Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему *Метод виявлення модифікованих фотографій облич людей нейромережевими засобами*

Галузь знань *12 – Інформаційні технології*
Шифр і назва галузі знань

Спеціальність *122 – Комп'ютерні науки*
Шифр і назва спеціальності

Освітня програма *Комп'ютерні науки*
Назва освітньої програми

Виконав: *студент групи КНм-23-1* _____ *Андрій ПОХИТУН*
Група виконавця | Підпис | Ім'я, ПРІЗВИЩЕ

Керівник: *зав. каф. КН, д.т.н., проф.* _____ *Олександр БАРМАК*
Науковий ступінь, посада | Підпис | Ім'я, ПРІЗВИЩЕ

Нормоконтроль: *к.т.н., доц. каф. КН* _____ *Руслан БАГРІЙ*
Науковий ступінь, посада | Підпис | Ім'я, ПРІЗВИЩЕ

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор _____ *Олександр БАРМАК*
Підпис | Ім'я, ПРІЗВИЩЕ

16 грудня 2024 р.

Хмельницький 2024

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет *інформаційних технологій*
Кафедра *комп'ютерних наук*
Освітній ступінь *магістр*
Галузь знань *12 – Інформаційні технології*
Спеціальність *122 – Комп'ютерні науки*

ЗАТВЕРДЖУЮ
Завідувач кафедри *комп'ютерних наук*

_____
(підпис)
д.т.н., професор Олександр БАРМАК
«*02*» *вересня* *2024* року

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА

1. Тема кваліфікаційної роботи магістра: *«Метод виявлення модифікованих фотографій облич людей нейромережевими засобами»*

2. Завдання видано студенту *Андрію ПОХИТУНУ*
(прізвище, ім'я, по батькові)

3. Керівник роботи *зав. кафедри КН, д.т.н, проф. Олександр БАРМАК*
(прізвище, ім'я, по батькові)

4. Затверджені наказом університету від «*26*» *Серпня* *2024* р. № *60.*

5. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Для досягнення мети, що полягає у підвищенні точності виявлення модифікованих фотографій облич людей, необхідно провести аналіз подібного програмного забезпечення; дослідити доступні в мережі набори даних, та підготувати власний датасет для навчання згорткової нейронної мережі; провести аналіз та обрати метод виявлення та виокремлення обличчя на світлинах; розробити метод виявлення модифікованих зображень облич людей; провести тестування та аналіз поданого методу виявлення модифікованих зображень облич людей. Метод вихідними даними має виконувати визначення зони обличчя людини, визначення ймовірності модифікації фото та визначення ймовірностей для кожного типу модифікації фото.

Реферат

Кваліфікаційна робота магістра присвячена розробці методу виявлення модифікованих облич людей нейромережевими засобами, який дозволяє підвищити точність виявлення модифікованих фотографій облич людей.

Актуальність теми. Сучасний світ характеризується досить стрімким розвитком цифрових технологій, пов’язаних із обробкою зображень. Проте із стрімким розвитком цифрових технологій почали з’являтись проблеми із неправдивою, підробленою інформацією, та способами її виявлення. Зокрема, методи створення модифікованих зображень облич, такі як морфінг, дипфейк, та різні алгоритми генерації штучних зображень, використовуються не тільки у розважальних цілях, а й у злочинних.

Особливо це важливо в умовах гібридної війни, яку веде російська федерація проти України, оскільки мережа, стає полем для боротьби, де інформаційні атаки, включаючи неправдиві, модифіковані фото та відео, можуть мати значний вплив на моральний стан суспільства.

Розробка методу виявлення модифікованих зображень облич людей, викликає декілька складних проблеми, а саме:

– швидкий розвиток алгоритмів модифікування зображень;

– велика кількість різноманітних зображень із різною складністю модифікацій;

– обмежена кількість навчальних даних;

– високі вимоги до обчислювальних ресурсів;

Тому, дослідження предметної області модифікованих зображень облич людей, є досить актуальним на даний час.

Отже, розробка методу для виявлення модифікованих облич людей може значно покращити стан у багатьох сферах життя та допоможе боротись із неправдивими новинами, сприятиме підвищенню довіри до соціальних мереж та може стати потужним інструментом у боротьбі із рядом інших проблем.

Мета і задачі роботи. Мета кваліфікаційної роботи полягає у підвищенні точності виявлення модифікованих фотографій облич людей. Для досягнення мети, обрано наступні пункти для дослідження:

– провести аналіз подібного програмного забезпечення;

– дослідити доступні в мережі набори даних, та підготувати власний датасет для навчання згорткової нейронної мережі;

– провести аналіз та обрати метод виявлення та виокремлення обличчя на фото;

– розробити метод виявлення модифікованих зображень облич людей;

– провести тестування та аналіз поданого методу виявлення модифікованих зображень облич людей.

Об’єкт дослідження – процес виявлення модифікованих облич людей на зображенні.

Предмет дослідження – моделі, методи та засоби для виявлення модифікованих облич людей на зображенні за допомогою нейромережевої класифікації

Методи дослідження, використанні для вирішення поставлених задач: методи аналізу зображень, згорткові нейронні мережі, методи виявлення та виокремлення облич людей на зображеннях.

Наукова новизна одержаних результатів. Створено новий метод виявлення модифікованих зображень облич людей, який дозволяє виявляти не лише наявність модифікації зображення обличчя, а і спосіб її одержання.

Практичне значення одержаних результатів. Було створено інформаційну систему автоматизованого виявлення модифікованих зображень облич людей, яка дозволяє за фотографією оцінити ступінь наявності модифікації обличчя людини та її вид і є програмною реалізацією створеного методу виявлення модифікованих зображень облич людей.

Інформаційна структура системи складається із набору даних у вигляді датасету, та трьох підсистем: «Підсистема навчання, донавчання нейромережі», «Підсистема використання попередньо навченої нейромережі», та головна

«Підсистема класифікації зображення», яка визначає наявність та тип модифікації на фото, яке завантажив користувач. Дані підсистеми реалізують основний функціонал методу виявлення модифікованих зображень облич людей. Виконані дослідження підтвердили зростання точності ідентифікації модифікації облич на зображеннях, при цьому метрики становили: Accuracy 0.99, Precision 0.98, Recall 0.98, F1 0.98, що є підвищенням на понад 0.04 для кожної метрики.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні наукові й практичні результати роботи доповідались у доповіді «Method for Neural Network Detecting Changed Images of People's Faces Using CNN» на I Міжнародній науково-практичній конференції «New Horizons in Scientific Research: Challenges and Solutions» (Marseille, France) 21-23 жовтня 2024 року та у доповіді «Підхід до формування датасету для нейромережевого виявлення модифікованих фотографій облич людей» на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп’ютерних наук» (м. Хмельницький) 15-16 листопада 2024 року. За темою кваліфікаційної роботи магістра автором виконано три наукові публікації.

Структура та обсяг роботи. Кваліфікаційна робота магістра складається з реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань із 50 найменувань та 5 додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 86 сторінок. У роботі наведено 74 рисунків та 10 таблиць.

Ключові слова: модифіковані зображення, згорткова нейронна мережа, глибинне машинне навчання.

Зміст

Перелік скорочень

Скорочення, термін, позначення	Пояснення
ЗНМ	Згорткова нейронна мережа
НМ	Нейронна мережа
ШІ	Штучний інтелект
КН	Комп’ютерні науки
AMSL	Average Mean Shape Learning
ПЗ	Програмне забезпечення
JSON	JavaScript object notation
BIN	Binary files
ІС	Інформаційна система

Вступ

Актуальність теми. Сучасний світ характеризується досить стрімким розвитком цифрових технологій, пов’язаних із обробкою зображень. Проте із стрімким розвитком цифрових технологій почали з’являтись проблеми із неправдивою, підробленою інформацією, та способами її виявлення. Зокрема, методи створення модифікованих зображень облич, такі як морфінг, дипфейк, та різні алгоритми генерації штучних зображень, використовуються не тільки у розважальних цілях, а й у злочинних.

Особливо це важливо в умовах гібридної війни, яку веде російська федерація проти України, оскільки мережа, стає полем для боротьби, де інформаційні атаки, включаючи неправдиві, модифіковані фото та відео, можуть мати значний вплив на моральний стан суспільства.

Розробка методу виявлення модифікованих зображень облич людей, викликає декілька складних проблеми, а саме:

– швидкий розвиток алгоритмів модифікування зображень;

– велика кількість різноманітних зображень із різною складністю модифікацій;

– обмежена кількість навчальних даних;

– високі вимоги до обчислювальних ресурсів;

Тому, дослідження предметної області модифікованих зображень облич людей, є досить актуальним на даний час.

Отже, розробка методу для виявлення модифікованих облич людей може значно покращити стан у багатьох сферах життя та допоможе боротись із неправдивими новинами, сприятиме підвищенню довіри до соціальних мереж та може стати потужним інструментом у боротьбі із рядом інших проблем.

Мета і задачі роботи. Мета кваліфікаційної роботи полягає у підвищенні точності виявлення модифікованих фотографій облич людей. Для досягнення мети, обрано наступні пункти для дослідження:

– провести аналіз подібного програмного забезпечення;

– дослідити доступні в мережі набори даних, та підготувати власний датасет для навчання згорткової нейронної мережі;

– провести аналіз та обрати метод виявлення та виокремлення обличчя на фото;

– розробити метод виявлення модифікованих зображень облич людей;

– провести тестування та аналіз поданого методу виявлення модифікованих зображень облич людей.

Об’єкт дослідження – процес виявлення модифікованих облич людей на зображенні.

Предмет дослідження – моделі, методи та засоби для виявлення модифікованих облич людей на зображенні за допомогою нейромережевої класифікації

Методи дослідження, використанні для вирішення поставлених задач: методи аналізу зображень, згорткові нейронні мережі, нейромережеві методи виявлення та виокремлення обличчя.

Наукова новизна одержаних результатів. Створено новий метод виявлення модифікованих зображень облич людей, який дозволяє виявляти не лише наявність модифікації зображення обличчя, а і спосіб її одержання.

Апробація результатів кваліфікаційної роботи магістра та публікації. Основні наукові й практичні результати роботи доповідались у доповіді «Method for Neural Network Detecting Changed Images of People's Faces Using CNN» на I Міжнародній науково-практичній конференції «New Horizons in Scientific Research: Challenges and Solutions» (Marseille, France) 21-23 жовтня 2024 року та у доповіді «Підхід до формування датасету для нейромережевого виявлення модифікованих фотографій облич людей» на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп’ютерних наук АПКН-2024» (м. Хмельницький) 15-16 листопада 2024 року.

За темою кваліфікаційної роботи магістра автором виконано три наукові публікації:

1. Pokhytun A., Mazurets O., Molchanova M., Tyschenko O. Method for Neural Network Detecting Changed Images of People's Faces Using CNN. New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International scientific and practical conference. October 21-23, 2024. Marseille, France. 2024. Pp. 35-40.

2. Похитун А.В., Мазурець О.В., Молчанова М.О., Бармак О.В. Підхід до формування датасету для нейромережевого виявлення модифікованих фотографій облич людей. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп’ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 428-433.

3. Похитун А.В., Мазурець О.В., Дидо Р.А., Молчанова М.О. Програмна архітектура для нейромережевого виявлення модифікованих фотографій облич людей. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1 (Довідка з редакції).

Структура та обсяг роботи. Кваліфікаційна робота магістра складається з реферату, завдання, змісту, переліку скорочень, вступу, 4 розділів, висновків, переліку посилань із 50 найменувань та 5 додатків. Обсяг основного тексту кваліфікаційної роботи магістра становить 86 сторінок. У роботі наведено 74 рисунків та 10 таблиць.

РОЗДІЛ 1 Характеристика предметної області та постановка задачі

1.1 Аналіз предметної області виявлення модифікованих облич людей

Автоматизація процесів виявлення модифікованих облич людей є важливою задачею інформаційних технологій, яка дозволить підвищити швидкість та надійність ідентифікації.

У сучасному світі змінити зовнішній вигляд по фото не є важкою задачею. Одним із найбільш впливових факторів є поява генеративних моделей які здатні створювати досить реалістичні зображення облич людей, або змінювати їх.

Серед великої кількості видів модифікацій, можна виділити декілька, які користуються найбільшою популярністю

– цифрові фільтри та косметичні зміни;
– глибинні фейки;
– морфінг;
– алгоритм обробки зображень.

Цифрові фільтри – один із найпоширеніших різновидів модифікацій фото у сучасному суспільстві. Даний вид модифікацій фото отримав свою популярність через зростання кількості користувачів соціальних мереж. За допомогою фільтрів користувачі могли змінити свою зовнішність всього за декілька натискань по екрану (Рисунок 1.1).

Рисунок 1.1 – приклад застосування цифрових фільтрів [1]

Глибинні фейки – це технологія, що дозволяє змінити обличчя на фото чи відео за допомогою нейронної мережі [2]. Алгоритми які використовуються у цій технології дозволяють не лише створювати нові зображення а й можуть накладати зображення на зображення, що дозволяє замінити на всьому фото лише обличчя людини (Рисунок 1.2).

Рисунок 1.2 – Приклад застосування глибинних фейків [3]

Морфінг – технологія, що дозволяє створити фейкові зображення за допомогою плавного переходу фото від одного до іншого [4]. Морфінг може використовуватися зловмисниками для підробки ідентифікаційних документів. Приклад використання технології зображено на Рисунку 1.3.

Рисунок 1.3 – Приклад використання морфінгу [5]

Алгоритми обробки зображень – алгоритми що включають в себе простіші методи зміни зображень за допомогою програмного забезпечення [6]. Одним із найвідоміших прикладів програмного забезпечення є усім відомий Adobe Photoshop.

Adobe Photoshop – це потужний графічний редактор який зазвичай використовують для роботи із растровими зображеннями [7].

Алгоритми обробки зображень також часто використовуються в кінематографії. За допомогою різного програмного забезпечення можна змінити риси обличчя, текстуру шкіри, колір шкіри, колір волосся і багато чого іншого Рисунок 1.4.

Рисунок 1.4 – Приклад застосування алгоритму обробки зображень [8]

В сучасному світі існує досить багато можливостей для коригування, редагування та обробки зображень, тому важливо розрізняти фейкові зображення від реальних оскільки модифіковані фото можуть використовуватись не лише для розважальних цілей а й для підробки документів, цькування, що може мати досить негативні наслідки.

Отже, в роботі буде автоматизовано процес виявлення цифрових фільтрів та косметичних змін, глибинних фейків, морфінгу та застосування алгоритмів обробки зображень.

1.2 Особливості застосування згорткових нейронних мереж при роботі із зображеннями обличчя

Нейронна мережа – це застосунок, або обчислювальна модель машинного навчання, яка приймає рішення подібно до людського мозку та складається з великої кількості взаємопов’язаних елементів – нейронів [9]. Нейронні мережі використовуються для розв’язання різноманітних завдань, зокрема в обробці зображень, розпізнавання мови, тексту та у багатьох інших сферах.

Генеративні моделі – це види машинного навчання, що можуть генерувати нові данні подібні до тих, на яких була навчена модель [10]. Це можуть бути не лише зображення а й фото, тексти, відео тощо.

Згорткова нейронна мережа – це категорія моделі машинного навчання, а саме тип алгоритму глибинного навчання, який в свою чергу добре підходить для аналізу візуальних даних [11].

Згорткові нейронні мережі є досить потужним інструментом при роботі із зображеннями, зокрема обличчям. Завдяки своїй архітектурі, яка досить добре підходить для виявлення та аналізу просторових і часових шаблонів у візуальних даних, ЗНМ є одним із кращих варіантів для роботи із фото.

Згорткові нейронні мережі виконують операцію згортки, які дають змогу виявити локальні особливості в зображеннях. Шари згортки здатні автоматично

навчатись на основі даних. Згортка – це проста математична операція, яка зазвичай використовується для обробки та розпізнавання зображення [12] (Рисунок 1.5).

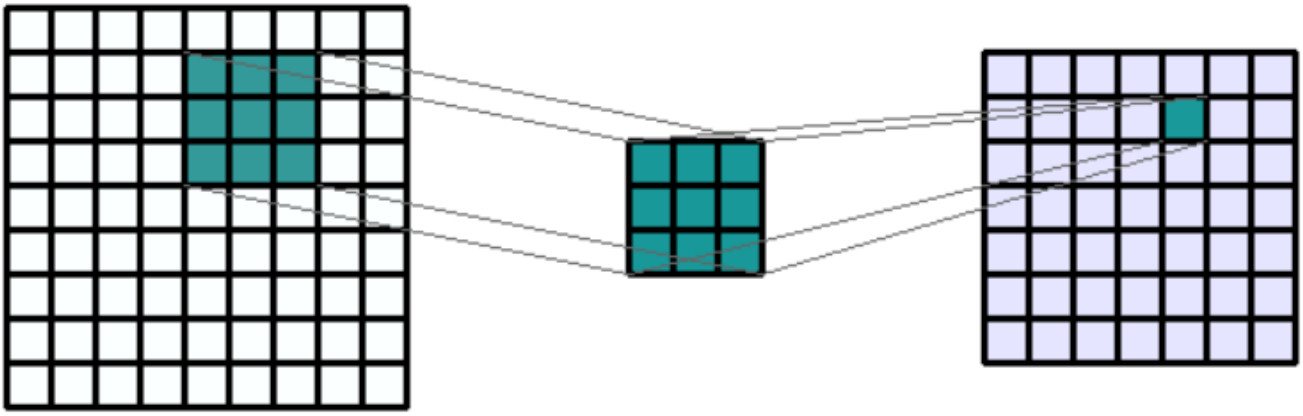


Рисунок 1.5 – Візуалізація процесу згортки [13]

Кінцева мета згортки полягає у тому, щоб виділити специфічні особливості із зображення та зберегти їх у вигляді вектора ознак. Оскільки в ході розробки методу для виявлення модифікованих зображень обличчя будуть використовуватись кольорові зображення замість фільтру 3x3, як зображено на рисунку 1.1, буде використовуватись фільтр 3x3x3, що дозволяє обробляти всі три канали(R,G,B).

Після виконання згортки, застосовується функція активації. Функції активації в згорткових нейронних мережах застосовуються для вивчення нелінійності між входом та виходом. Зазвичай ця функція застосовується до входу кожного нейрона, який приймає зменшену кількість вхідних даних. Найбільш поширеною функцією активації в згорткових нейронних мережах є ReLU. ReLU – функція активації, яка має тенденцію створювати розріджені представлення вхідних даних (Рисунок 1.6) [14].

Отже функції активації додають моделі нелінійності та дозволяє нейронній мережі навчитись більш складним патернам.

Пулігові шари – в згорткових нейронних мережах призначені для зменшення розмірності данних, що покращує ефективність обчислень і зберігає важливі ознаки зображення [15]. Шари пулінгу бувають декількох видів, а саме максимальний, глобальний та середній.

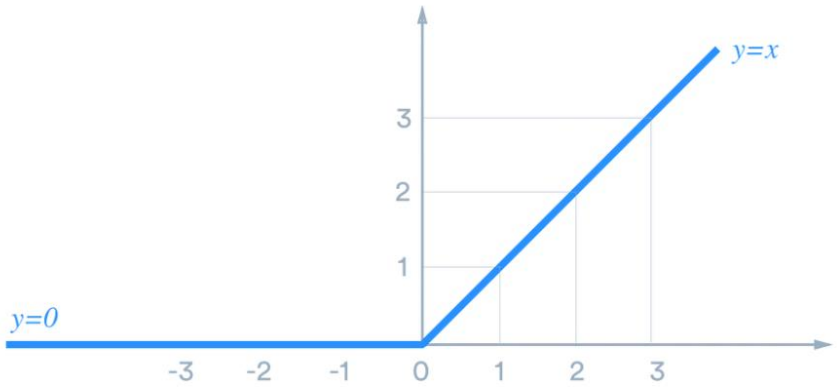


Рисунок 1.6 – Візуалізація функції ReLU [16]

Максимальна вибірка (MaxPooling) – операція об’єднання, яка обчислює максимальне значення фрагменту серед елементів у кожному фільтрі (Рисунок 1.7) [17].

