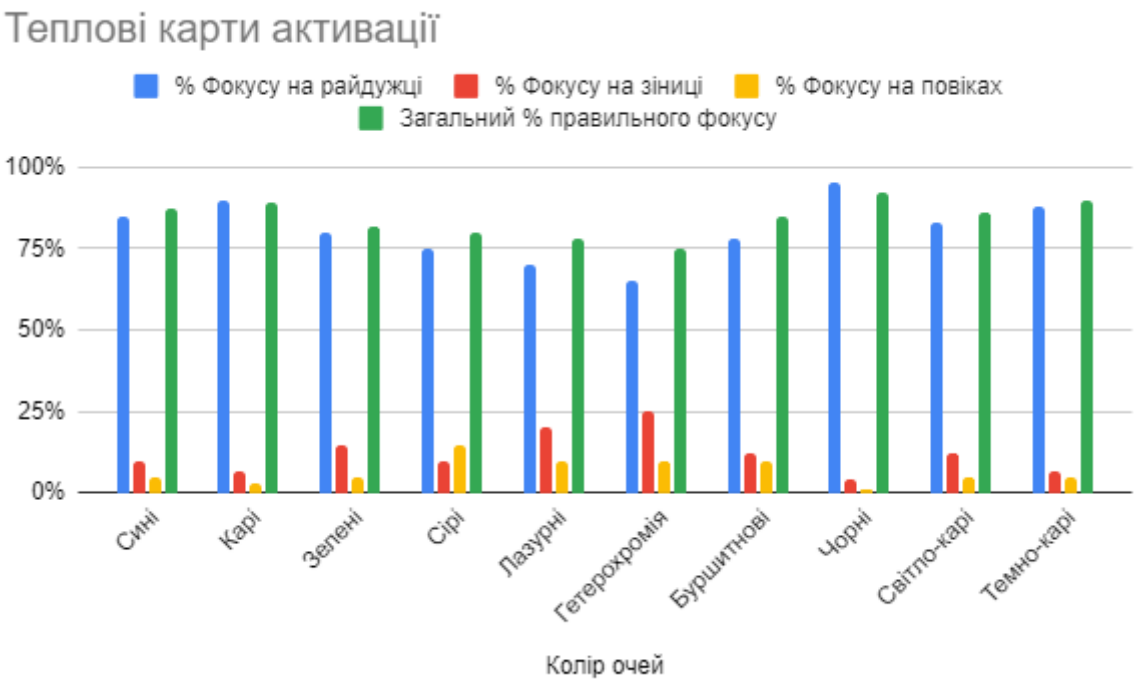
Значення ROC-AUC визначає здатність моделі диференціювати між класами незалежно від порогу класифікації, що демонструє її загальну надійність і ефективність у різних умовах роботи.

З отриманих результатів можна зробити висновок, що модель демонструє високу ефективність у розпізнаванні деяких кольорів очей, особливо синіх та карих, з високими значеннями у всіх метриках. Водночас,

для кольорів як гетерохромія модель показує нижчі результати, що може свідчити про потребу додаткової оптимізації або перегляду підходів до тренування.

Також було проведено тестування за допомогою теплових карт активації для аналізу того, які області зображень очей найбільше активізуються при роботі нейронної мережі.

Дослідження дозволило визначити, чи модель належним чином фокусується на ключових анатомічних рисах, таких як райдужка, зіниця і повіки, які є важливими для визначення кольору очей, теплові карти активації та їх проценти представлені на графіку 3.6.



Графік 3.6 – Теплові карти активації.

Модель найкраще фокусується на райдужці для всіх кольорів очей, особливо високий відсоток фокусу на райдужці у чорних очей (95%), що свідчить про високу здатність моделі диференціювати більш темні кольори.

Найменший загальний відсоток правильного фокусу спостерігається у випадках гетерохромії (75%), що може бути пов'язано з складністю класифікації цього унікального і рідкісного варіанта кольору очей.

У деяких випадках (наприклад, сірі очі) модель має тенденцію до більшого фокусу на повіках, що може вказувати на потребу подальшого налаштування моделі для зменшення уваги до менш релевантних областей.

3.5 Додаткові функції. Розширення моделі.

Також були розглянуті додаткові функції та стратегії для розширення можливостей існуючої моделі згорткової нейронної мережі, аби вона могла класифікувати не тільки колір очей, а й інші біометричні характеристики.

За допомогою схожих архітектур можна додати класифікацію таких характеристик, як форма обличчя, розмір і форма носа, відстань між очима, а також текстура і колір шкіри.

Використання однієї моделі для одночасної класифікації декількох біометричних характеристик може покращити загальну точність завдяки спільному використанню внутрішніх ознак між різними задачами. Це допоможе моделі краще загальнізувати та ідентифікувати складніші зв'язки в біометричних даних [43].

Для швидшого розвитку і покращення результатів можна використати вже навчені моделі, які були обучені на великих і різноманітних датасетах. Трансферне навчання дозволить перенести знання з інших сфер, наприклад, від загального розпізнавання облич до більш специфічних задач, таких як ідентифікація біометричних особливостей.