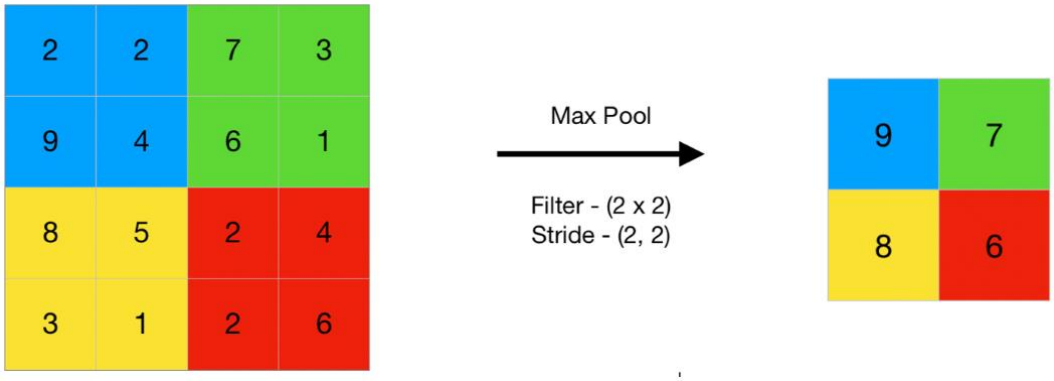


Рисунок 1.7 – Візуалізація максимальної вибірки [17]

Середня вибірка(Average Pooling) – працює аналогічно максимальній вибірці, проте обирає не найбільше значення а середнє (Рисунок 1.8).

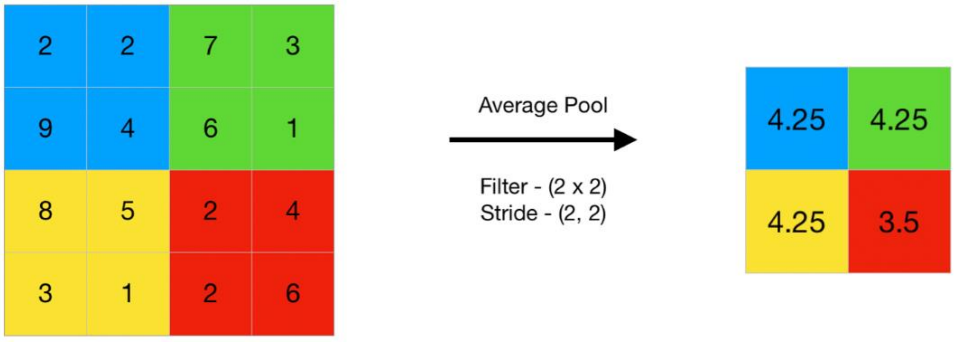


Рисунок 1.8 – Візуалізація середньої вибірки [18]

Таким чином, було розглянуто особливості згорткових нейронних мереж, які будуть використані у якості нейромережевих моделей для виявлення модифікацій та виявлення обличчя на фотозображенні.

1.3 Аналіз сучасних наукових публікацій з проблеми виявлення модифікованих фотографій облич людей

Розвиток цифрових технологій призвів до створення програм, які можуть становити загрозу для демократії, національної безпеки та конфіденційності, зокрема через технології deepfake. Контент типу «deep fake», зокрема у вигляді зображень, поширюється з небаченою швидкістю. Такий фальшивий контент створюється за допомогою передових алгоритмів глибокого навчання, таких як GAN, автокодери та варіаційні автокодери. Це явище сприяє поширенню дезінформації, що суттєво впливає на суспільство, знижуючи рівень довіри до контенту в соціальних мережах, тому дана тема приваблює увагу науковців. Наприклад, дослідження [19] присвячене аналізу різних методів виявлення deep fake, які навчаються на малих вибірках даних. Запропонована робота демонструє ефективну модель CNN та три попередньо навчені моделі CNN, які використовують метод переносу навчання на великому наборі даних з Kaggle, що містить 140 тисяч зображень облич. Запропонована модель CNN досягла точності 96%, в той час як DenseNet121 – 97%.

Дослідження [20] надає огляд літератури щодо методів виявлення deep fake за допомогою DL-алгоритмів, категоризуючи їх за застосуваннями: відео, зображення, аудіо та гібридні мультимедійні методи. Метою є допомогти читачам краще зрозуміти, як генеруються та виявляються deep fake, останні досягнення в цій сфері, слабкі місця існуючих методів безпеки та напрямки для подальших досліджень. Результати показують, що найбільш поширеним методом у публікаціях є використання згорткових нейронних мереж.

У дослідженні [21] було застосовано унікальну активну судово-експертну стратегію на основі архітектури Compact Ensemble-дискримінаторів з

використанням глибоких умовних генеративних суперечливих мереж (CED-DCGAN) для виявлення deep fake в реальному часі під час відеоконференцій. DCGAN зосереджується на виявленні deep fake у відео, яке розбивається по кадрам, аналізуючи характеристики, оскільки технології створення переконливих підробок швидко розвиваються.

Аналіз сучасних розробок та публікацій, пов’язаних із виявленням модифікованих фотографій облич людей, зосереджується на кількох ключових напрямках, які показують як розвиток технологій так і нові виклики та проблеми в даній галузі. Процес розпізнавання обличчя стикається із багатьма проблемами, такими як варіації поз обличчя, зміна освітлення, більшість алгоритмів значно піддаються впливу цих змін [22].

Освітлення є однією із основних проблем з якою стикаються розробники методів для виявлення не лише модифікованих фотографій облич людей, а й облич в цілому [23]. Метод розпізнавання обличчя користувача не може його розпізнати через надмірне освітлення. Освітлення, затемнення частини фото, наявність додаткових елементів, все це спричиняє оклюзію. Оклюзія – це явище, при якому один чи декілька об’єктів повністю або частково перекриває інший об’єкт [24]. Приклади оклюзії зображені на Рисунку 1.9.

Рисунок 1.9 – Приклади оклюзії [25]

У 2022 році, було проведено експеремент, в якому перевіряли як будуть працювати найпопулярніші, на той час, алгоритми розпізнавання облич на фото та відео із оклюзією. Якщо людина була у масці, то похибка становила від 20% до 50%, що є досить критично [26].

Ще однією проблемою в сфері розпізнавання облич людей – проблема із наборами даних. Існує постійна потреба у більш різноманітних та репрезентативних наборах даних, щоб зменшити зміщення в алгоритмах розпізнавання облич [27]. Інша проблема пов'язана зі змінами обличчя внаслідок старіння, оскільки цей процес впливає на його форму та структуру. Крім того, важливо забезпечити баланс між точністю розпізнавання, конфіденційністю та безпекою.

Під час розпізнавання обличчя на зображенні, макіяж може значно вплинути на результат. Макіяж можна назвати оптичним видом оклюзії, що змінює сприйняття природних рис обличчя. У 2023 році було проведено дослідження, під час якого було зроблене фото людини з макіяжем та без нього. Із повним макіяжем на обличчі метод видавав 5% похибки, порівнюючи із зображенням без макіяжу [28]. Також було помічено, що наявність помади на обличчі ніяк не впливає на результат роботи методу.

Отже, дана тематика є актуальною, про що свідчить висока увага науковців. Однак, розглянуті дослідження мають недоліки та є недосконалими, що свідчить про необхідність подальших наукових досліджень.

1.4 Аналіз існуючого програмного забезпечення для виявлення модифікацій фотографій облич людей

На сьогоднішній день існує велика кількість програм для редагування фотографій, активно почали розроблятись також різноманітні програми для виявлення цих змін. Щоб перевірити функціональність сайтів було вибрано кілька зображень із датасету (описаного нижче). Обрані зображення різного рівня складності – від найпростіших, модифікації яких можна легко помітити

«неозброєним» оком, до більш складних і якісно модифікованих фото (Рисунок 1.10).

Рисунок 1.10 – Набір фото для тестування готового програмного забезпечення

Одним із таких сервісів є imageedited. Imageedited – зручний сервіс, що має лише одну функцію, а саме завантаження зображення та перевірка, чи було фото модифіковане (Рисунок 1.11) [29].

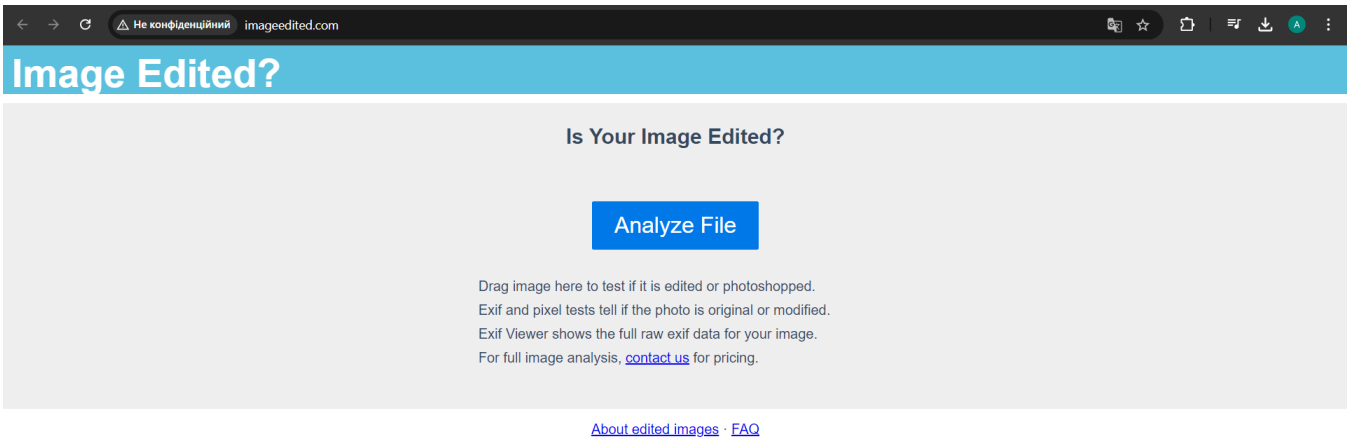


Рисунок 1.11 – Головна сторінка веб-сайту imageedited [29]

Після того, як зображення було завантажене, сервіс відображає результат (Рисунок 1.12).

Сайт має простий дизайн, проте виявив два із трьох зображень як «модифіковані», що є досить хорошим показником.

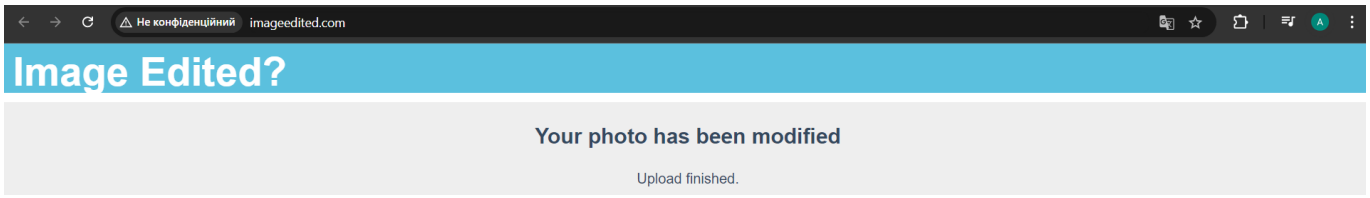


Рисунок 1.12 – Результат роботи сайту imageedited [29]

Іншим подібним, але більш багатофункціональним ресурсом є fakeimagedetector. Fakeimagedetector – онлайн-сервіс, призначений для виявлення підроблених зображень [30]. Сайт має привабливий дизайн і містить власний блог для користувачів (Рисунок 1.13).

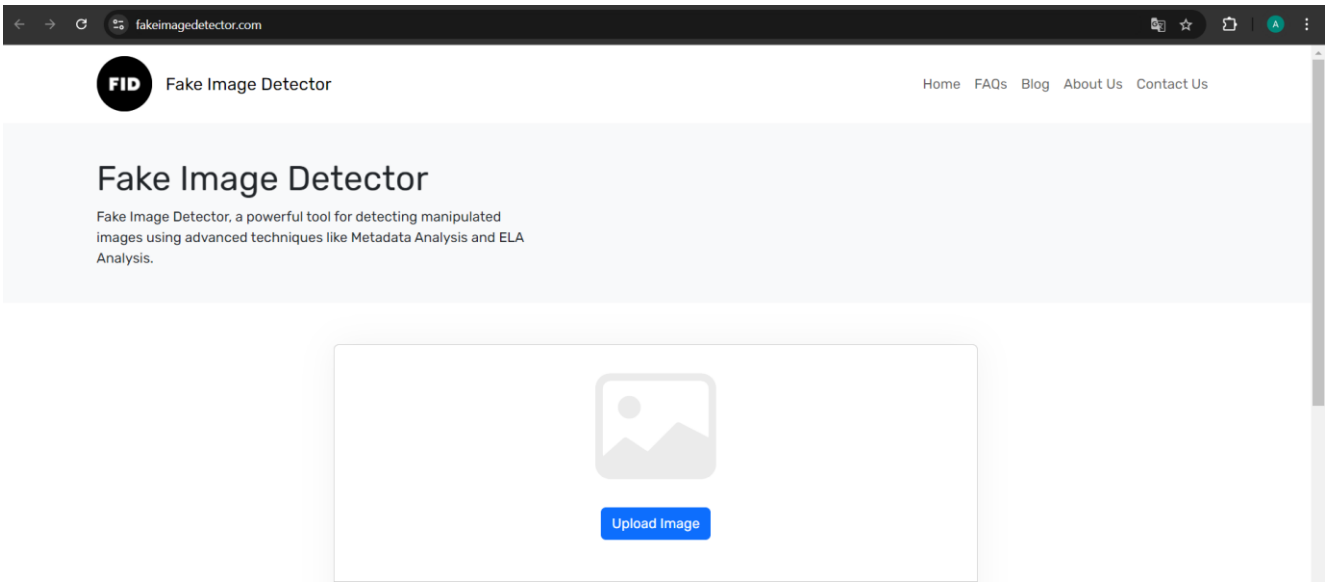


Рисунок 1.13 – Головна сторінка сайту Fakeimagedetector [30]

Fakeimagedetector здатний не лише надавати результати у текстовому форматі, а й у вигляді графіків, що відображають можливі змінені елементи зображення (Рисунок 1.14).

Після завершення тестування система вірно ідентифікувала три зображення з трьох, що є відмінним досягненням.

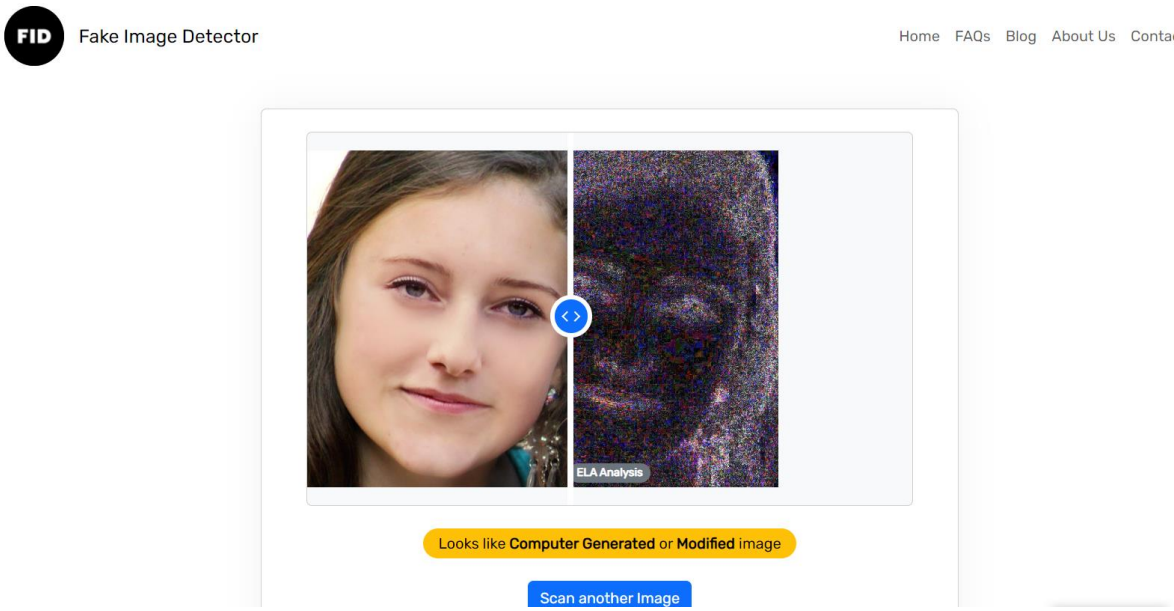


Рисунок 1.14 – Результат роботи сервісу Fakeimagedetector [30]

Таким чином, проаналізувавши вже існуючі програми, можна стверджувати, що деякі з них демонструють доволі високу точність, тоді як інші мають значніший рівень похибки. Однак, жоден з цих інструментів не забезпечує абсолютну точність і може давати помилки, тому розробка програмного забезпечення, яке дозволить покращити відсоток виявлення модифікацій облич є актуальною.

1.5 Постановка задачі

Мета кваліфікаційної роботи магістра полягає у підвищенні точності виявлення модифікованих фотографій облич людей. Для досягнення мети, визначені наступні задачі:

– провести аналіз вже готових рішень із класифікації модифікованих зображень за допомогою нейронної мережі;

– дослідити предметну область виявлення модифікованих зображень облич людей;

– знайти існуючі набори даних модифікованих фото та підготувати навчальний та тренувальний набір зображень для нейронної мережі;

– дослідити проблеми виявлення модифікованих зображень облич;

– проаналізувати готові рішення для виявлення та виокремлення обличчя на фото;

– розробити архітектуру нейронної мережі, дослідити та налаштувати гіперпараметри;

– обрати засоби розробки методу виявлення модифікованих зображень облич людей;

– розробити схему інформаційної системи виявлення модифікованих зображень облич;

– спроєктувати механізм обробки винятків та помилок роботи інформаційної системи;

– описати функціональні можливості розробленої інформаційної системи;

– провести тестування розробленого методу виявлення модифікованих зображень облич людей.

РОЗДІЛ 2 Метод виявлення модифікованих фотографій облич людей нейромережевими засобами

2.1 Підготовка зразків робочих даних для класифікатора

Для розробки методу виявлення модифікованих зображень облич людей, критично важливо мати структуровані вхідні дані. В глобальній мережі доступна досить велика кількість датасетів, проте для вирішення поставленої задачі немає такого датасету, який би одночасно міг бути використаним для навчання нейромережі для ідентифікації наявності модифікації, та для класифікації його виду.

Датасет – набір даних, який використовується для навчання та тестування моделей машинного навчання [31].

Під час розробки методу виявлення модифікованих зображень облич людей, у відкритому доступі, було знайдено декілька датасетів з яких було сформовано один, структурований набір даних із різними класами.

Зокрема для створення набору даних, були використані, вже готові, наступні датасети:

– FRLL-morphs (500 зображень);

– deepfake_faces (100 зображень);

– набір даних розроблений департаментом комп’ютерних наук університету Йонсей (400 зображень 300 із яких модифіковані та 100 без модифікацій).

Із датасету FRLL-morphs було створено 5 підкласів модифікованих зображень, кожен з яких містить по 100 фото (amsl, facemorpher, opnecv, stylegan2, webmorpher). Назви даних підкласів відповідають алгоритмам, за допомогою яких було модифіковане фото.

FRLL-morphs – набір структурованих даних, сформованих на основі даних взятих із Face Research London Lab [32]. Однією із ключових переваг даного датасету є те, що модифіковані зображення створювались різними методами (Рисунок 2.1).

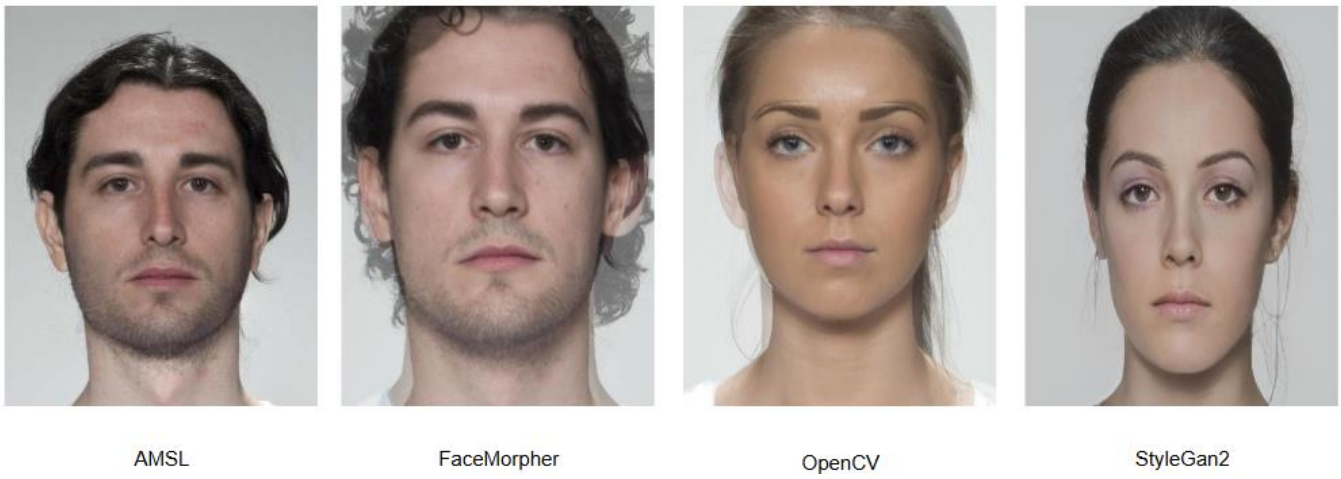


Рисунок 2.1 – Приклад модифікованих зображень різними алгоритмами

На Рисунку 2.1 видно, що алгоритми AMSL та StyleGan2 досить важко розпізнати модифікацію, а ПЗ FaceMorpher та алгоритм OpenCV мають характерне розмиття на фоні.

Deepfake_faces – набір даних, розмірністю 224 на 224 пікселі, що є перевагою, оскільки саме на таку розмірність існує досить багато навчених моделей [33]. На деяких зразках даного датасету присутні шуми (Рисунок 2.2).

Рисунок 2.2 – Приклад зображень датасету depfake_faces

Набір даних розроблений департаментом комп’ютерних наук університет Йонсей містить декілька рівнів модифікацій, від простіших – відразу помітних, до більш складних, модифікації на яких важко помітити (Рисунок 2.3).

Рисунок 2.3 – Приклад зображень, створених департаментом КН університету Йонсей

Із даного набору даних, у власний датасет додано по 100 модифікованих фото (підкласи easy, mid, hard), та обрано 100 зображень без модифікацій.

На основі вищеописаних датасетів, було створено власний датасет із різними та рівнями модифікацій (Рисунок 2.4).

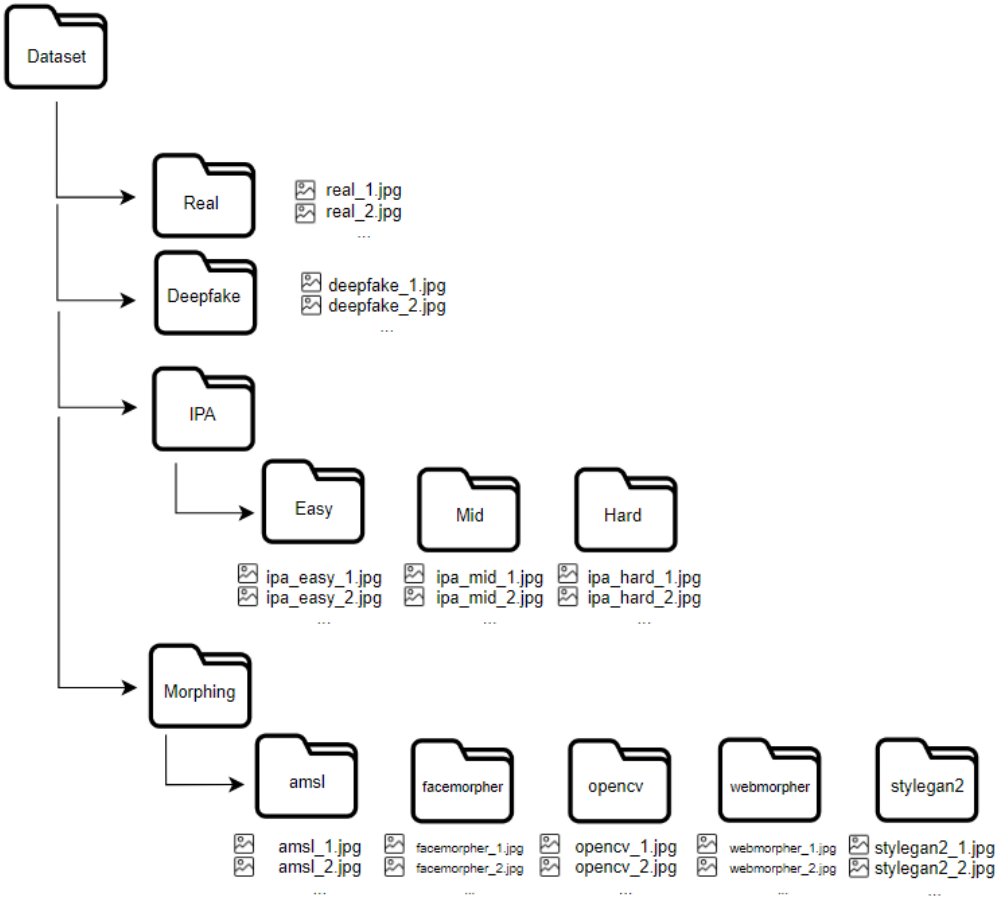


Рисунок 2.4 – Структура створеного датасету

Отже, в даному пункті було проаналізовано готові набори даних, та створено власний датасет. Створений набір даних містить наступні класи (підкласи):

– deepfake;

– real;

– ipa (easy,mid,hard);

– morphing (amsl, facemorfer, opencv, webmorpher, stylefan2).

Кожен із підкласів містить приблизно однакову кількість зображень різної розмірності. Зведення до одного розміру реалізується програмно.

2.2 Схема та кроки методу виявлення модифікованих фотографій облич людей

Метод виявлення модифікованих зображень облич людей призначений для перетворення вхідних даних у вигляді зображення, у вихідні дані у вигляді результату класифікації, а саме тип, складність та алгоритм за допомогою якого було модифіковане фото (Рисунок 2.5).

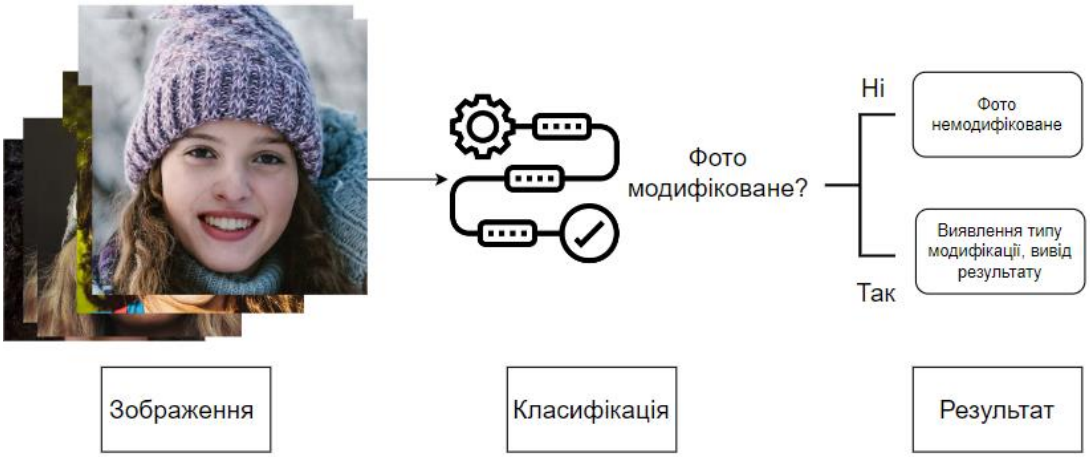


Рисунок 2.5 – Візуалізація роботи методу визначення модифікованих зображень

Для вхідних даних було створено датасет із приблизно 1000 зображеннями різних класів, а також тестова вибірка зображень для оцінки коректності (Рисунок 2.6).

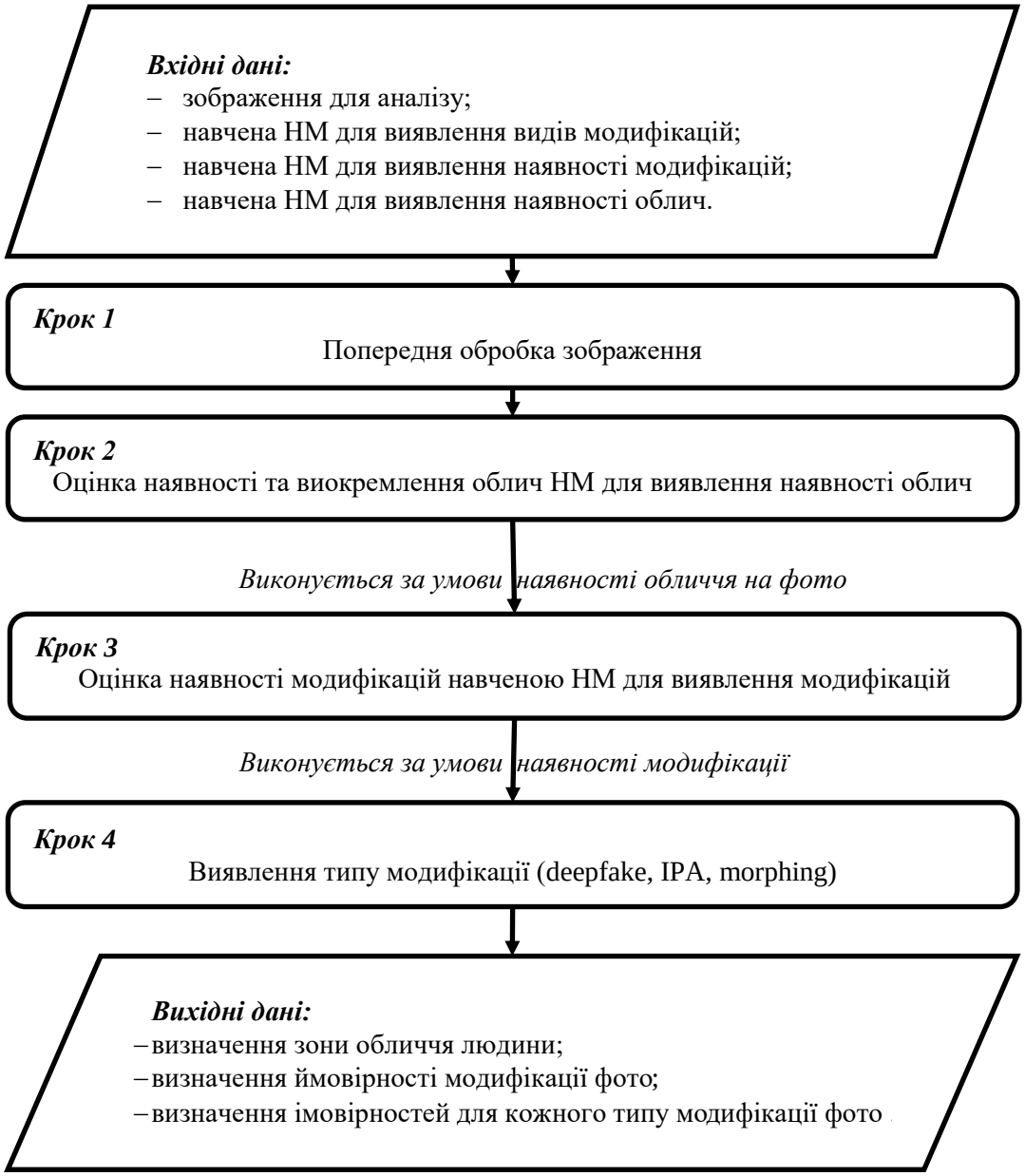


Рисунок 2.6 – Схема методу виявлення модифікованих зображень облич людей

Процес виявлення модифікованого зображення обличчя можна розділити на декілька етапів: попередня обробка зображення, оцінка виявлення та виокремлення обличчя на зображенні, оцінка наявності модифікації, виявлення типу модифікації.

Перший етап – попередня обробка зображень. На даному етапі фото, завантажене користувачем, проходить попередню обробку, а саме зміна розмірності до 224x224, нормалізація та конвертація у тензор.

Другий етап – виявлення та виокремлення облич на фото. На даному етапі відбувається перевірка, чи взагалі присутнє обличчя на зображенні і чи є сенс в подальшій перевірці на модифікації.

Третій етап – оцінка виявлення модифікації на фото. На даному етапі визначається чи піддавалось фото будь-яким модифікаціям.

Завершальний етап, виявлення типу модифікації, складності та алгоритму, за допомогою якого фото було модифіковане.

Вихідними даними методу виявлення модифікованих зображень облич, є результат у вигляді ймовірності належності зображення до конкретної модифікації.

Отже, було розроблено метод виявлення модифікованих зображень облич людей. Метод розроблений для перетворення вхідних даних у вигляді зображення для аналізу, навченої НМ для виявлення видів модифікацій, навчена НМ для виявлення наявності модифікацій та навчена НМ для виявлення наявності облич у вихідні дані – визначення зони обличчя людини, визначення ймовірності модифікації фото та визначення ймовірностей для кожного типу модифікації фото. Розроблений метод відрізняється від існуючих тим, що дозволяє виявляти не лише наявність модифікації обличчя, а і спосіб її походження.

2.3 Виявлення й виокремлення облич на фото за допомогою BlazeFace

Перш ніж виконувати процес виявлення типу модифікації, потрібно переконатись що на фото присутнє обличчя. В ході розробки методу виявлення модифікованих зображень облич людей було прийнято рішення використовувати готову модель – BlazeFace.

BlazeFace – потужна модель для розпізнавання обличчя на фото, розроблена GoogleResearch [34]. Дана модель використовує архітектуру, що базується на MibileNetV1, кориговану для задач розпізнавання обличчя. Спрощена схема роботи BlazeFace зображена на Рисунку 2.7

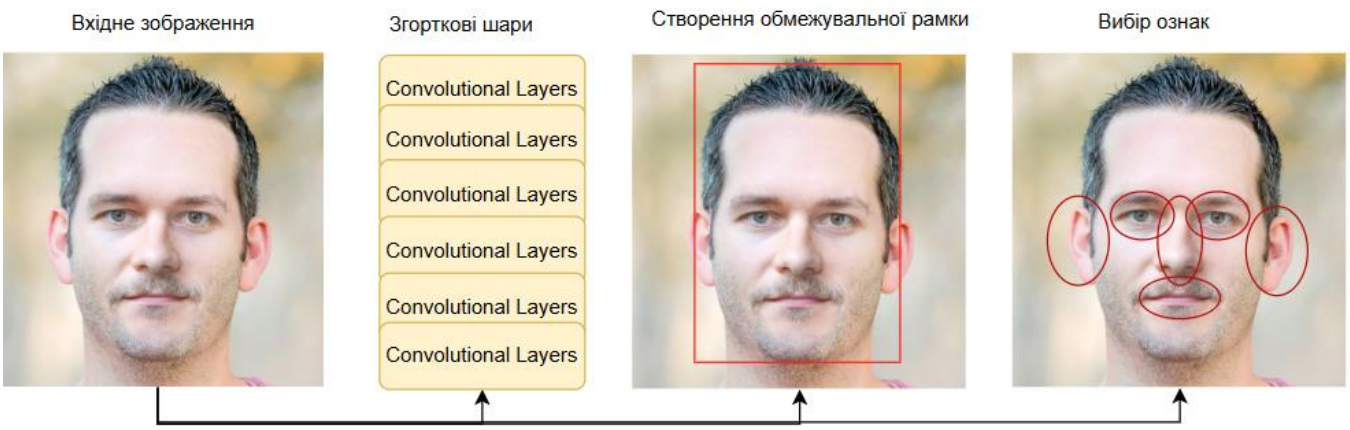


Рисунок 2.7 – Візуалізація роботи BlazeFace

На вхід дається зображення, після чого застосовуються згорткові шари, створюється обмежувальна рамка із координатами ymin, xmin, ymax, xmax, що описують кути рамки. Модель визначає ключові точки на обличчі, а саме: очі, ніч, вуха, та рот. Тобто модель визначає 12 точок, по 2 на кожну частину обличчя, після чого і відбувається процес визначення ймовірності наявності обличчя на фото.

В контексті методу для виявлення модифікованих зображень облич, blazeface виступає посередником, який не допускає класифікації зображення на якому відсутнє обличчя (Рисунок 2.8).

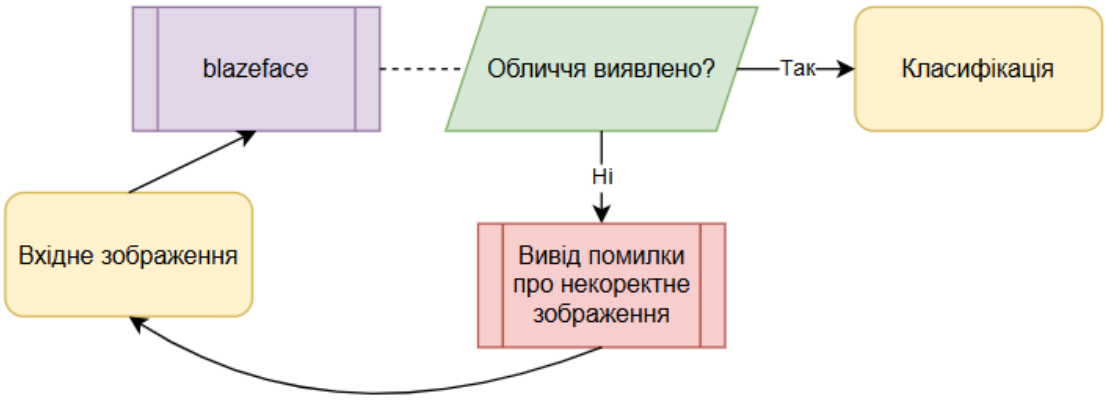


Рисунок 2.8 – Схема роботи blazeface

Модель ніяк не впливає на завантажене зображення, тобто розмірність та якість, після детекції обличчя, зміненні не будуть.

Як результат, blazeface повертає масив, де кожен елемент містить координати певної ключової точки(око, ніс, тощо). Якщо на фото буде виявлено більше ніж одне обличчя, на виході буде отримано список об’єктів із координатами. В випадку якщо обличчя не виявлено – модель вертає пустий масив.

Отже, BlazeFace досить добре підходить для розпізнавання обличчя на фото і буде використано елементом вхідних даних методу виявлення модифікованих зображень облич людей, оскільки модель досить проста в використанні та не вимагає додаткових навичок та дій від розробника.

2.4 Навчання нейронної мережі

Навчання нейронної мережі – невід’ємний етап в розробці методу для виявлення модифікованих зображень облич, адже від даного етапу залежать наскільки точно модель зможе класифікувати зображення.

Процес навчання нейроних мереж зображено на Рисунку 2.9.