Включення більшої кількості і більш різноманітних датасетів може допомогти моделі краще адаптуватися до різноманітності людських біометричних особливостей.

Комбінація кількох моделей може підвищити загальну надійність та точність системи. Енсамблі, які використовують різні підходи до аналізу та класифікації біометричних даних, забезпечують більш стабільні та надійні результати.

3.6 Висновки до розділу

У третьому розділі дипломної роботи було здійснено розробку та детальний аналіз нейронної мережі для класифікації кольору очей. Була створена та налаштована модель глибокого навчання, що включає згорткові, пулінгові та повнозв'язні шари, з використанням методів регуляризації для підвищення загальної ефективності та уникнення перенавчання.

Значна увага була приділена підготовці даних, що включала збір датасету, його аугментацію та попередню обробку для забезпечення репрезентативності та різноманітності. Особлива увага була зосереджена на точності та надійності тренувальних і валідаційних процедур, що дозволило виявити оптимальні параметри для тренування моделі.

Оцінка моделі проводилася за допомогою декількох метрик, таких як точність, F1-score, Precision, Recall та ROC-AUC, які допомогли оцінити її ефективність у реальних умовах. Було проведено також перехресну валідацію для забезпечення надійності результатів, а аналіз чутливості та помилок забезпечив глибше розуміння обмежень та можливостей моделі.

Тестування моделі в реальних умовах показало, що мережа здатна ефективно класифікувати колір очей, виявляючи високу точність у більшості випадків, з особливою увагою до критичних аспектів, таких як захист персональних даних та відповідність регулятивним вимогам. Результати підтверджують потенціал подальшого вдосконалення та розширення моделі для класифікації інших біометричних характеристик.

Третій розділ підсумовує значний обсяг досліджень та розробок у сфері використання нейронних мереж для біометричної автентифікації та підкреслює важливість продовження досліджень у цій області, зокрема в контексті забезпечення приватності та безпеки.

4 РОЗРОБКА СИСТЕМИ АВТЕНТИФІКАЦІЇ НА ОСНОВІ БІОМЕТРИЧНИХ ДАНИХ ТА ПЕРСОНАЛЬНОЇ ІНФОРМАЦІЇ

Четвертий розділ присвячений розробці та реалізації системи автентифікації, яка використовує як основу біометричні дані та персональну інформацію.

Мета полягає у створенні надійної системи, здатної верифікувати ідентичність особи на основі її біологічних та особистісних характеристик, що дозволяє забезпечити високий рівень безпеки доступу до різноманітних ресурсів.

Завдання цього розділу включають детальне вивчення та аналіз існуючих методів автентифікації, розробку архітектури системи, яка інтегрується з базою даних для збереження інформації про користувачів, а також реалізацію механізмів шифрування та ключів для захисту даних.

4.1 Структура планованої системи автентифікації

Система автентифікації, що розробляється, базується на комплексному підході до верифікації особистості користувача, використовуючи як біометричні, так і особистісні дані.

Ключовими компонентами системи є модуль біометричного розпізнавання, база даних користувачів, модуль обробки та зберігання персональних даних, а також інтерфейс для взаємодії з користувачами.

Модуль біометричного розпізнавання включає застосування згорткових нейронних мереж для аналізу та класифікації біометричних характеристик, таких як колір очей. Інформація використовується як один із ключових факторів для ідентифікації особи [60].

База даних користувачів зберігає не тільки біометричні дані, але й персональну інформацію користувачів, таку як ім'я, прізвище, дата

народження. Всі дані захищені відповідними механізмами шифрування, щоб забезпечити їх конфіденційність і безпеку.

Модуль обробки даних виконує функції валідації введених даних і зіставлення їх з даними, що зберігаються в базі даних. В разі успішної валідації користувач отримує доступ до свого особистого кабінету або інших захищених ресурсів [61].

Користувацький інтерфейс надає зручні та інтуїтивно зрозумілі засоби для взаємодії з системою, включаючи авторизацію та автентифікацію. Також включає механізми збору згоди на обробку персональних даних, відповідно до законодавчих вимог.

4.2 Архітектура системи автентифікації

Архітектура розробленої системи автентифікації є багаторівневою та інтегрованою, що дозволяє ефективно обробляти і зберігати велику кількість даних, одночасно забезпечуючи швидкий доступ і високий рівень безпеки.

4.2.1 Клієнтська частина

- Веб-інтерфейс дозволяє користувачам реєструватися, вводити свої персональні та біометричні дані.
- Мобільний додаток надає альтернативний доступ до системи через мобільні пристрої.

Система використовує веб-сервіси для прийому персональної інформації та зображень від користувачів. Далі, нейронна мережа обробляє зображення для визначення кольору очей, а сервер автентифікації перевіряє відповідність введених персональних даних з даними в базі. Успішна автентифікація дозволяє користувачу доступ до свого персонального кабінету.

4.2.2 Сервер автентифікації

Модуль прийому даних отримує вхідні дані від користувачів через клієнтські інтерфейси.