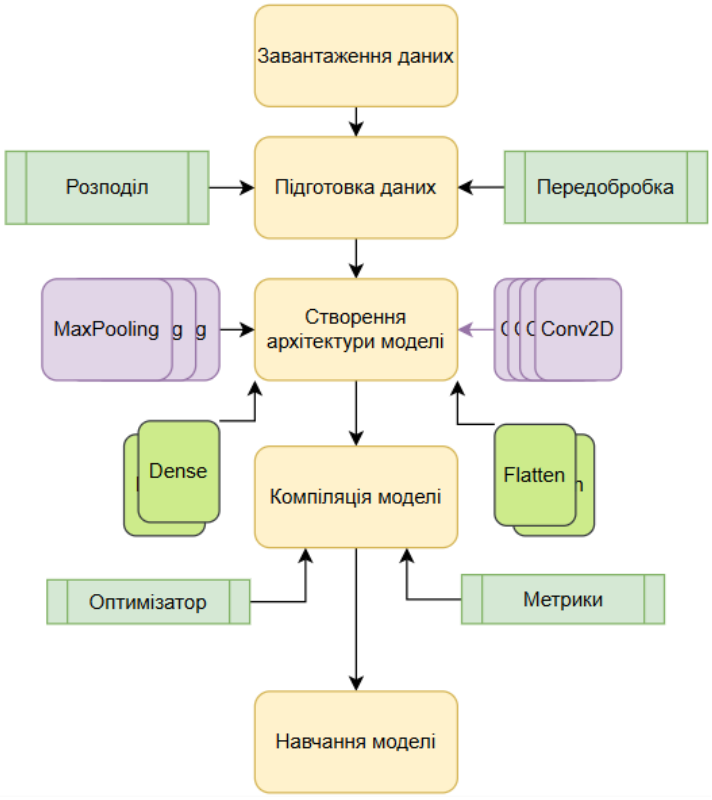


Рисунок 2.9 – Схема навчання згорткової нейронної моделі

Перший крок – завантаження та підготовка вхідних даних. На даному етапі завантажуються зображення та розподіляються відповідні мітки. Дані підлягають обробці та нормалізації.

Наступним етапом є створення архітектури нейромережі, тобто визначається різноманітні шари (згортки, пулінгу тощо), та яка кількість буде додана до моделі.

На етапі компіляції обирається оптимізатор, метрики та функція втрат які будуть використані у моделі.

Завершальним є етап навчання моделі. Застосовуються всі параметри, функції оптимізації та активації та відбувається процес навчання.

Отже, в даному розділі було розглянуто етапи навчання нейромережевих моделей. Дотримання даних кроків допоможе ефективно навчити нейромережеву модель для використання у розробці методу виявлення модифікованих зображень облич людей.

2.5 Розробка архітектури нейронної мережі для виявлення модифікованих зображень облич людей

2.5.1 Використання бібліотеки TensorFlow для виявлення модифікованих облич людей

TensorFlow – бібліотека для машинного навчання розроблена компанією Google [35]. Дана бібліотека досить добре підходить для побудови та навчання згорткових нейроних мереж та надає численні переваги у роботі із зображеннями.

TensorFlow має велику кількість готових модулів для роботи із зображеннями та нейромережевими моделями в цілому:

– tf.keras.preprocesing.image – модуль для попередньої обробки зображень, включаючи аугментацію, що є досить корисною функцією, адже це дозволить збільшити кількість вхідних даних;

– tf.keras.applications – модуль, що надає попередньо навчені моделі та готові архітектури нейроних мереж;

– tf.keras.layers – містить шари для побудови нейроних мереж (згорткові, пулінгові тощо);

– tf.io – набір функцій, для завантаження та збереження зображень.

Бібліотек tensorflow, містить вже готові функції, які спростять процес навчання згорткової нейронної мережі (Рисунок 2.10).

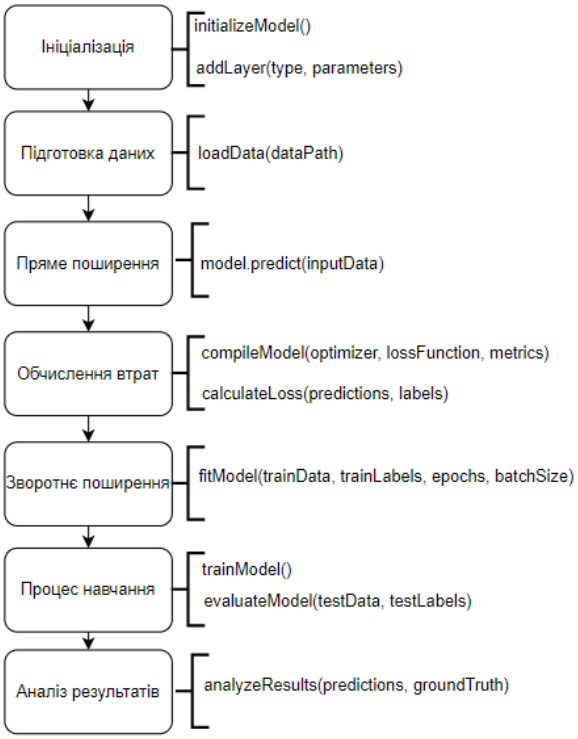


Рисунок 2.10 – Функції tensorflow для навчання НМ

Список функцій, та їх значення:

– initializeModel – ініціалізує модель та створює базову структуру для додання шарів в майбутньому:

– addLayer – додавання шару до моделі, приймає параметр тип(згортковий, пулінг, тощо);

– loadData – виконує функцію завантаження даних із вказаного шляху;

– model.predict – виконує процес передбачення, на основі вхідних даних;

– compileModel – компілює модель, задаючи функцію втрат, оптимізатор та метрики;

– calcualteLoss – обчислює втрати моделі;

– fitModel – навчає модель, на попередньо підготовлених тренувальних даних;

– trainModel – запускає повний процес навчання із вказаними даними;

– evaluateModel – оцінює продуктивність моделі на тестовій вибірці даних;

– analyzeResults – аналізує результати моделі.

Отже, бібліотека tensorflow містить велику кількість готових рішень для роботи із зображеннями та побудови архітектури нейромереж в цілому, що надає певні переваги при розробці методу для виявлення модифікованих облич людей.

2.5.2 Нормалізація та аугментація вхідних даних

Важливим етапом в ході розробки методу виявлення модифікованих зображень облич людей є підготовка вхідних даних. Для коректного виявлення типу модифікацій потрібно зібрати, нормалізувати дані.

Нормалізація даних – процес масштабування числових значень у набір даних до певного діапазону значень [36]. Нормалізація вхідних даних перед навчанням не потрібно упускати, оскільки це допоможе скоротити час і обчислювальні ресурси при навчанні нейромережевої моделі, а також підвищить точність розпізнавання (Рисунок 2.11).

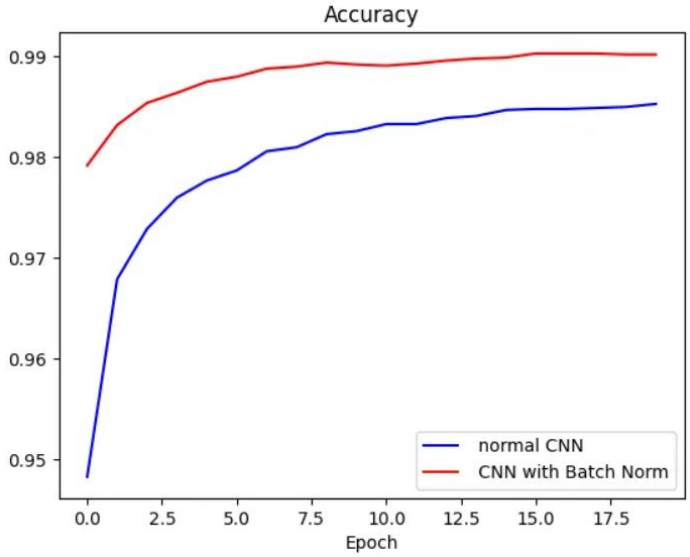


Рисунок 2.11 – Точність роботи CNN з нормалізацією даних та без [36]

TensorFlow має декілька зручних методів та функцій для реалізації нормалізації, що допомагають спростити та покращити процес навчання нейронної мережі:

– tf.keras.layers.BatchNormalization – функція, що нормалізую вхідні дані в рамках одного пакету(batch);

– tf.keras.layers.LayerNormalization – метод, що проводить нормалізацію по вектору, у кожному окремому шару незалежно від партії;

– tensorflow_addons – нормалізує вхідні дані в межах одного каналу.

Аугментація – збільшення розмірності вхідних даних за допомогою різних трансформацій [37]. Даний підхід дозволяє значно збільшити кількість даних адже із одного зразка зображення можна отримати значно більшу кількість (Рисунок 2.12).

Рисунок 2.12 – Візуалізація аугментації вхідних даних

В даному розділі було розглянуто важливість нормалізації та аугментації вхідних даних при розробці методу виявлення модифікованих зображень облич людей. Також було описано функції, які надає біліотека tensorflow, для реалізації даних методів.

2.5.3 Налаштування гіперпараметрів моделі

Налаштування гіперпараметрів критично важливий етап під час навчання згорткової нейронної мережі, оскільки правильно підібрані параметри напряму впливають на працездатність моделі.

Кількість шарів у згортковій нейронній мережі є одним із найважливіших гіперпараметрів, оскільки від їх числа залежить наскільки складні ознаки та закономірності зможе засвоїти модель (Рисунок 2.13).

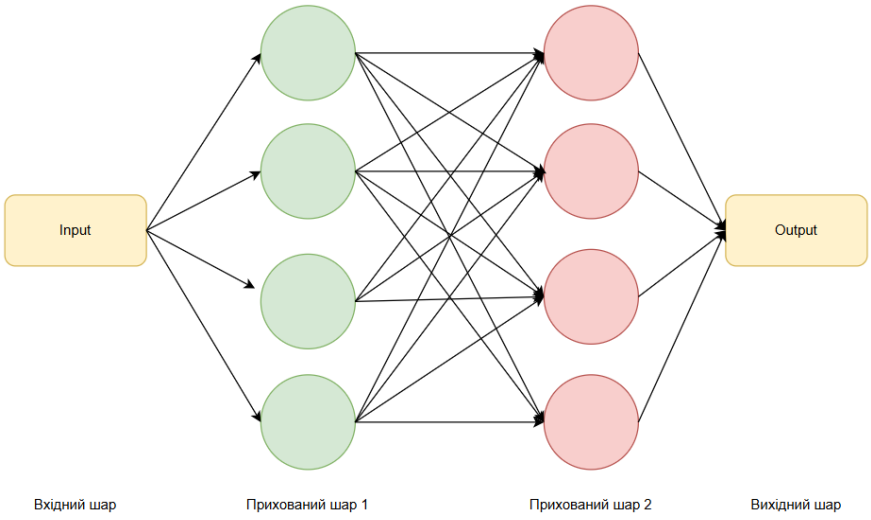


Рисунок 2.13 – Базова архітектура нейромережі із двома прихованими шарами

Розмір фільтра – важливий параметр, який визначає розмір отриманої інформації із вхідного зображення [38] (Рисунок 2.14).

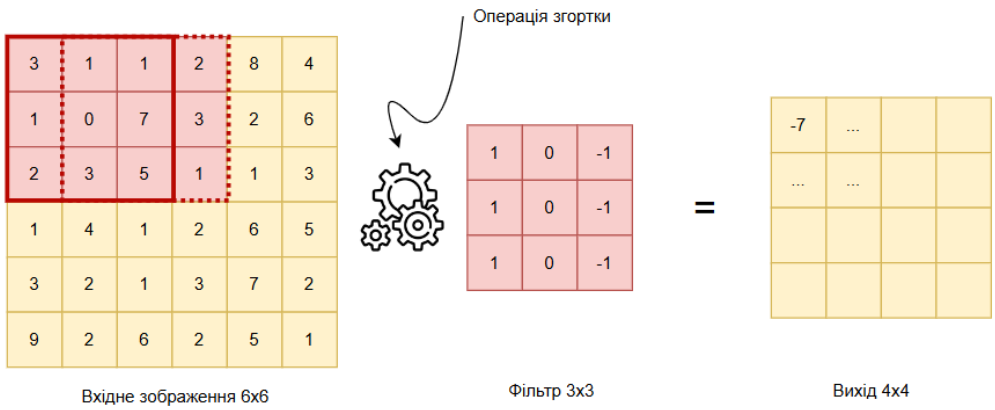


Рисунок 2.14 – Візуалізація роботи фільтра в НМ

Більший розмір фільтра зможе отримати більше інформації із вхідного зображення, проте якщо вказати занадто велике значення це може призвести до погіршення продуктивності моделі та втрати локальних ознак.

Крок фільтра – це гіперпараметр, який визначає на яку кількість пікселів буде зміщатись фільтр по вхідному зображені (Рисунок 2.15).

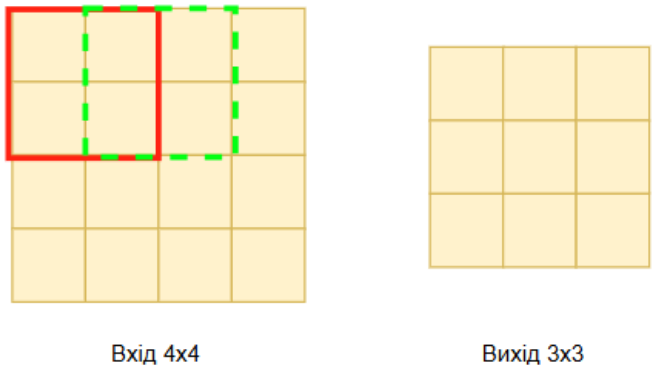


Рисунок 2.15 – Візуалізація кроку фільтра

Більше значення кроку фільтра може зменшити мапу ознак, що зазвичай приводить до втрати інформації. Менший значення кроку зберігає більшу кількість ознак, проте збільшує час обчислення.

Відступ (padding) – техніка, що дозволяє зберегти просторові розміри зображення, під час використання загорткових шарів. Даний параметр передбачає додавання додаткових нулів навколо вхідного зображення [39] (Рисунок 2.16).

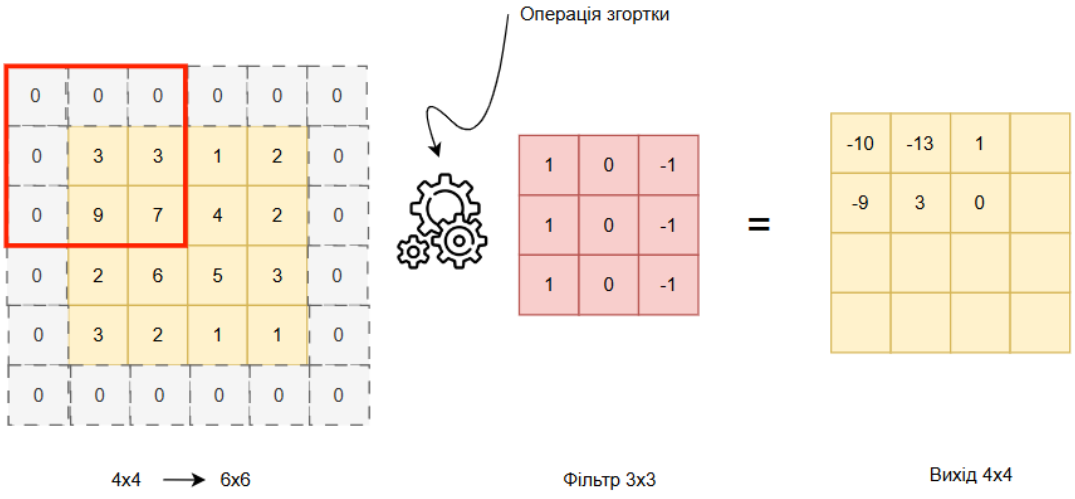


Рисунок 2.16 – Візуалізація відступу в НМ

Швидкість кроку – важливий параметр, який вказує розмір оновлення ваг під час навчання. Якщо вказати значення занадто велике, навчання буде нестабільним, занадто низька швидкість кроку призведе до втрати якості нейромережевої моделі (Рисунок 2.17).

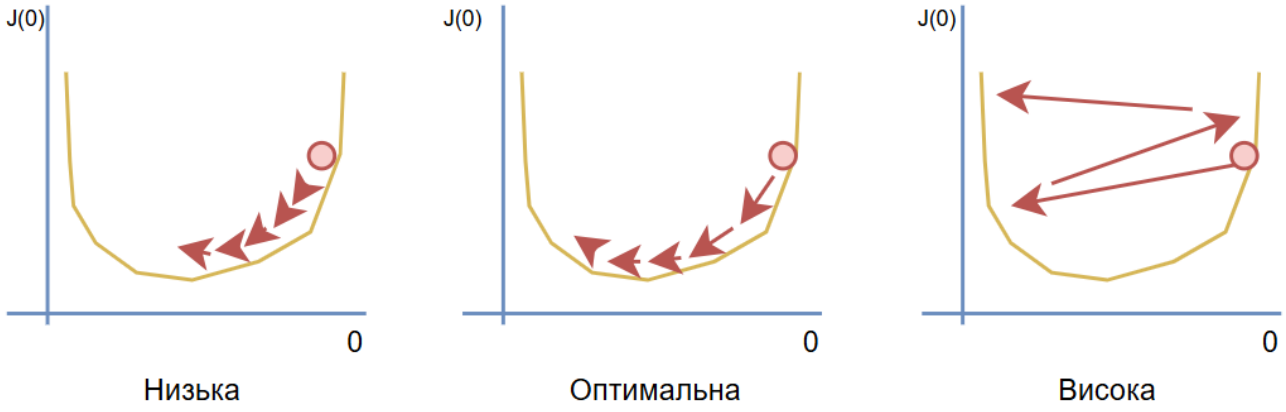


Рисунок 2.17 – Візуалізація швидкості оновлення ваг в НМ

Розмір пакету(batch size) – параметр, який визначає кількість зразків вхідних даних, які обробляються мережею за одну ітерацію. Більший розмір даного параметру покращує якість навчання, проте знижує швидкість. Менший розмір зменшує навантаження на пам’ять та збільшує швидкість навчання [40].

Якщо при побудові нейронної мережі вказати занадто велике значення розміру пакета то відбудеться перенавчання, занадто мале значення даного параметра призведе до недонавчання мережі (Рисунок 2.18).

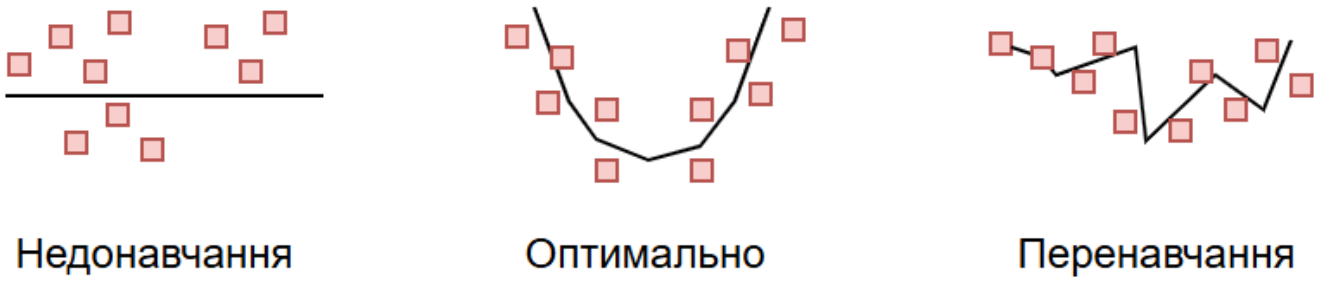


Рисунок 2.18 – Вплив розміру пакета на стабільність навчання

Для розробки методу виявлення модифікованих зображень облич людей обрано наступні гіперпараметри:

– batch_size – 16, невеликий розмір пакету дозволяє отримати баланс між швидкістю та навчанням;

– epochs – 10, оскільки невеликий розмір даного параметру зменшує навантаження при навчанні та дозволяє уникнути перенавчання;

– оптимізатор – adam, забезпечить швидке та стабільне навчання моделі;

– розмірність фільтрів – 32, типовий вибір для згорткових шарів, який забезпечить виділення основних ознак;

– розмір ядра – 3x3, забезпечить отримання локальних ознак із областей зображення.