- Модуль перевірки біометричних даних здійснює процес верифікації за допомогою нейронної мережі для аналізу біометричних характеристик.
- Модуль авторизації вирішує, чи надавати користувачу доступ до ресурсів системи згідно з результатами верифікації.

Біометричні дані використовуються для перевірки ідентичності особи на основі унікальних фізичних або поведінкових характеристик. У цьому випадку використовуються біометричні характеристики, такі як колір очей, який аналізується за допомогою нейронної мережі, розробленої в попередніх розділах. Класифікатор, навчений на багатокласовій класифікації кольору очей, використовується для екстракції цієї особливості з наданих користувачем зображень [64].

Персональні дані, такі як ім'я, прізвище та дата народження, використовуються для попередньої перевірки особи перед біометричною верифікацією. Ці дані порівнюються з інформацією, що зберігається в базі даних, для підтвердження правильності введених користувачем відомостей.

4.2.3 Модуль безпеки

- Шифрування даних гарантує безпеку персональних та біометричних даних під час їх зберігання та передачі.
- Механізми аудиту та моніторингу забезпечують виявлення та реагування на спроби несанкціонованого доступу.

Кожен з цих компонентів взаємодіє з іншими для забезпечення надійної та безпечної роботи системи. Архітектура розроблена таким чином, щоб оптимізувати процеси автентифікації та мінімізувати можливість помилок чи зловживань [61].

Код автентифікації користувача наданий в Додатку Г.

4.3 Розробка ключів та методів шифрування

В основі системи автентифікації лежить генерація унікальних ключів, які використовуються для захисту доступу до ресурсів. Ці ключі генеруються на основі комбінації біометричних даних і особистої інформації, забезпечуючи високий рівень безпеки.

Використання біометричних даних, таких як колір очей, разом з особистою інформацією (ім'я, прізвище, дата народження) дозволяє створювати складні, важкі для вгадування ключі.

Ключ генерується шляхом застосування криптографічних геш-функцій до даних, отриманих з біометричних сканерів і введених користувачем параметрів. Це може було реалізовано за допомогою алгоритму SHA-256, який перетворює вхідну інформацію в унікальний набір символів фіксованої довжини. Під час розробки системи була врахована потреба у збалансуванні між силою шифрування та швидкістю обробки запитів. Код шифрування наданий в Додатку Д.

Отримані ключі будуть використані для шифрування даних, що передаються між клієнтом та сервером, а також для створення електронних підписів, які підтверджують автентичність та цілісність переданих даних. Шифрування допомагає забезпечити конфіденційність і захист даних користувачів від несанкціонованого доступу [62].

4.4 Застосування гешування для забезпечення цілісності даних

Гешування є процес перетворення вхідних даних будь-якого розміру в невеликий, фіксований розмір біт, відомий як геш-сума, який є унікальним для кожного набору вхідних даних. Іншими словами, навіть мінімальні зміни в даних призводять до зовсім іншого, непередбачуваного шеш-значення.

Гешування застосовується для перевірки цілісності даних, адже перевірка геш-суми розрахованих даних дозволяє швидко ідентифікувати

будь-які зміни, які могли статися під час передачі або зберігання. Це особливо корисно в системах, де необхідно забезпечити незмінність вмісту, наприклад, в системах електронного документообігу, блокчейн технологіях, а також при зберіганні цифрових підписів [63].

Втілений код для гешування за допомогою SHA-256 наданий в додатку Ж.

4.5 Інтеграція системи з користувацьким інтерфейсом

Розробка користувацького інтерфейсу була спрямована на створення простої та зрозумілої взаємодії, яка дозволяє користувачам легко і надійно ідентифікувати себе без зайвих перешкод [56].

Для реалізації користувацького інтерфейсу системи автентифікації, був використаний Python з бібліотекою Tkinter, яка є стандартним інструментом для створення графічного інтерфейсу користувача в Python. Код інтерфейсу наданий в додатку Ж. Дія на кнопку запускає функцію authenticate, яка перевіряє введені дані.

4.6 Висновки до розділу

У четвертому розділі дипломної роботи розроблено повнофункціональну систему автентифікації, яка використовує біометричні дані та особисту інформацію для генерації унікальних ключів. Система об'єднує класичні методи автентифікації з інноваційними підходами на основі нейронних мереж, забезпечуючи високий рівень безпеки та надійність.

Основні переваги розробленої системи включають здатність ефективно використовувати біометричні дані, такі як колір очей, для створення складних механізмів захисту, що ускладнює несанкціонований доступ.

Інтеграція з користувацьким інтерфейсом забезпечує простоту та інтуїтивність використання, а впроваджені заходи захисту відповідають сучасним стандартам безпеки.

ВИСНОВКИ

У даній дипломній роботі була проведена розробка та аналіз системи автентифікації, що інтегрує біометричні дані з традиційними методами верифікації особистості. Дипломна робота охоплює створення згорткових нейронних мереж для розпізнавання біометричних характеристик, розробку системи автентифікації з використанням цих даних, а також застосування криптографічних методів для захисту персональної інформації.

Реалізація нейронних мереж демонструє значний потенціал у класифікації біометричних параметрів, зокрема кольору очей, що дозволяє системі автентифікації діяти більш точно та надійно. Вивчення та застосування різноманітних оптимізаційних алгоритмів і методів тренування підкреслили важливість детального підходу до підготовки та налаштування моделей.

Розроблена модель глибокого навчання включала в себе згорткові шари, пулінг та повнозв'язні шари, а також застосування методів регуляризації для оптимізації результатів та уникнення перенавчання.

Особлива увага була зосереджена на етапі підготовки даних, який охоплював збір, аугментацію та попередню обробку датасету для забезпечення його репрезентативності та різноманітності. Процеси валідації та тренування були виконані з високою мірою деталізації, що дозволило визначити найефективніші налаштування для тренування моделі.

Ефективність моделі була оцінена за допомогою різноманітних метрик, таких як точність, F1-score, Precision, Recall, та ROC-AUC. Проведення перехресної валідації додатково підтвердило надійність і точність результатів. Аналіз чутливості та помилок допоміг виявити ключові фактори, що впливають на точність класифікації, а також визначити обмеження моделі.

Тестування моделі у реальних умовах показало високу ефективність у класифікації кольору очей, з особливим акцентом на дотримання захисту персональних даних та відповідність юридичним нормам. Результати

підтвердили потенціал для подальшого розширення моделі для аналізу інших біометричних характеристик.

Розроблена система автентифікації включає заходи для забезпечення безпеки, такі як шифрування і гешування, що забезпечують цілісність та конфіденційність даних. Інтеграція системи з користувацьким інтерфейсом робить її доступною для кінцевих користувачів, дотримуючись при цьому сучасних стандартів юзабіліті та безпеки.

Однак, система має ряд обмежень, зокрема залежність від якості біометричних даних та необхідність постійного оновлення заходів безпеки у відповідь на нові кіберзагрози. Існують можливості для подальшого розвитку, включаючи розширення біометричних параметрів, використання глибших і складніших нейронних мереж для покращення точності розпізнавання.

Завершуючи, ця дипломна робота вносить значний вклад у область систем автентифікації, демонструючи, як інтеграція нейронних мереж та сучасних криптографічних методів може підвищити ефективність і безпеку автентифікаційних процесів. Ці знахідки можуть слугувати основою для майбутніх досліджень і розвитку у цій динамічно змінюваній і надзвичайно актуальній галузі.

ПЕРЕЛІК ДЖЕРЕЛ ТА ПОСИЛАНЬ

1. Закон України "Про захист інформації в інформаційно-телекомунікаційних системах" від 4 червня 2020 року № 681-IX. Режим доступу: https://ips.ligazakon.net/document/z008000?an=4779&ed=2020_06_04
2. Кіберзагрози в медійному просторі України - armyinform. Режим доступу: https://armyinform.com.ua/tag/kiberzagrozy/
3. Поняття і зміст кіберзагроз на сучасному етапі / І. Діордіца - 2017 - pgp. Режим доступу: http://pgp-journal.kiev.ua/archive/2017/4/22.pdf
4. Сучасні кіберзагрози: як захистити облікові записи працівників - 2023 - forbes. Режим доступу: https://forbes.ua/innovations/suchasni-kiberzagrozi-yak-zakhistiti-oblikovi-zapisi-pratsivnikiv-21122023-18049
5. ЩОДЕННІ КІБЕРЗАГРОЗИ / Міністерство оборони України - 2023 - mil. Режим доступу: https://www.mil.gov.ua/ukbs/shhodenni-kiberzagrozi/
6. Закон України "Про інформацію" від 1 січня 2023 року № 2657-XII. Режим доступу: https://zakon.rada.gov.ua/laws/show/2657-12#Text
7. Закон України "Про основні засади забезпечення кібербезпеки України" від 17 серпня 2022 року № 2163-VIII. Режим доступу: https://zakon.rada.gov.ua/laws/show/2163-19#Text
8. Квантова криптографія та алгоритми / Л. П. Данильченко, А. В. Голованьов, В. С. Заєць та ін. // Збірник наукових праць НТУ "ХПІ". Серія: Нові рішення в сучасних технологіях. - 2016. - Вип. 42. - С. 51-55.
9. Захист інформації в постквантовому світі: теорія та практика / В. М. Голуб, М. М. Кулаков, В. П. Новіков та ін. - Київ: Видавничий дім "Слово", 2021.
10. Closer Look at Authentication and Authorization Mechanisms for Web-based Applications / Sylvia Encheva - 2012 - researchgate. Режим доступу: https://www.researchgate.net/publication/250310860_A_Closer_Look_at_Authentication_and_Authorization_Mechanisms_for_Web-based_Applications