Крок фільтра(1) та швидкість оновлення ваг(0.001) залишено за замовчуванням. Відступи не задаються. Однак, вплив гіперпараметрів ще буде досліджено окремо.

Отже, в даному розділі було розглянуто та обрано значення найважливіших гіперпараметрів для згорткової нейронної мережі. Під час розробки методу виявлення модифікованих зображень облич, потрібно правильно налаштувати вищеописані параметри, оскільки від цього буде залежати швидкість, точність, функціональність даного методу.

2.5.4 Архітектура нейронної мережі для ідентифікації модифікації

Правильно побудована архітектура є важливим етапом у розробці методу виявлення модифікованих зображень, оскільки структура архітектури впливає на швидкість, точність та обчислювальну ефективність моделі (Рисунок 2.19).

Вхідний шар приймає вхідні дані у вигляді зображення розмірністю 224x224x3, для можливості роботи із кольоровими зображеннями.

Згорткові шари містять 64 фільтри розміром 3x3, та функцію активації ReLU. Після операції згортки, на вихід отримується тензор розмірністю 224x224x64, оскільки кількість фільтрів замінює кількість каналів. Функція активації ReLU додає нелінійність у модель, що допомагає прибрати негативні значення.

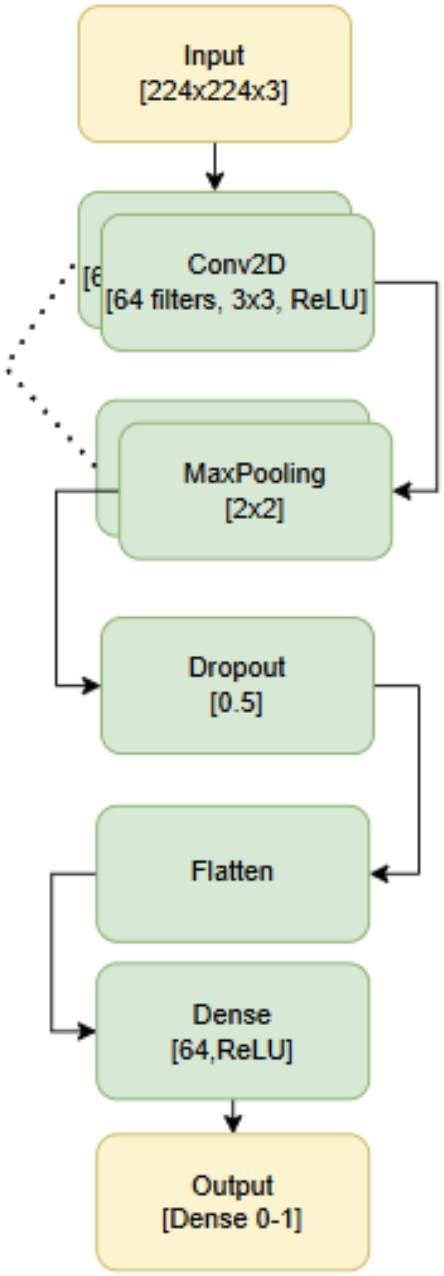


Рисунок 2.19 – Архітектура нейромережі для виявлення наявності модифікації

Шари пулінгу отримують на вхід тензор розмірністю 224x224x64, після чого відбувається зменшення розмірності вдвічі, завдяки ядру 2x2. Тобто на вихід передається тензор розмірністю 112x112x64.

Шар dropout виконує функції випадкового відключення, щоб запобігти перенавчанню моделі. Тобто на даному етапі відбувається відключення 50% нейронів (обрано випадково).

Шар flatten перетворю тривимірний тензор у одновимірний вектор, оскільки вихідний шар dense, приймають на вхід лише одновимірні вектори.

Вихідний шар dense, повертає один нейрон 0-1, для отримання результату класифікації, тобто даний шар визначає чи модифіковане зображення.

Отже, дана архітектура призначена для виявлення присутності модифікації на зображені.

2.5.5 Архітектура нейронної мережі для класифікації виду модифікації

Архітектура нейромережевої моделі для виявлення типу модифікації зображена на Рисунку 2.20.

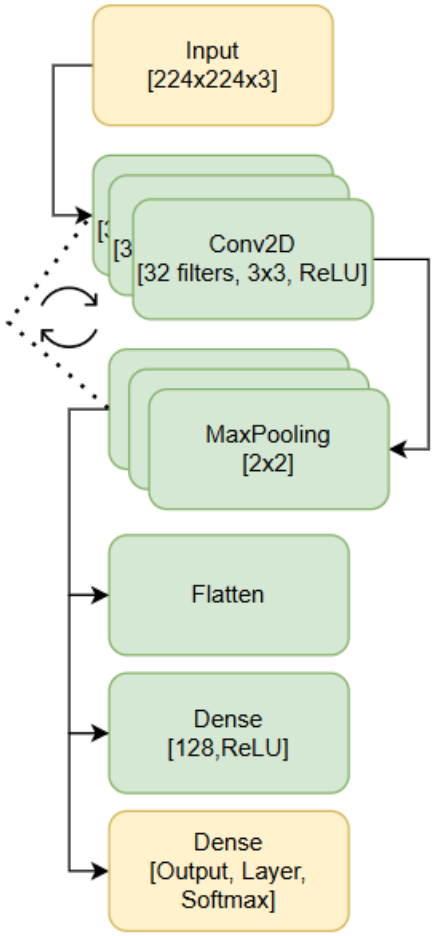


Рисунок 2.20 – Архітектура нейромережі для виявлення типу модифікації

Вхідний шар (Input) отримує зображення розміром 224x224 пікселі, із трьома каналами, оскільки зображення кольорові.

Згорткові шари (Conv2D), містять 32 фільтри розмірністю 3x3, що обробляють зображення. За допомогою даних шарів, модель навчається виявляти базові ознаки(краї, текстури).

Шари пулінгу (MaxPooling), використовують філбтри 2x2, які зменшують розмірність вхідного зображення вдвічі, отримуючи найважливіші ознаки, що знижує обчислювальну складність нейромережевої моделі.

Наступний шар перетворює багатомірний масив, отриманий із попередніх шарів, у одновимірний вектор даних, для того щоб підготувати дані для обробки у повнозв'язних шарах.

Повнозв'язний шар (Dense), обробляє отриманий вектор, що дає змогу виявляти складніші ознаки. Шар містить 128 нейронів, кожен із яких запускає функцію активаці.

Вихідний шар (Output), створює ймовірності для кожного із можливих класів модифікацій зображення та повертає той, де сума ймовірностей найбільша. Вихідний шар містить 9 виході відповідно до кожного класу, а саме deepfake, ipa (easy), ipa (mid), ipa (hard), morphing (amsl), morphing (facemorfer), morphing (opencv), morphing (webmorpher), morphing (stylefan2).

Функція оптимізації – метод або алгоритм, що використовується для оптимізації процесу навчання нейронних мереж [41].

Також слід зазначити, що при розробці методу виявлення модифікованих зображень облич людей було обрано алгоритм оптимізації «Adam».

Отже, в даному розділі були розглянуті ключові моменти, які слід врахувати на етапі навчання нейромережевої моделі та побудовано архітектуру нейромережі для методу виявлення модифікованих зображень облич людей. Дана архітектура нейромережі призначена для класифікації виду модифікації для зображень із наявними модифікаціями.

2.6 Оцінка продуктивності нейромережевих моделей

Під час розробки методу виявлення модифікованих зображень облич людей, важливим етапом є оцінка продуктивності побудованих нейромережевих моделей.

Метрики дозволяють визначити наскільки добре НМ виконують поставлене завдання. До найбільш значущих можна віднести наступні:

– accuracy;

– precision;

– recall;

– F1-міра.

Оскільки для розробки методу виявлення модифікованих зобрежнь облич, використовується НМ для виявлення типу модифікації та НМ для ідентифікації модифікації формули метрик буду відрізнятись.

Формули метрик для нейронної мережі ідентифікації модифікації описані нижче.

Accuracy розраховує відсоткове співвідношення правильно класифікованих фото серед усіх наявних. Дана метрика розраховується наступною формулою:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2.1)$$

де TP – кількість правильно класифікованих модифікованих фото, TN – кількість правильно класифікованих фото без модифікацій, FP – реальні обличчя помилково класифіковані як фото з модифікацією, FN – модифіковані фото, хибно класифіковані як реальні [42].

Precision – метрика, яка вимірює частку коректно класифікованих модифікованих облич, серед всіх, які були класифіковані як модифіковані. Precision розраховується наступною формулою:

$$Precision = \frac{TP}{TP + FP}, \quad (2.2)$$

де TP – кількість правильно класифікованих модифікованих фото, FP – реальні обличчя помилково класифіковані як фото з модифікацією [43].

Recall – показує здатність моделі виявляти модифіковані фото обличчя, розраховується наступною формулою:

$$Recall = \frac{TP}{TP + FN}, \tag{2.3}$$

де *TP* – кількість правильно класифікованих модифікованих фото, *FN* – модифіковані фото, хибно класифіковані як реальні [41].

F1-міра – розраховує середнє між precision та recall, розраховується наступною формулою:

$$F1 = \frac{Precision \cdot Recall}{Precision + Recall}, \tag{2.4}$$

де *Precision* та *Recall*, метрики, які були розраховані формулами 2.2 та 2.3 [42].

Формули розрахунку метрик нейронної мережі виявлення типу модифікацій описані нижче.

Accuracy визначається наступною формулою:

$$Accuracy = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FP_i + FN_i)}, \tag{2.5}$$

де *TP(i)* – кількість правильно класифікованих модифікованих фото,як конкретний клас модифікації, *FP(i)* – кількість зображень, які хибно класифіковані як певний тип модифікацій, *FN(i)* – кількість зображень, які класифіковані хибно як певний тип модифікацій, хоча вони належать до цього класу.

Precision визначається наступною формулою:

$$Precision_i = \frac{TP_i}{TP_i + FP_i}, \tag{2.6}$$

де *TP(i)* – кількість правильно класифікованих зображень, *FP(i)* – кількість зображень, які хибно класифіковані як певний тип модифікацій.

Метрика Recall розраховується наступною формулою:

$$Recall_i = \frac{TP_i}{TP_i + FN_i}, \tag{2.7}$$

де TP_i – кількість правильно класифікованих фото, , $FN(i)$ – кількість зображень, які класифіковані невірно, як певний тип модифікацій, хоча вони належать до цього класу.

$$F1_i = 2\frac{TP_i}{TP_i + FN_i}, \tag{2.3}$$

де TP_i – кількість правильно класифікованих зразків фото, $FN(i)$ – кількість зображень, які класифіковані хибно як певний тип модифікацій, хоча вони належать до цього класу.

Різниця в обчисленнях даних метрик, полягає у кількості класів. У нейронній мережі визначення наявності модифікацій, їх лише два (реальне фото, модифіковане фото). У НМ для визначення типу модифікації, метрики обчислюються для кожного класу окермо.

Отже, в даному розділі розглянуто та описано метрики, які будуть використовуватись для оцінки точності нейромереж в рамках методу виявлення модифікованих зображень облич людей, а саме: accuracy, precision, recall, f1.

Висновки до розділу 2

Отже, в даному розділі було проаналізовано готові вибірки даних, знайдені у відкритому доступі та створено власний датасет для розробки методу виявлення модифікованих зображень облич людей. Створений набір даних містить класи та мітки класів, а саме:

– deepfake;

– iPA;

– morphing.

Кожен клас має особливості, такі як різні алгоритми створення модифікацій, чи різна складність для виявлення модифікованих облич.

Розглянуто важливість нормалізації та аугментації, їх переваги, та недоліки. Дані підходи були застосовані до створеного датасету, що значно розширює та покращує створений набір даних.

Було проведено ознайомлення із бібліотекою BlazeFace, яка допоможе у розробці методу для виявлення модифікацій на етапі виокремлення обличчя на зображені.

В ході проектування архітектури нейромережевої моделі, використано бібліотеку tensorflow, виділено переваги та недоліки на етапі створення згорткової нейронної мережі.

В процесі ознайомлення із tensorflow досліджено важливість гіперпараметрів та їх вплив на точність нейронної мережі.

Створено власну архітектуру згорткової нейронної мережі для розробки методу виявлення модифікованих облич людей.

Обрано до використання метрики, які будуть використовуватись для оцінки ефективності нейромереж в рамках методу виявлення модифікованих зображень облич людей, а саме: accuracy, precision, recall, f1.

РОЗДІЛ 3 Проєктування інформаційної системи виявлення модифікованих зображень облич людей

3.1 Визначення комбінації засобів розробки інформаційної системи

Інтерфейс є невід'ємною частиною будь-якого програмного забезпечення. В ході розробки інтерфейсу для методу виявлення модифікованих зображень облич людей було використано HTML для каркасу сторінки та CSS для стилізації всіх елементів. HTML – мова розмітки, що використовується для побудови каркасу веб-сторінок за допомогою тегів [46]. Дана мова розмітки, є базовою для створення структури веб-додатків.

За допомогою HTML у інформаційній системі створено поля для вхідних даних, кнопки управління, та поля для виводу результатів. CSS – мова стилів, яка задає стилізацію елементам HTML [47]. CSS надає змогу створити різного роду анімації та зробити інтерфейс візуально привабливим. До прикладу, на Рисунку 3.1, зображено частину інтерфейсу користувача із використанням CSS та без.

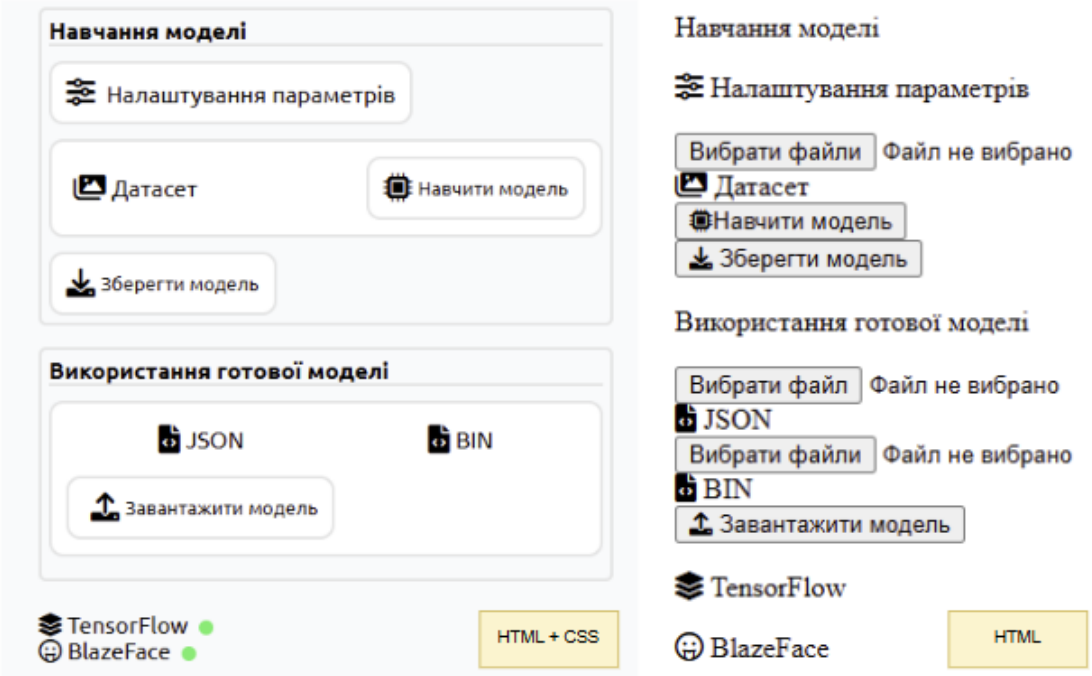


Рисунок 3.1 – Порівняння інтерфейсів із використанням CSS та без

Для реалізації основного методу, використана мова програмування javascript та. Javascript є основною мовою програмування, яка створює інтерактивність та функціональність методу виявлення модифікованих зображень облич людей.

У якості бібліотеки машинного навчання обрано TensorFlow. Дана бібліотека дозволяє запускати нейронні мережі безпосередньо у браузерів, що є вагомою перевагою.

Для попереднього виявлення та виокремлення обличчя використана попередньо навчена модель – BlazeFace. Дана модель має високу швидкість роботи, компактний розмір та високу точність виявлення обличчя на зображені. Також вагомою перевагою є гнучка інтеграція із бібліотекою TensorFlow, що дозволяє реалізувати функціонал виявлення та виокремлення облич досить просто.

В якості середовища розробки коду, використовувався Visual Studio Code, оскільки він безкоштовний, має велику кількість програмних розширень, а також не використовує багато ресурсів ПК [48] (Рисунок 3.2).

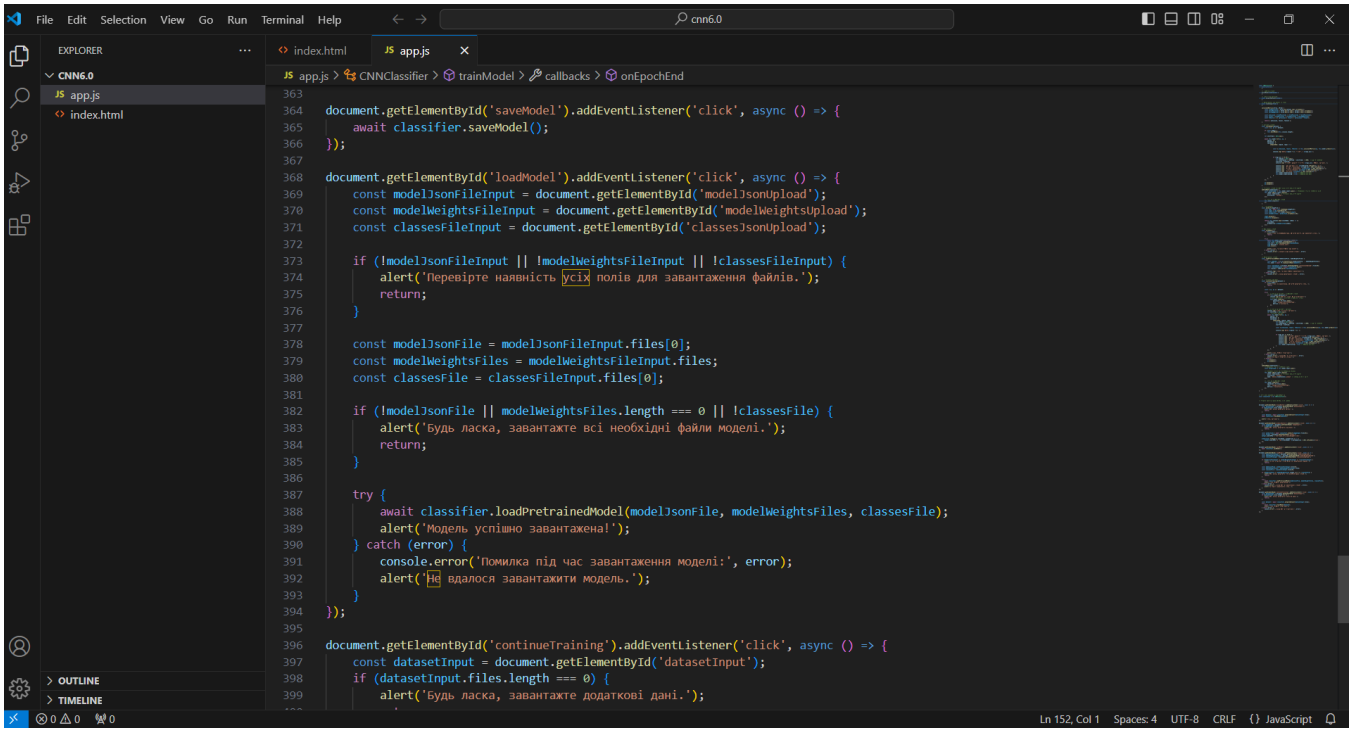


Рисунок 3.2 – Інтерфейс Visual Studio code

Отже, поєднання всіх вищеописаних технологій надає змогу розробити метод виявлення модифікованих зображень облич людей, а також побудувати та стилізувати візуально привабливий та зрозумілий інтерфейс.