11. A Review on Authentication Methods - 2015 - researchgate. Режим доступу: https://www.researchgate.net/publication/281109747_A_Review_on_Authentication_Methods
12. Network Security, Threats, Authentication, Authorization, and Securing Devices / L. Wenbin - cdn. Режим доступу: https://cdn.stmarytx.edu/wp-content/uploads/2020/10/Network-Security-Threats-Authentication-Authorization-and-Securing-Devices.pdf
13. Identification, Authentication and Authorization on the World Wide Web / E. Kabay - mekabay. Режим доступу: https://www.mekabay.com/infosecmgmt/iaawww.pdf
14. Authentication and Authorization Models / V. More - cscjournals. Режим доступу: https://www.cscjournals.org/manuscript/Journals/IJCSS/Volume5/Issue1/IJCSS-408.pdf
15. Password based authentication/ M. Drasar - is. Режим доступу: https://is.muni.cz/th/98998/fi_m/thesis.pdf
16. CAS - Central Authentication Service Login - mff. Режим доступу: https://www.mff.cuni.cz/cs/spravni-oddeleni/navody/skoleni-ridicu-mff-uk/?fileType=pdf
17. Secure Authentication System based on Multi-Factor Authentication - researchgate. Режим доступу: https://www.researchgate.net/publication/377224776_Secure_Authentication_System_based_on_Multi-Factor_Authentication
18. P 2 Auth: Two-Factor Authentication Leveraging PIN and Keystroke-Induced PPG Measurements / Yuchen Su, Guoqing Jiang, Yicong Du, Yuefeng Chen, Hongbo Liu - 2023 – researchgate. Режим доступу: https://www.researchgate.net/publication/374655849_P_2_Auth_Two-Factor_Authentication_Leveraging_PIN_and_Keystroke-Induced_PPG_Measurements

19. Reimagining Authentication: A User-Centric Two-Factor Authentication with Personalized Image Verification - researchgate. Режим доступу: https://www.researchgate.net/publication/379101135_Reimagining_Authentication_A_User-Centric_Two-Factor_Authentication_with_Personalized_Image_Verification
20. Two Factor Authentication: Voice Biometric and Token-Based Authentication - researchgate. Режим доступу: https://www.researchgate.net/publication/377056656_Two_Factor_Authentication_Voice_Biometric_and_Token-Based_Authentication
21. Two-Factor Authentication Approach Based on Behavior Patterns for Defeating Puppet Attacks - researchgate. Режим доступу: https://www.researchgate.net/publication/377679279_Two-Factor_Authentication_Approach_Based_on_Behavior_Patterns_for_Defeating_Puppet_Attacks
22. Implementation of two-factor user authentication in computer systems / Wenhao Wang, Guyue Li, Zhiming Chu, Haobo Li, Daniele Faccio – 2024 - researchgate. Режим доступу: https://www.researchgate.net/publication/378821207_Implementation_of_two-factor_user_authentication_in_computer_systems
23. Two-Factor Authentication for Internet of Drones Using PUF and Blockchain / Ayan Kumar Das – 2024 - researchgate. Режим доступу: https://www.researchgate.net/publication/378460762_Two-Factor_Authentication_for_Internet_of_Drones_Using_PUF_and_Blockchain
24. Enhancing 802.1X authentication with identity providers using EAP-OAUTH and OAuth 2.0 / Duarte Mortágua, André Zúquete, Paulo Salvador – 2024 - researchgate. Режим доступу: https://www.researchgate.net/publication/379003883_Enhancing_8021X_authentication_with_identity_providers_using_EAP-OAUTH_and_OAuth_20
25. Open Identity Certification with OpenID Connect / Jonas Primbs, Michael Menth – 2024 - researchgate. Режим доступу:

https://www.researchgate.net/publication/372784921_OIDC2_Open_Identity_Certification_with_OpenID_Connect

26. SAML 2 - researchgate. Режим доступу: https://www.researchgate.net/publication/365515053_SAML_2
27. Methods and Benchmark for Detecting Cryptographic API Misuses in Python / Miles Frantz, Ya Xiao, Tanmoy Sarkar Pias, Na Meng, Danfeng (Daphne) Yao - 2024 - researchgate. Режим доступу: https://www.researchgate.net/publication/379066722_Methods_and_Benchmark_for_Detecting_Cryptographic_API_Misuses_in_Python
28. History of Cryptographic Key Sizes - researchgate. Режим доступу: https://www.researchgate.net/publication/356142827_11_-_History_of_Cryptographic_Key_Sizes
29. Neural network (machine learning) – wikipedia. Режим доступу: https://en.wikipedia.org/wiki/Neural_network_(machine_learning)
30. Neural Model - an overview – sciencedirect. Режим доступу: https://www.sciencedirect.com/topics/engineering/neural-model
31. Neural network – wikipedia. Режим доступу: https://en.wikipedia.org/wiki/Neural_network
32. What is a Neural Network? – ibm. Режим доступу: https://www.ibm.com/topics/neural-networks
33. What is a neural network? | Types of neural networks – cloudfare . Режим доступу: https://www.cloudflare.com/learning/ai/what-is-neural-network/
34. Explained: Neural networks – mit. Режим доступу: https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414
35. What Is a Neural Network? – aws. Режим доступу: https://aws.amazon.com/what-is/neural-network/
36. What is a neural network? – geekforgeeks. Режим доступу: https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/
37. What Is a Neural Network? - MATLAB & Simulink – mathworks. Режим доступу: https://www.mathworks.com/discovery/neural-network.html