3.2 Проєктування модулів інформаційної системи

Щоб описати принцип функціонування інформаційної системи виявлення модифікованих зображень облич людей, створено структурну схему (Рисунок 3.3).

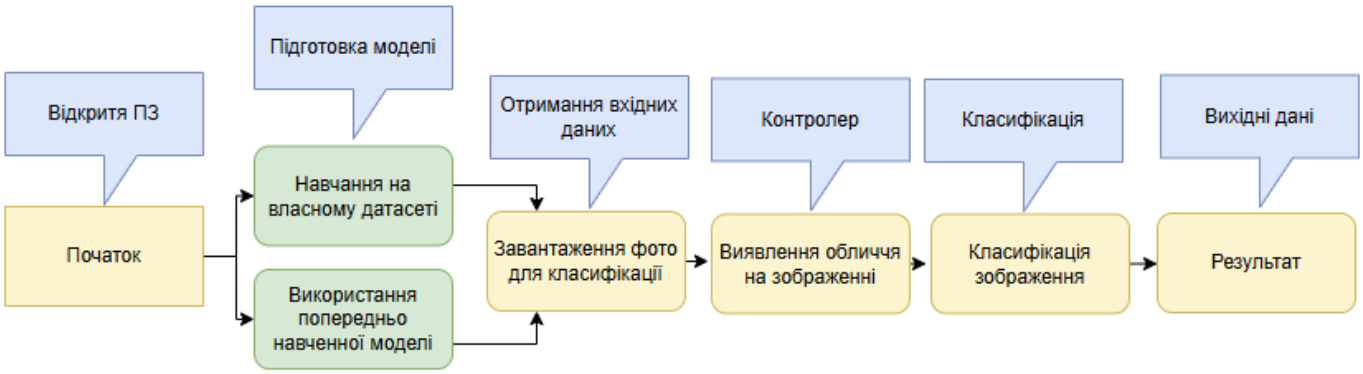


Рисунок 3.3 – Схема функціонування інформаційної системи

Дана схема містить етапи від відкриття до отримання результатів класифікації, не враховуючи валідацію вхідних даних та статус підключення зовнішніх бібліотек.

Загальний принцип наступний, користувач відкриває інформаційну систему, після чого навчає модель на власному датасеті, чи завантажує попередньо навчену модель. Після успішної підготовки моделі – завантажується фото для класифікації. Перед запуском функції виявлення модифікації, посередник (blazeface), перевіряє наявність обличчя на фото. У разі успішного виявлення обличчя – запускається процес класифікації та вивід результатів. Якщо ж обличчя не виявлено, користувачу пропонується завантажити інше зображення.

Отже, у даному розділі розглянуто схему інформаційної системи виявлення модифікованих зображень облич людей та поетапно описані дії користувача від відкриття ПЗ до отримання результату.

3.3 Проєктування компонентної схеми інформаційної системи

Під час розробки методу виявлення модифікованих зображень облич людей, було програмно реалізовано декілька компонентів, кожен із яких має свої функції (Рисунок 3.4).

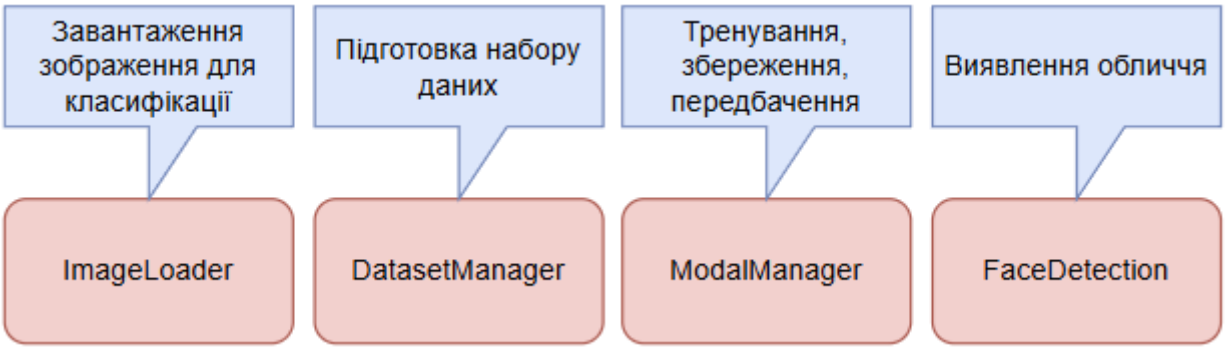


Рисунок 3.4 – Список компонентів ІС

В цілому, ланцюжок послідовних дій, від завантаження датасету до отримання результатів класифікації зображено на Рисунку 3.5.

На даній схемі зображений варіант із завантаженням датасету, проте якщо користувач вирішить не навчати модель самостійно, а завантажити попередньо навчену, тоді етап із навчанням пропускається.

Початок процесу відбувається після того як користувач завантажує вхідний набір даних – датсет. Після чого відбувається перетворення зображень у тензор.

Наступним етапом є навчання моделі, куди подаються тензори зображень та гіперпараметри за замовчуванням або ж ті, які вказав користувач. В якості вихідних даних отримана навчена нейромережева модель.

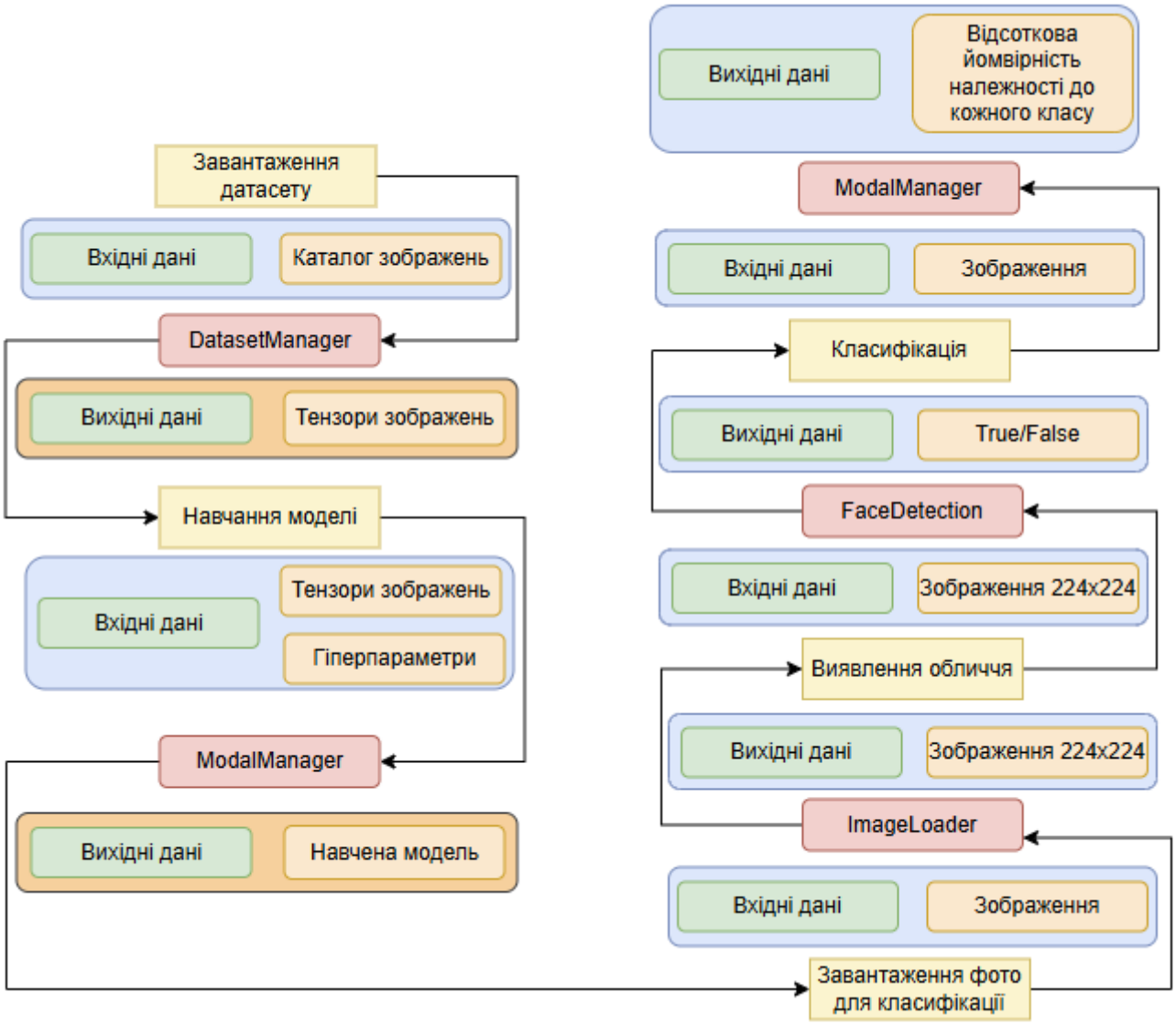


Рисунок 3.5 – Взаємодія компонентів інформаційної системи

Після того, як модель готова до класифікації, користувач завантажує зображення, після чого відбувається зміна розміру фото (якщо потрібно) до 224x224.

Після приведення до потрібного розміру, виконується посередник (blazeface) для виявлення та виокремлення обличчя на зображені. Якщо результат позитивний (зображення містить обличчя) запускається процес класифікації та виведення результатів у вигляді відсоткової ймовірності належності до всіх класів. Якщо ж обличчя не виявлено користувачу отримує сповіщення, що обличчя не знайдено та прохання завантажити друге фото.

Отже було спроєктовано схему компонентів та зображено послідовність їх виклику при використанні інформаційної системи виявлення модифікованих зображень облич людей.

3.4 Проєктування механізму обробки помилок та винятків

Для коректної роботи методу для виявлення модифікованих зображень облич, перед початком класифікації, потрібно передбачити, на якому етапі може виникнути помилка. Для уникнення непередбачуваних помилок потрібно реалізувати декілька функцій для перевірки коректності введення та завантаження даних користувачем (Рисунок 3.6).

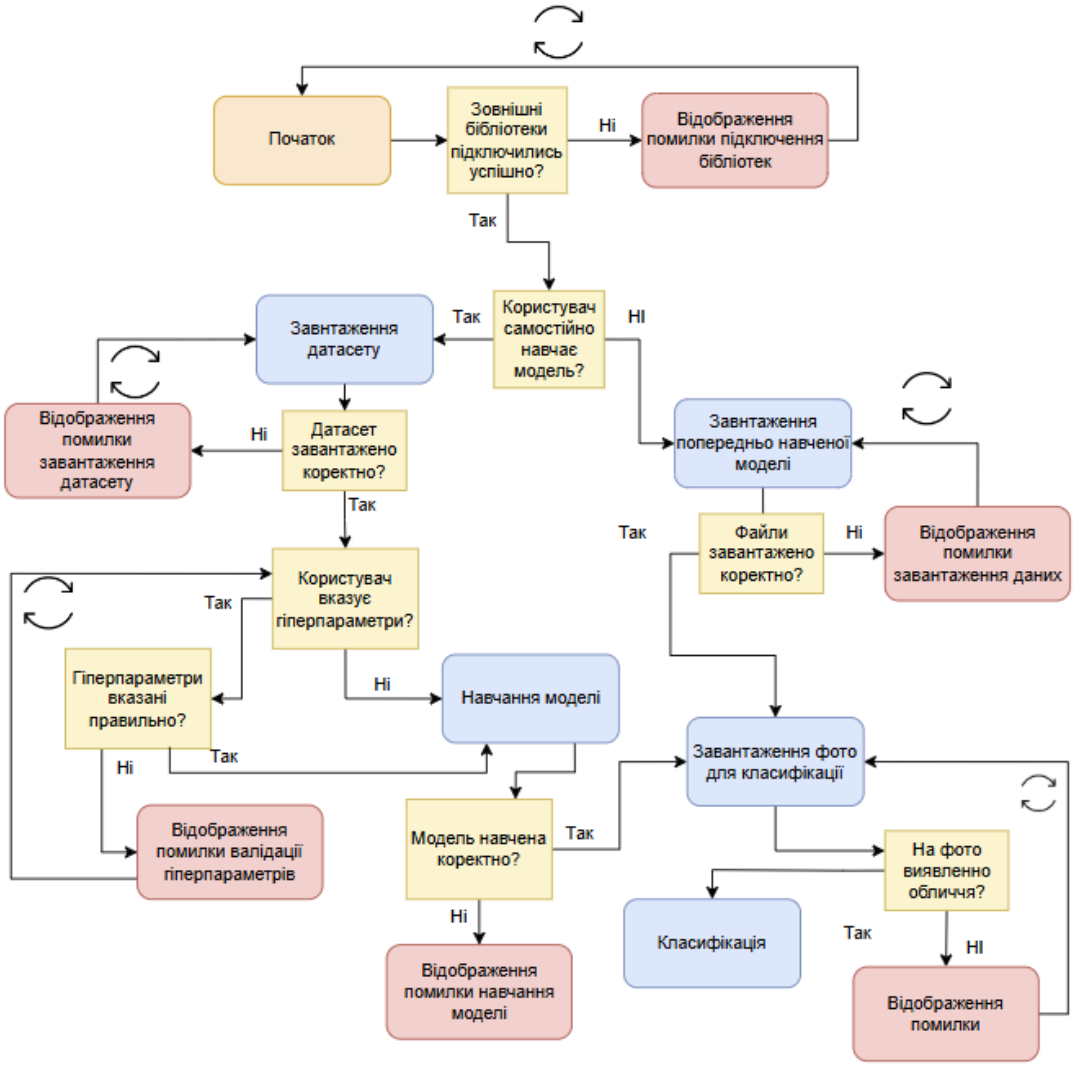


Рисунок 3.6 – Схема функцій для перевірки даних

Перша помилка з якою може зіткнутись користувач – помилка підключення зовнішніх бібліотек. Найчастіші помилки при використання зовнішніх пакетів, мають статус 4xx або 5xx. Статус 4xx – помилка викликана на стороні клієнта. Оскільки запит вказано статично, а TensorFlow не потребує додаткових дій від користувача, помилка із статусом 4xx виникати не буде. Статус 5xx – помилка на стороні сервера. Якщо ж запит до TensorFlow поверне статус 5xx, бібліотека підключена не буде та метод функціонувати не буде. Якщо запит до BlazeFace поверне статус 5xx, метод буде працювати некоректно, оскільки буде пропущений етап виявлення обличчя на фото.

Перевірка завантаження датасету, є досить важлива, оскільки напряму впливає на навчання моделі. Помилка під час завантаження може бути викликана декількома факторами, такими як, невірний формат зображень, завантаження пустого датасету, тощо.

Також важливо реалізувати наступні перевірки:

– перевірка гіперпараметрів;

– коректність навчання нейромережевої моделі;

– правильність завантаження попередньо навченої моделі.

Отже в даному розділі було розроблено схему можливих помилок під час використання інформаційної системи виявлення модифікованих зображень облич людей. Розглянуто можливі сценарії коли і як можуть виникнути певні помилки та як їх уникнути.

3.5 Проєктування інтерфейсів користувача та опис їх функціональності

Проєктування та створення інтерфейсу користувача є невід’ємним етапом під час розробки методу виявлення модифікованих зображень облич людей. Інтерфейс повинен бути інтуїтивно зрозумілим, не викликати складнощів у нового користувача та водночас має містити всю потрібну інформацію.

Під час розробки інтерфейсу потрібно реалізувати наступні частини:

– шапка (header);
– навчання моделі;
– модальне вікно введення гіперпараметрів;
– блок використання попередньо навченої моделі;
– статус підключення зовнішніх бібліотек;
– блок завантаження зображення для класифікації;
– блок для виводу результатів.

Структура елементів інформаційної системи зображена на Рисунку 3.7.

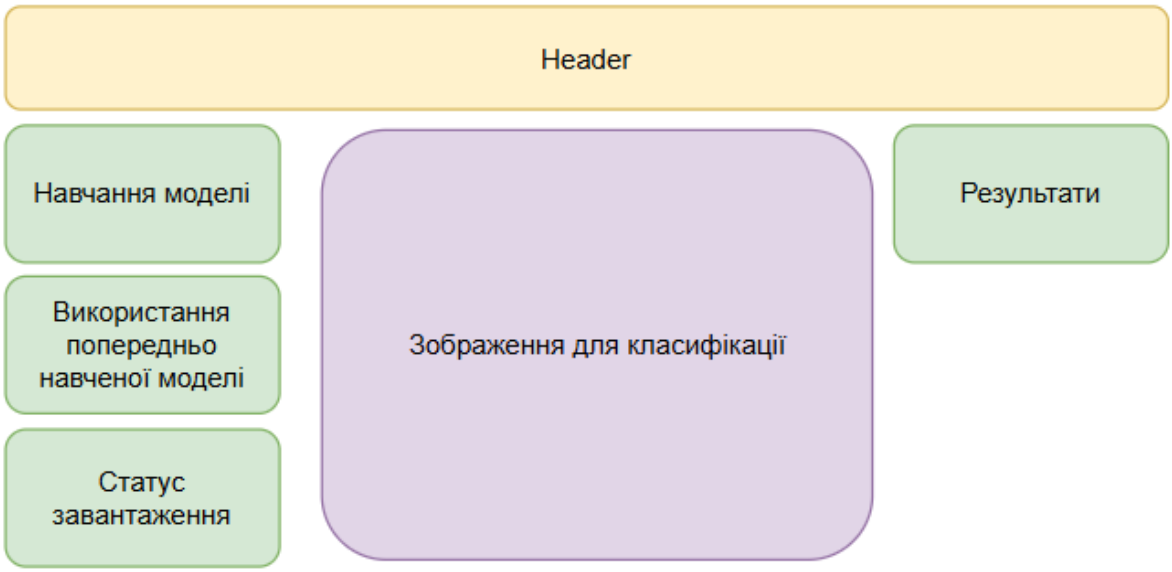


Рисунок 3.7 – Структура елементів інформаційної системи

Верхня частина інформаційної системи, шапка (header), не є критично важливою, проте туди можна вивести інформацію про розробника чи назву проєкту.

Блок, який відповідає за навчання моделі має містити вхідне поле, куди користувач завантажує датасет та кнопка яка активує процес навчання. Також даний блок містить кнопку, яка виконує функцію завантаження навченої моделі.

Частина інтерфейсу, яка відповідає за використання попередньо навченої моделі, має містити два вхідних поля для JSON та BIN файлів, а також кнопку, яка запускає функцію для завантаження.

Блок із статусом завантаження зовнішніх даних містить назви бібліотек, а також індикатор, який сповіщає про успішне підключення.

Для завантаження робочого зображення реалізовано вхідне поле, куди користувач вивантажує фото, та кнопку, яка запускає процес класифікації.

Для отримання результатів реалізовано окрему частину у інтерфейсі, куди виводяться усі вихідні дані.

Для того, щоб користувач мав можливість вводити гіперпараметри, потрібно реалізувати блок для введення значень (Рисунок 3.8).