38. Python AI: How to Build a Neural Network & Make Predictions – realpython. Режим доступу: https://realpython.com/python-ai-neural-network/
39. Neural network models (supervised) – skitit-learn. Режим доступу: https://scikit-learn.org/stable/modules/neural_networks_supervised.html
40. Your First Deep Learning Project in Python with Keras – machinelearningmastery. Режим доступу: https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/
41. Implementing Artificial Neural Network in Python from Scratch – analyticsvidhya. Режим доступу: https://www.analyticsvidhya.com/blog/2021/10/implementing-artificial-neural-networkclassification-in-python-from-scratch/
42. Deep Learning with Python: Neural Networks – nnfs. Режим доступу: https://nnfs.io/
43. Create Your First Neural Network with Python – intel. Режим доступу: https://www.intel.com/content/www/us/en/developer/articles/technical/create-first-neural-network-with-python-tensorflow.html
44. How To Create a Neural Network In Python – activestate. Режим доступу: https://www.activestate.com/resources/quick-reads/how-to-create-a-neural-network-in-python-with-and-without-keras/
45. TensorFlow 2 quickstart for beginners – tensorflow. Режим доступу: https://www.tensorflow.org/tutorials/quickstart/beginner
46. TensorFlow 2 Tutorial: Get Started in Deep Learning – machinelearningmastery. Режим доступу: https://machinelearningmastery.com/tensorflow-tutorial-deep-learning-with-tf-keras/
47. Tutorials – tensorflow. Режим доступу: https://www.tensorflow.org/tutorials
48. Convolutional Neural Network (CNN) | TensorFlow Core – tensorflow. Режим доступу: https://www.tensorflow.org/tutorials/images/cnn
49. An Adaptive Neural Network Model for Clinical Face Mask Detection – researchgate. Режим доступу:

https://www.researchgate.net/publication/374768550_An_Adaptive_Neural_Network_Model_for_Clinical_Face_Mask_Detection

50. semantic neural model approach for face recognition from sketch – researchgate. Режим доступу: https://www.researchgate.net/publication/370469385_semantic_neural_model_approach_for_face_recognition_from_sketch
51. The Neural Dynamics of Face Ensemble and Central Face Processing– researchgate. Режим доступу: https://www.researchgate.net/publication/376838117_The_neural_dynamics_of_face_ensemble_and_central_face_processing
52. Neural network – researchgate. Режим доступу: https://www.researchgate.net/publication/379621297_Gabor_wavelet_and_neural_network_face_detection
53. Gabor wavelet and neural network face detection – researchgate. Режим доступу: https://en.wikipedia.org/wiki/Neural_network
54. Comparison of Convolutional Neural Network (CNN) Models in Face Classification of Papuan and Other Ethnicities – researchgate. Режим доступу: https://www.researchgate.net/publication/369441314_Comparison_of_Convolutional_Neural_Network_CNN_Models_in_Face_Classification_of_Papuan_and_Other_Ethnicities
55. In between faces: Childhood adversity is associated with reduced threat-safety discrimination during facial expression processing in adolescence – researchgate. Режим доступу: https://www.researchgate.net/publication/380276789_In_between_faces_Childhood_adversity_is_associated_with_reduced_threat-safety_discrimination_during_facial_expression_processing_in_adolescence
56. Early Experience of Threat is Associated with Altered Neural Sensitivity for Facial Expressions in Young Adults with Emerging Psychiatric Symptoms – researchgate. Режим доступу: https://www.researchgate.net/publication/380277823_Title_Early_Experience_

of_Threat_is_Associated_with_Altered_Neural_Sensitivity_for_Facial_Expressions_in_Young_Adults_with_Emerging_Psychiatric_Symptoms

57. Real Time Face Mask Detection with TensorFlow and Python – researchgate. Режим доступу: https://www.researchgate.net/publication/379659269_Real_Time_Face_Mask_Detection_with_TensorFlow_and_Python
58. Masked face age and gender identification using CAFFE-modified MobileNetV2 on photo and real-time video images by transfer learning and deep learning techniques – researchgate. Режим доступу: https://www.researchgate.net/publication/377358747_Masked_face_age_and_gender_identification_using_CAFFE-modified_MobileNetV2_on_photo_and_real-time_video_images_by_transfer_learning_and_deep_learning_techniques
59. Intelligent Face Recognition Based Students' Attendance System– researchgate. Режим доступу: https://www.researchgate.net/publication/380528380_Intelligent_Face_Recognition_Based_Students%27_Attendance_System
60. Face Mask Recognition Menggunakan Model CNN (Convolutional Neural Network) Berbasis Python dan OpenCV – researchgate. Режим доступу: https://www.researchgate.net/publication/375081629_Face_Mask_Recognition_Menggunakan_Model_CNN_Convolutional_Neural_Network_Berbasis_Python_dan_OpenCV
61. Secured shared authentication key with two-way clock synchronization over multiparty quantum communication – researchgate. Режим доступу: https://www.researchgate.net/publication/375695678_Secured_shared_authentication_key_with_two-way_clock_synchronization_over_multiparty_quantum_communication
62. Face Recognition Authentication System with CNN and Blink Detection Algorithm – researchgate. Режим доступу:

https://www.researchgate.net/publication/372618579_Face_Recognition_Authentication_System_with_CNN_and_Blink_Detection_Algorithm

63. Improvised Multi-Factor Authentication for End-User Security in Cyber Physical System – researchgate. Режим доступу: https://www.researchgate.net/publication/373571979_Improvised_Multi-Factor_Authentication_for_End-User_Security_in_Cyber_Physical_System
64. Implementation of Data Layer In Blockchain Network Using SHA256 Hashing Algorithm – researchgate. Режим доступу: https://www.researchgate.net/publication/380230645_Implementation_of_Data_Layer_In_Blockchain_Network_Using_SHA256_Hashing_Algorithm

Додаток А

Код нейронної моделі на основі бібліотеки Tensorflow

```
import os
import tensorflow as tf
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout

dataset_path = 'C:\\Users\\misha\\Desktop\\Bachelor\\pictures'

img_width, img_height = 64, 64

train_datagen = ImageDataGenerator(rescale=1./255, validation_split=0.2)

train_generator = train_datagen.flow_from_directory(
    dataset_path,
    target_size=(img_width, img_height),
    batch_size=32,
    class_mode='categorical',
    subset='training')

validation_generator = train_datagen.flow_from_directory(
    dataset_path,
    target_size=(img_width, img_height),
    batch_size=32,
    class_mode='categorical',
    subset='validation')
```

```
        model = Sequential()
        model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(img_width,
img_height, 3)))
        model.add(MaxPooling2D(2, 2))
        model.add(Conv2D(64, (3, 3), activation='relu'))
        model.add(MaxPooling2D(2, 2))
        model.add(Flatten())
        model.add(Dense(64, activation='relu'))
        model.add(Dropout(0.5))
        model.add(Dense(5, activation='softmax'))

        model.compile(loss='categorical_crossentropy',
                      optimizer='adam',
                      metrics=['accuracy'])

        model.fit(
           train_generator,
           steps_per_epoch=train_generator.samples // train_generator.batch_size,
           validation_data=validation_generator,
           validation_steps=validation_generator.samples                    //
validation_generator.batch_size,
           epochs=10)

model.save('eye_color_model.h5')



        import tensorflow as tf
        from tensorflow.keras.preprocessing import image
        import numpy as np
```

```
        model = tf.keras.models.load_model('C:\\Users\\misha\\Desktop\\Bachelor\\eye_color_model.h5')

        def predict_eye_color(img_path):
            img = image.load_img(img_path, target_size=(64, 64))
            img_array = image.img_to_array(img)
            img_array = np.expand_dims(img_array, axis=0) / 255.0

            prediction = model.predict(img_array)
            class_indices = ['Blue', 'Brown', 'Green', 'Grey', 'Yellow']
            predicted_class = class_indices[np.argmax(prediction)]

            return predicted_class

        test_image = 'try.jpg'
print(predict_eye_color(test_image))
```

Додаток Б

Код нейронної моделі на основі бібліотеки Torch

```
import torch
from torchvision import transforms, datasets
from torch.utils.data import DataLoader

dataset_path = 'C:\\Users\\misha\\Desktop\\Bachelor\\pictures'

transform = transforms.Compose([
    transforms.Resize((64, 64)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
])

dataset = datasets.ImageFolder(root=dataset_path, transform=transform)
train_size = int(0.8 * len(dataset))
val_size = len(dataset) - train_size
train_dataset, val_dataset = torch.utils.data.random_split(dataset, [train_size, val_size])

train_loader = DataLoader(train_dataset, batch_size=32, shuffle=True)
val_loader = DataLoader(val_dataset, batch_size=32, shuffle=False)

import torch.nn as nn
import torch.nn.functional as F

class EyeColorNet(nn.Module):
```

```
        def __init__(self):
            super(EyeColorNet, self).__init__()
            self.conv1 = nn.Conv2d(3, 32, 3, padding=1)
            self.pool = nn.MaxPool2d(2, 2)
            self.conv2 = nn.Conv2d(32, 64, 3, padding=1)
            self.fc1 = nn.Linear(64 * 16 * 16, 128)
            self.fc2 = nn.Linear(128, 5)

        def forward(self, x):
            x = self.pool(F.relu(self.conv1(x)))
            x = self.pool(F.relu(self.conv2(x)))
            x = x.view(-1, 64 * 16 * 16)
            x = F.relu(self.fc1(x))
            x = self.fc2(x)
    return x

    import torch
    import torch.optim as optim
    from model import EyeColorNet
    from data_loader import train_loader, val_loader

    device = torch.device("cuda" if torch.cuda.is_available() else "cpu")

    model = EyeColorNet().to(device)
    criterion = nn.CrossEntropyLoss()
    optimizer = optim.Adam(model.parameters(), lr=0.001)

    for epoch in range(10):
        model.train()
```

```
            for inputs, labels in train_loader:
                inputs, labels = inputs.to(device), labels.to(device)
                optimizer.zero_grad()
                outputs = model(inputs)
                loss = criterion(outputs, labels)
                loss.backward()
                optimizer.step()

            model.eval()

torch.save(model.state_dict(), 'eye_color_model.pth')
```