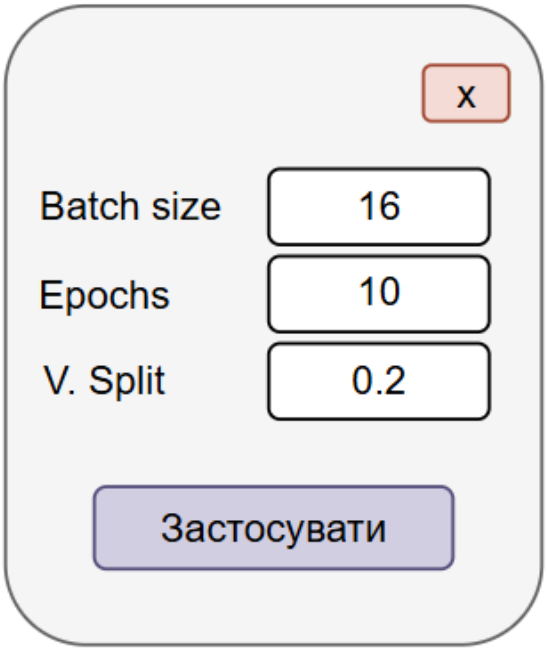


Рисунок 3.8 – Модальне вікно введення гіперпараметрів

Модальне вікно містить три вхідни поля, куди вводяться дані, та кнопку застосування. Також потрібно передбачити кнопку для закриття модального вікна.

Вікно для отримання результатів класифікації має бути динамічним, оскільки є декілька варіантів отримання результату роботи методу, а саме:

– обличчя не виявлено;
– обличчя на фото виявлено, фото модифіковане;
– обличчя на фото виявлено, фото не модифіковане.

Перший випадок застосовується тоді, коли на зображенні, завантажене користувачем обличчя не виявлено (Рисунок 3.9).

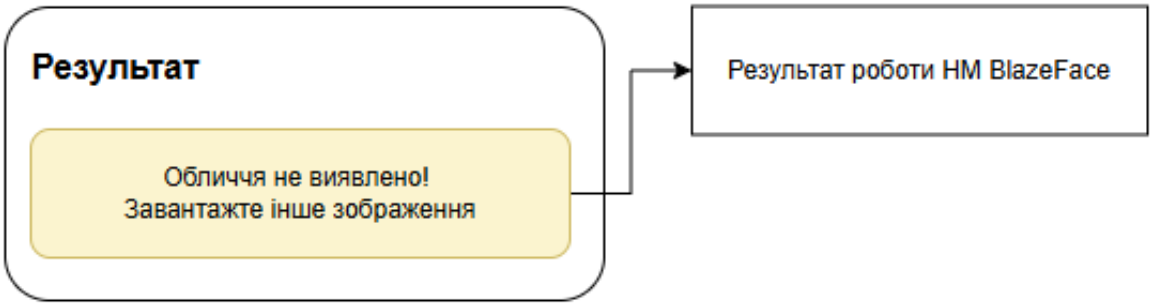


Рисунок 3.9 – Інтерфейс роботи НМ BlazeFace

В такому випадку, застосовується лише одна нейронна мережа BlazeFace, яка сповіщає користувача що на зображенні обличчя не виявлено.

Якщо ж на зображенні присутнє обличчя та фото модифіковане, блок виведення інформації матиме наступний вигляд (Рисунок 3.10).

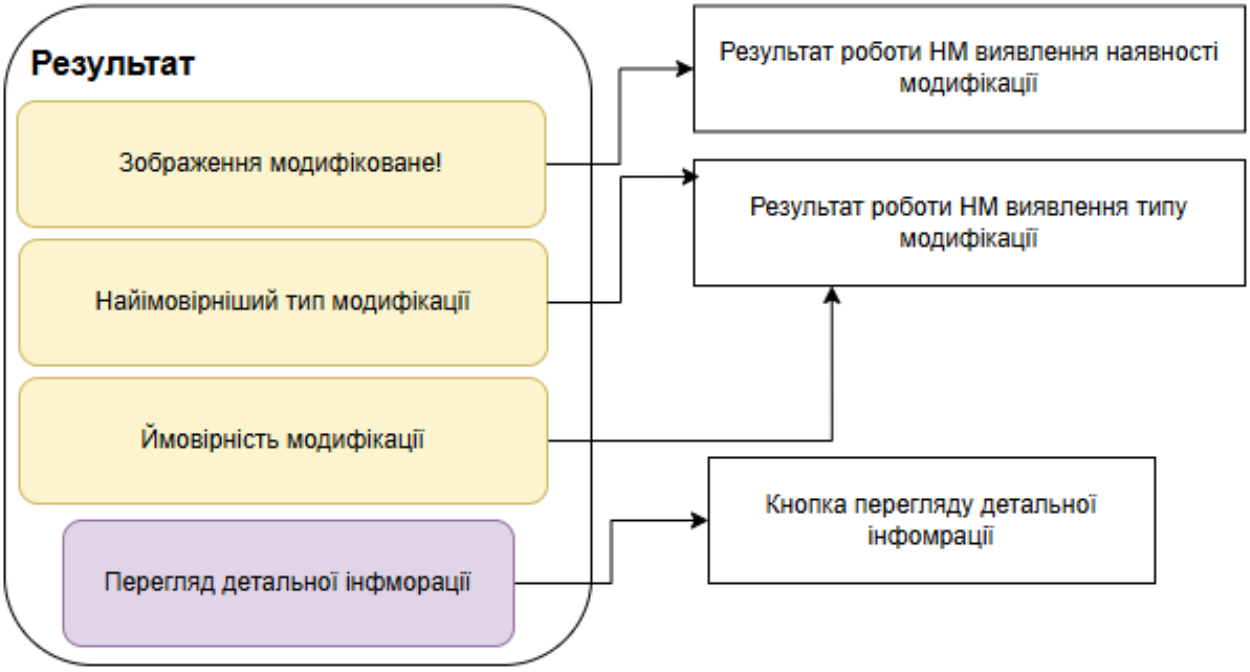


Рисунок 3.10 – Блок виведення інформації у випадку виявлення модифікації

Якщо ж натиснути на кнопку «Перегляд детальної інформації», користувачу стане доступним наступне вікно, в якому описано ймовірність належності фото до певної модифікації (Рисунок 3.11).

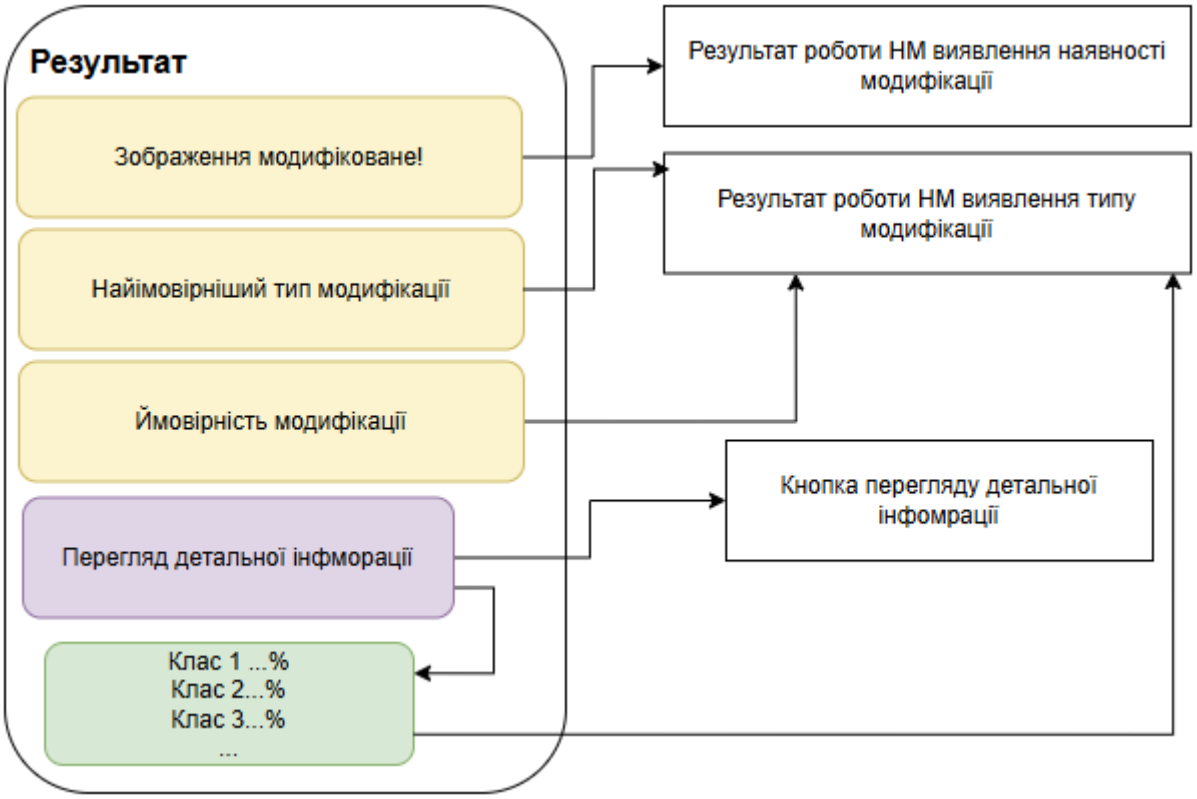


Рисунок 3.11 – Інтерфейс детального перегляду результату

У випадку, якщо ж обличчя виявлено, проте фото не модифіковане, користувачу стане доступним наступний блок (Рисунок 3.12).

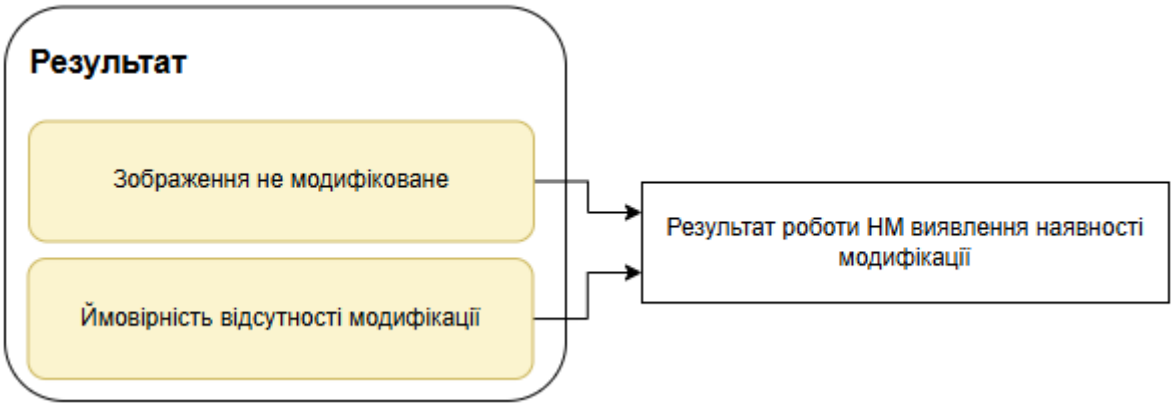


Рисунок 3.12 – Інтерфейс користувача у випадку відсутності модифікації

Отже, у даному розділі спроєктовано структурну схему елементів інформаційної системи, яка використовує метод виявлення модифікованих зображень облич людей та розписано їх значення.

Висновки до розділу 3

В даному розділі було обрано засоби розробки методу виявлення модифікованих зображень облич людей, а саме для каркасу сторінки – HTML, для стилізації елементів – CSS, програмна реалізація методу написана на мові програмування JavaScript. Також задіяні бібліотека TensorFlow для створення архітектури, навчання та тестування нейромережі та модель BlazeFace для попереднього виявлення обличчя. В якості редактора програмного коду обрано – Visual studio code.

Розроблено загальну схему інформаційної системи та поетапно розписано кроки від відкриття програмної системи до отримання результатів.

Спроєктовано компоненту схему, в якій зображено програмні компоненти, де вони задіяні та що отримують на вхід та віддають на вихід.

Дані схеми візуалізують етапи, які були програмно написані, для реалізації методу виявлення модифікованих зображень облич людей.

Побудовано схему можливих помилок та винятків які потрібно врахувати при розробці, задля коректної роботи методу.

РОЗДІЛ 4 Експериментальне дослідження методу виявлення модифікованих зображень облич людей нейромережевими засобами

4.1 Структура модулів для програмної реалізації методу виявлення модифікованих зображень облич людей

Під час розробки програмної реалізації методу виявлення модифікованих зображень облич людей, було створено наступні класи:

– classifier;

– datasetManager;

– faceDetection;

– imageLoader.

Окрім класів, був створений окремий файл, в якому описана логіка взаємодії методів із інфомраційною системою (Рисунок 4.1).

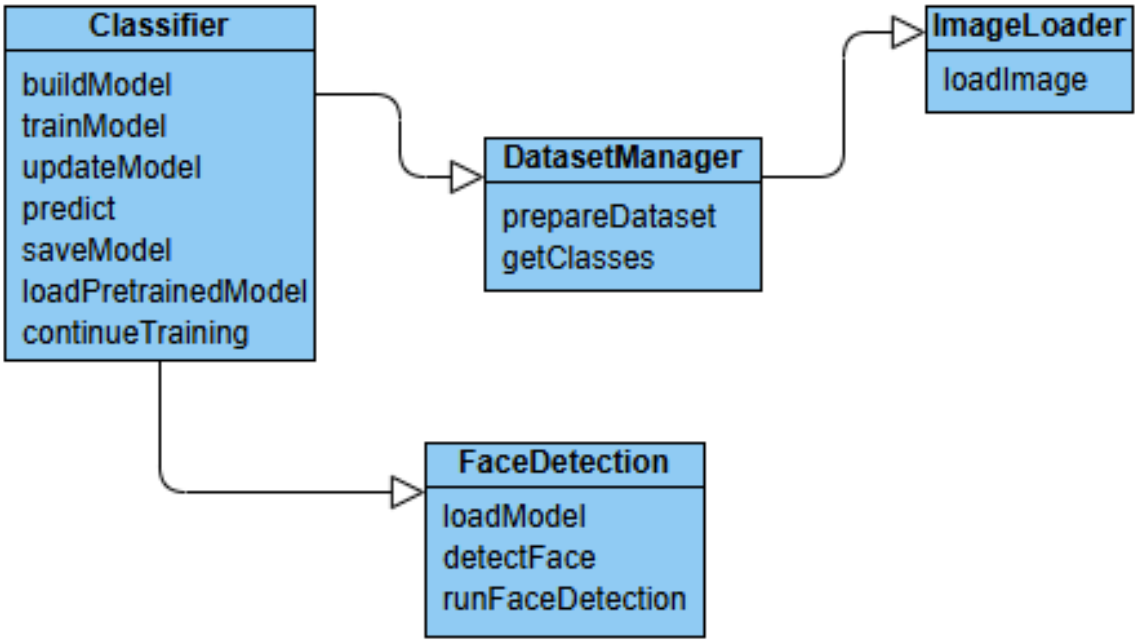


Рисунок 4.1 – Діаграма класів програмної реалізації методу виявлення модифікованих зображень облич людей

Оскільки для розробки методу використовувалась бібліотека tensorflow та модель blazeface, перед класифікацією потрібно перевірити коректність

підключення. Для цього було реалізовано функції які виконують перевірка статусу підключення та надають користувачу відповідні повідомлення.

Для виявленням та виокремлення облич, було створений клас «FaceDetection» із наступними методами:

– loadModel – завантаження моделі, перевірка статусу підключення;
– detectFaces – виконує перевірку на наявність обличчя;
– runImageForFaces – основний метод, який запускає гілку процесів для виявлення обличчя.

Для роботи із датасетом був створений клас «DatasetManager», конструктор якого приймає наступні праметри:

– labels – масив міток для класифікації;
– imageTensor – масив для зберігання тензорві зображень;
– datasetLabels – масив для зберігання відповідних міток для зображень.

Клас містить метод «prepareDataset», який зчитує файли зображень, створює тензори та мітки, а також метод «getClasses», який створює повертає список клсів, відповідно структурі датасету. Даний метод автоматично визначає клас модифікації, відповідно до назви папки. Це досить зручно оскільки використовуючи даний підхід, можна збільшувати масив, та кількість класів не змінюючи програмний код.

Також даний клас містить метод, для перевірки коректності завантаження датасету.

Клас «Classifier» відповідає за створення моделі для виявлення модифікованих зображень облич.

Метод «buildModel» створює нейромережеву модель із кількох шарів:

– згортковий шар;
– шар пулінгу;
– шар flatten, для перетворення матриці пікселів;
– шар dense, для обробки ознак;
– вихідний шар із функцією активації «softmax».

Метод «trainModel» відповідає за навчання моделі з заданими користувачем параметрами. Якщо ж гіперпараметри не задані, в якості значень приймають параметри за замовчуванням, а саме:

– epochs – 10;

– batchsize – 16;

– validationSplit – 0.2.

Метод «saveModel» відповідає за завантаження навченої моделі, а метод «loadPretrainedModel» за використання без повторного навчання.

Метод «countinueTraining» відповідає за подальше навчання моделі на нових вхідних даних, не втрачаючи попередні.

Для роботи із завантаженням зображень створено клас «imageLoader», який відповідає за попередню обробку та завантаження фото.

Також було написано декілька функцій для виводу інформації та перевірки коректності введених даних користувачем, а саме:

– trainButton;

– predictButton;

– saveModel;

– loadModel;

– countinueTraining;

TrainButton – кнопка яка запускає процес навчання нейромережевої моделі, отримавши вхідні дані із поля dataset, попередньо перевіривши коректність завантаження датасету. В разі успіху вивід повідомлення користувачу про успішне навчання, в випадку неуспішного тренування – вивід коду помилки.

PredictButton відповідає за запуск функції класифікації зображення завантаженого користувачем, включаючи функцій для обробки зображень та перевірки коректності завантаження фото. В разі успішного проходження всіх перевірок, запускається процес класифікації після чого виводиться результат у відповідне поле.

SaveModel виконує перевірку чи відбувся процес тренування моделі, запускає функції збереження попередньо навченої моделі у відповідних файлах.

LoadModel – кнопка яка отримує JSON та BIN файли завантажені користувачем у відповідні поля, попередньо перевіряючи коректність завантаження. Якщо файли отримані коректно запускає процес компіляції нейромережевої моделі. В випадку помилки виводить код та відповідне повідомлення.

Кнопка continueTraining відповідає за запуски декількох функції для продовження навчання попередньо навченої моделі. Попередньо перевіряється чи користувач завантажив навчену модель та вхідні дані для донавчання.

В даному розділі описані основні реалізовані класи та функції які були написані для реалізації методу виявлення модифікованих зображень облич людей. Набір даних функцій реалізує основний функціонал, такий як, створення та навчання нейромережевої моделі.

4.2 Організація прикладного донавчання нейромережевої моделі для виявлення модифікованих зображень облич людей

Оскільки для навчання класифікатора методу виявлення модифікованих зображень облич людей потрібна велика кількість вхідних даних, виникла проблема в нестачі обчислювальних ресурсів комп’ютера. Коли для навчання завантажувалось відразу 1000 зразків зображеннь, виникала помилка втрати контексту WebGL (Рисунок 4. 2).