Додаток В

Використання оптимізаторів у нейронній мережі на основі Tensorflow

```
from tensorflow.keras.optimizers import Adam, SGD, RMSprop

model.compile(loss='categorical_crossentropy',
              optimizer=Adam(learning_rate=0.001),
              metrics=['accuracy'])

model.compile(loss='categorical_crossentropy',
              optimizer=SGD(learning_rate=0.001),
              metrics=['accuracy'])

model.compile(loss='categorical_crossentropy',
              optimizer=RMSprop(learning_rate=0.001),
    metrics=['accuracy'])
```

Додаток Г

Створення методу автентифікації

```
import tensorflow as tf
from tensorflow.keras.models import load_model
from tensorflow.keras.preprocessing import image
import numpy as np

model = load_model('eye_color_model.h5')

def authenticate_user(name, surname, dob, eye_image_path):
    user_record = database.get_user_record(name, surname, dob)
    if user_record is None:
        return False, "User not found"

    img = image.load_img(eye_image_path, target_size=(64, 64))
    img_array = image.img_to_array(img)
    img_array = np.expand_dims(img_array, axis=0) / 255.0

    predictions = model.predict(img_array)
    predicted_color = class_indices[np.argmax(predictions)]

    if predicted_color == user_record['eye_color']:
        return True, "User authenticated"
    else:
        return False, "Authentication failed"

authenticate_user("Mykhailo", "Kononenko", "2003-09-01",
"Kononenko_Eye.jpg")
```

Додаток Д

Генерація ключів за допомогою SHA256

```
from cryptography.hazmat.primitives.ciphers import Cipher, algorithms, modes
from cryptography.hazmat.primitives.kdf.pbkdf2 import PBKDF2HMAC
from cryptography.hazmat.primitives import hashes
from cryptography.hazmat.backends import default_backend
from os import urandom

def generate_key(password: bytes, salt: bytes) -> bytes:
    kdf = PBKDF2HMAC(
        algorithm=hashes.SHA256(),
        length=32,
        salt=salt,
        iterations=100000,
        backend=default_backend()
    )
    return kdf.derive(password)

def encrypt_data(data: bytes, key: bytes) -> bytes:
    iv = urandom(16)  # Генерація випадкового IV (ініціалізаційного вектора)
    cipher = Cipher(algorithms.AES(key), modes.CFB(iv), backend=default_backend())
    encryptor = cipher.encryptor()
    encrypted_data = encryptor.update(data) + encryptor.finalize()
    return iv + encrypted_data  # Повертаємо IV разом з зашифрованими даними
```

```
        def decrypt_data(encrypted_data: bytes, key: bytes) -> bytes:
            iv = encrypted_data[:16]
            cipher = Cipher(algorithms.AES(key), modes.CFB(iv),
backend=default_backend())
            decryptor = cipher.decryptor()
            return decryptor.update(encrypted_data[16:]) + decryptor.finalize()

        password = b'my_very_secret_password'
        salt = urandom(16)
        key = generate_key(password, salt)

        original_data = b'Hello, this is a secret message!'
        encrypted_data = encrypt_data(original_data, key)
        decrypted_data = decrypt_data(encrypted_data, key)

        print("Original:", original_data)
        print("Encrypted:", encrypted_data)
print("Decrypted:", decrypted_data)
```

Додаток Е

Гешування ключів за допомогою SHA256

```
import hashlib


def generate_hash(data: bytes) -> str:
    hash_object = hashlib.sha256()
    hash_object.update(data)
    return hash_object.hexdigest()
```

Додаток Ж

Інтерфейс користувача розроблений на основі Tkinter.

```
import tkinter as tk
from tkinter import messagebox

def authenticate():
    name = name_entry.get()
    surname = surname_entry.get()
    dob = dob_entry.get()
    eye_image_path = eye_image_path _entry.get()
    def authenticate_user(name, surname, dob, eye_image_path):
        messagebox.showinfo("Login Success", "You are now logged in.")
    else:
        messagebox.showinfo("Login Failed", "Incorrect username or password.")

root = tk.Tk()
root.title("Authentication System")
root.geometry("300x150")

tk.Label(root, text="Username:").pack()
username_entry = tk.Entry(root, width=25)
username_entry.pack()

tk.Label(root, text="DOB:").pack()
dob_entry = tk.Entry(root, width=25)
dob_entry.pack()
```

```
        tk.Label(root, text="Surname:").pack()
        surname_entry = tk.Entry(root, width=25)
        surname_entry.pack()

        tk.Label(root, text="Path:").pack()
        eye_image_path_entry = tk.Entry(root, width=25)
        eye_image_path_entry.pack()



        login_button = tk.Button(root, text="Login", command=authenticate)
        login_button.pack()

root.mainloop()
```