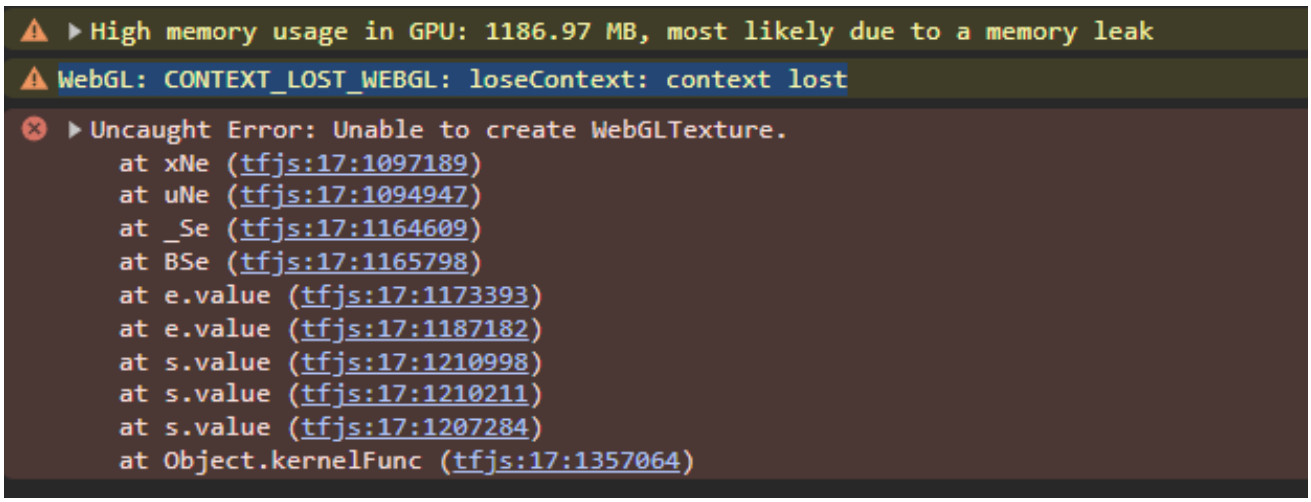


Рисунок 4.2 – Помилка втрати контексту WebGl

Дана помилка виникає тоді, коли метод вимагає занадто велику кількість ресурсів графічного процесора. Зміна гіперпараметрів та розмірності вхідних зображень частково зменшує навантаження, проте цього недостатньо. Було прийнято рішення реалізувати функціонал поступового навчання моделі. Тобто не завантажувати відразу 1000 зображень, а розділити датасет на рівномірну кількість фото, наскільки дозволяють параметри комп'ютера та поступово навчати модель.

Для цього була реалізована функція завантаження та вивантаження навченої моделі. Тобто користувач, після навчання, може завантажити та використовувати навчену модель, без потреби в повторному навчанні.

Після того як була реалізована функція завантаження навченої моделі, тобто отримано файли із вагами та структурою нейронної мережі, було реалізовано функцію покращення (донавчання) моделі. Даний метод складається із наступних кроків:

- підготовка моделі;
- перевірка наявності моделі;
- підготовка даних для навчання;
- компіляція моделі;
- запуск процесу навчання;

Крок підготовки моделі отримує завантаженні користувачем файли, після чого відбувається перевірка коректності завантаження файлів та чи готова модель для подальшого навчання.

Після підготовки даних до навчання, відбувається компіляція моделі, оскільки якщо пропустити даний крок, буде отримана помилка, яка вказує що для подальшого навчання потрібно визначення моделі та метод втратить функціональність.

Після компіляції викликається метод «fit», бібліотеки tensorflow, який запускає процес навчання.

Під час дослідження ефективності, виявлено що для досягнення бажаної точності, в середньому потрібно 8 епох. Оскільки в інформаційній системі методу

виявлення модифікованих зображень облич людей, передбачений функціонал введення гіперпараметрів користувачем, потрібно передбачити можливість зупинки навчання класифікатора, після того як модель досягнула бажаної точності, щоб передбачити перенавчання та некоректне використання обчислювальних ресурсів (Рисунок 4.3).

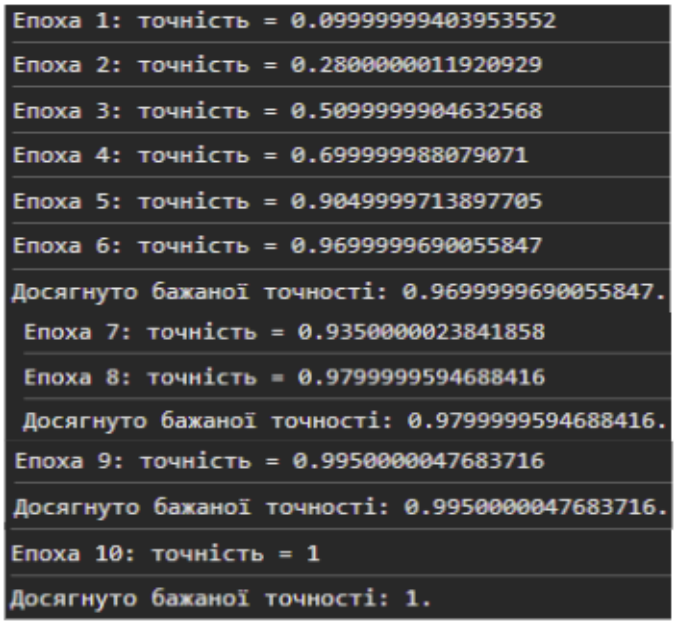


Рисунок 4.3 – Точність на різних етапах навчання

Проаналізувавши Рисунок 4.3, видно, що класифікатор досягнув бажаної точності ще на 7 епосі, проте продовжив навчання, оскільки немає точки зупинки.

За допомогою готової функції tensorflow, було реалізовано зупинку навчання класифікатора при досягнені бажаної точності (Рисунок 4.4).

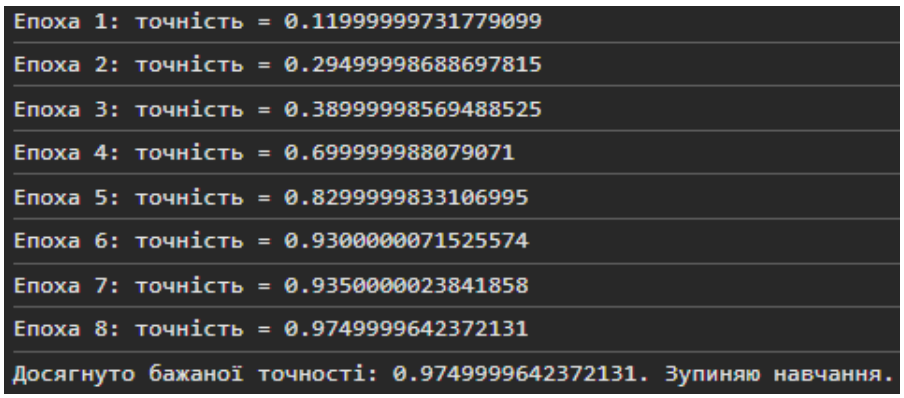


Рисунок 4.4 – Результат роботи функції зупинки

Для кращої масштабованості методу виявлення модифікованих зображень облич людей, розроблений функціонал автоматичного визначення кількості класів вхідного набору даних. Для цього написано функцію, яка автоматично враховує структуру датасету відповідно до назв папок (Рисунок 4.5).

Рисунок 4.5 – Автоматичне створення масиву класів

Завдяки набору функцій javascript, методу враховує стркутуру датасету, та додає до масиву унікальні назви папок.

Якщо ж датасет містить підкласи, будуть додані відповідні мітки класів до масиву (Рисунок 4.6).

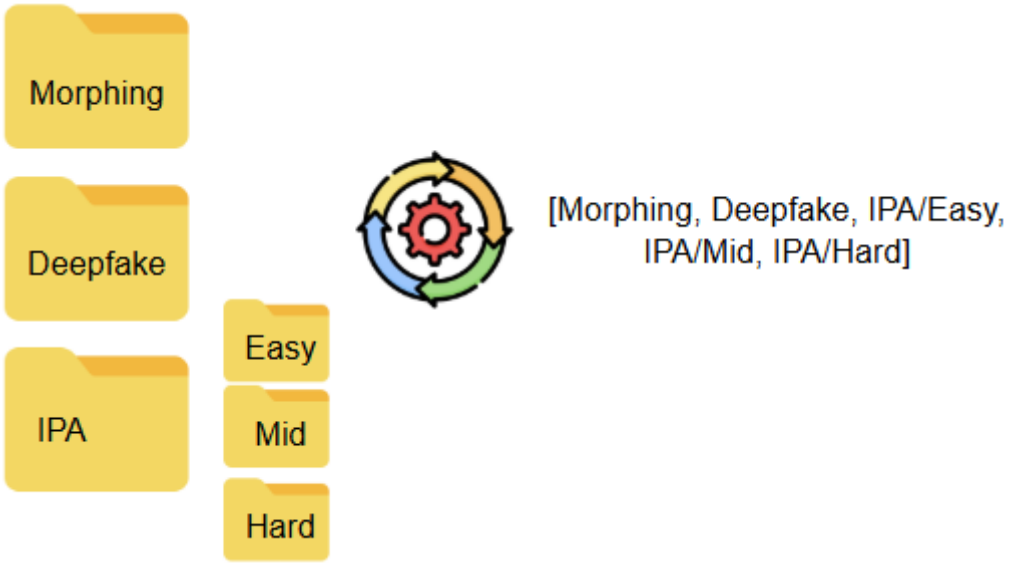


Рисунок 4.6 – Врахування підкласів

Отже в даному розділі розглянуто реалізацію функціоналу донавчання нейромережевої моделі. Як результат, отриману навчену модель, на більшій кількості зображень, без виникнення помилок із нестачею обчислювального ресурсу. Для навчання класифікатору методу виявлення модифікованих зображень облич людей, датасет було розділено на 5 частин по 200 зображень у кожній, які поступово покращували модель, уникаючи помилки втрати контексту WebGl. Також реалізовано функціонал зупинки навчання нейромережевої моделі, при досягненні бажаної точності.

4.3 Дослідження функціональних можливостей експериментальної інформаційної системи

Під час розробки методу для виявлення модифікованих зображень облич людей, важливо зробити візуально привабливий та інтуїтивно зрозумілий інтерфейс (Рисунок 4.7).

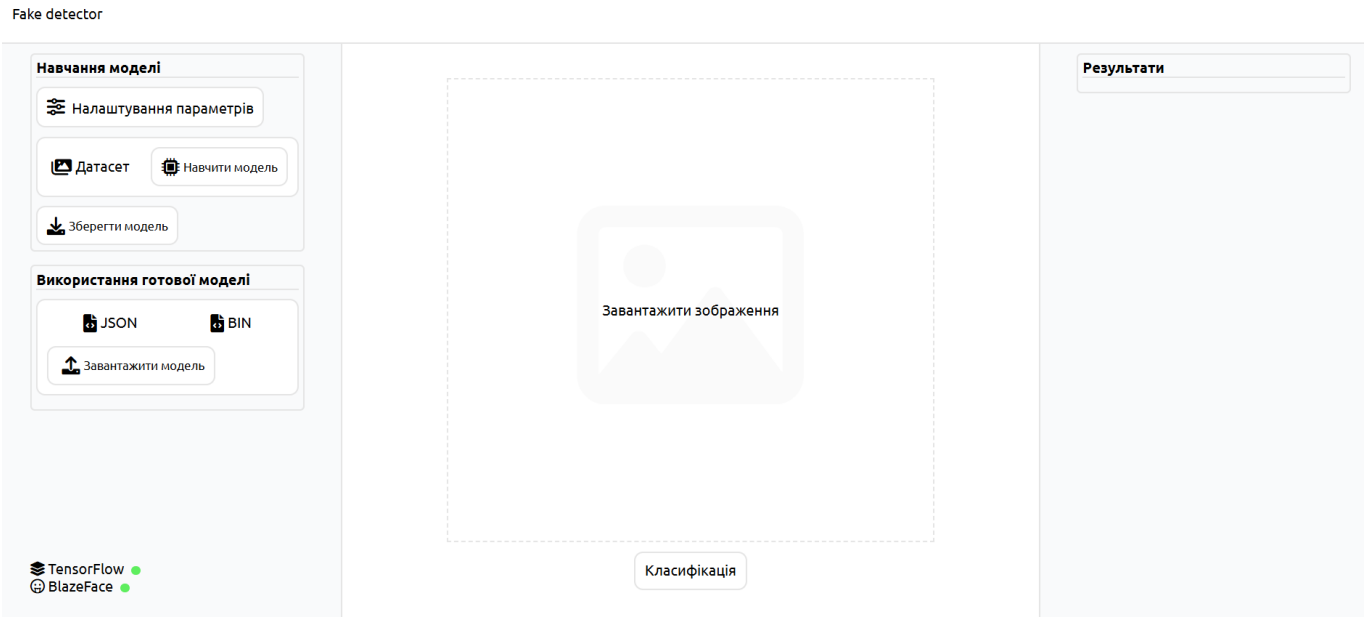


Рисунок 4.7– Інтерфейс інформаційної системи

Оскільки використовується зовнішнє підключення до бібліотеки tensorflow та моделі balzeface, перед початком класифікації, важливо переконатись що

завантаження та підключення відбулось успішно. Для цього було створено блок, на якому відображаються статуси підключень. У разі успішного завантаження та підключення, користувач зможе побачити зелену мітку, та відповідне повідомлення у консолі (Рисунок 4.8).

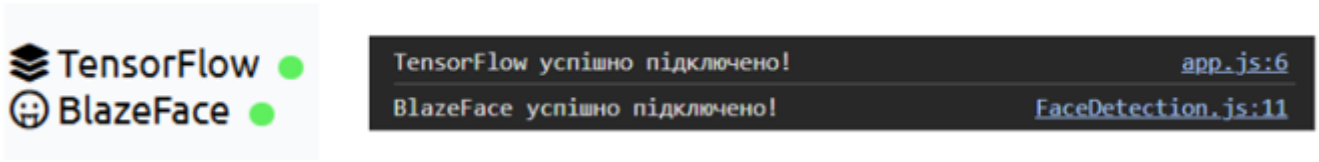


Рисунок 4.8 – Успішний статус підключення зовнішніх бібліотек

В противному випадку користувачу відображається червона мітка, та повідомлення про помилку (Рисунок 4.9).

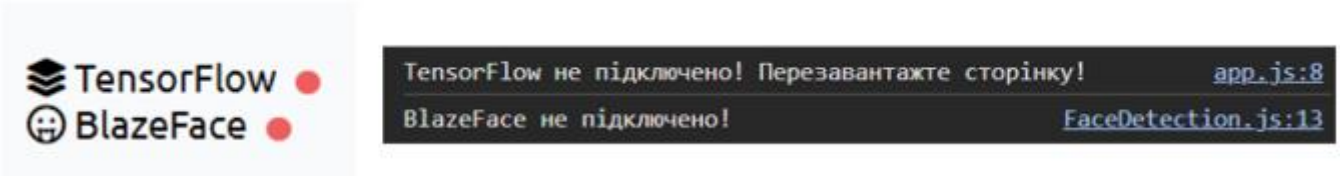


Рисунок 4.9 – Помилка при підключенні зовнішніх бібліотек

Блок «Навчання моделі» створений для тренування моделі на власному датасеті, попередньо задавши гіперпараметри для навчання (Рисунок 4.10).

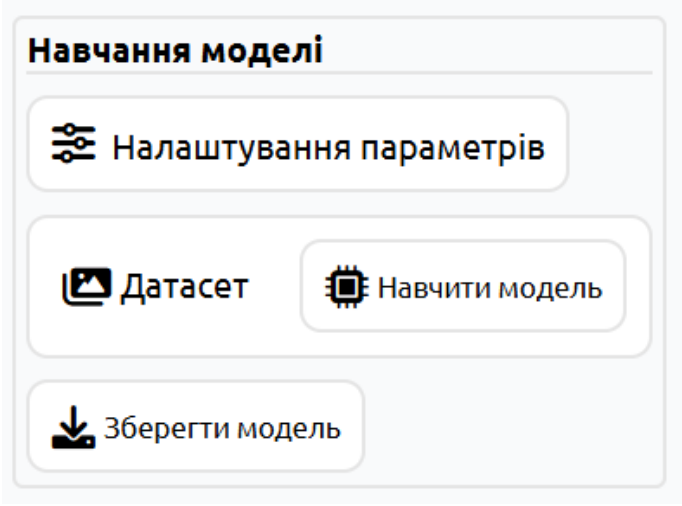


Рисунок 4.10 – Блок «Навчання моделі»

Натиснувши на кнопку «Налаштування параметрів», користувач зможе змінити гіперпараметри для навчання моделі (Рисунок 4.11).

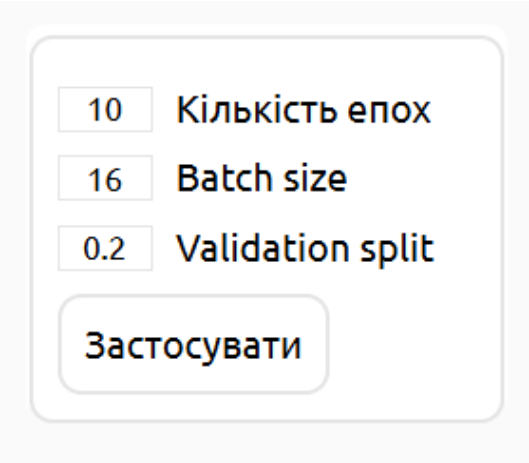


Рисунок 4.11 – Блок «налаштування гіперпараметрів»

Після навчання нейромережевої моделі, користувач має змогу зберегти навчену модель. Данна можливість дає зберегти файли у форматі json та bin, в яких міститься вся потрібна інформація, для подальшого використання навченої моделі без повторного тренування (Рисунок 4.12).

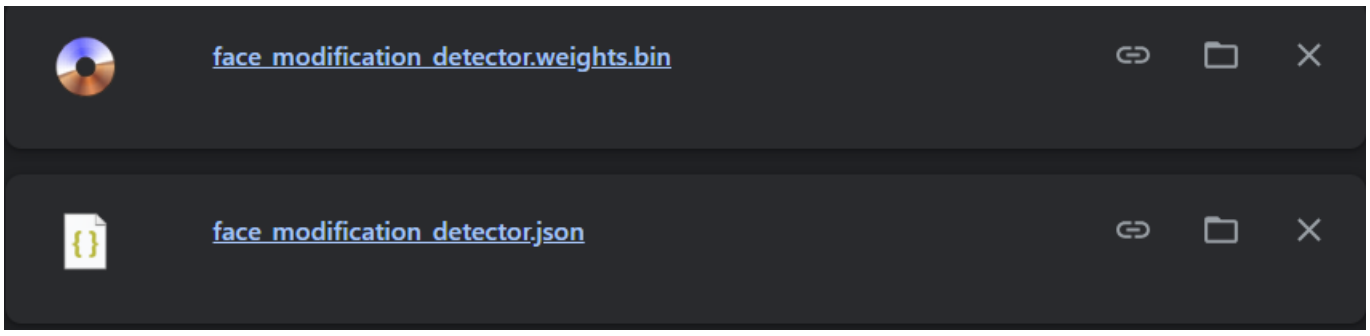


Рисунок 4.12 – Завантаження навченої моделі

Блок «Використання готової моделі» дає змогу користувачу використати вже готову, навчену, модель без потреби додаткового навчання (Рисунок 4.13).

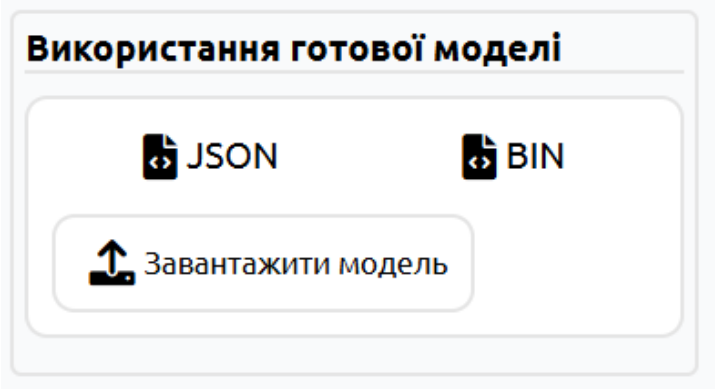


Рисунок 4.13 – Блок «Використання готової моделі»

Користувачу необхідно вибрати JSON файл що описує структуру моделі і метадані та BIN файл, що містить ваги моделі, які були створені під час тренування.

Перед класифікації зображення, відбувається перевірка на наявність обличчя на фото. Якщо обличчя не виявлено, відображається відповідне повідомлення (Рисунок 4.14).

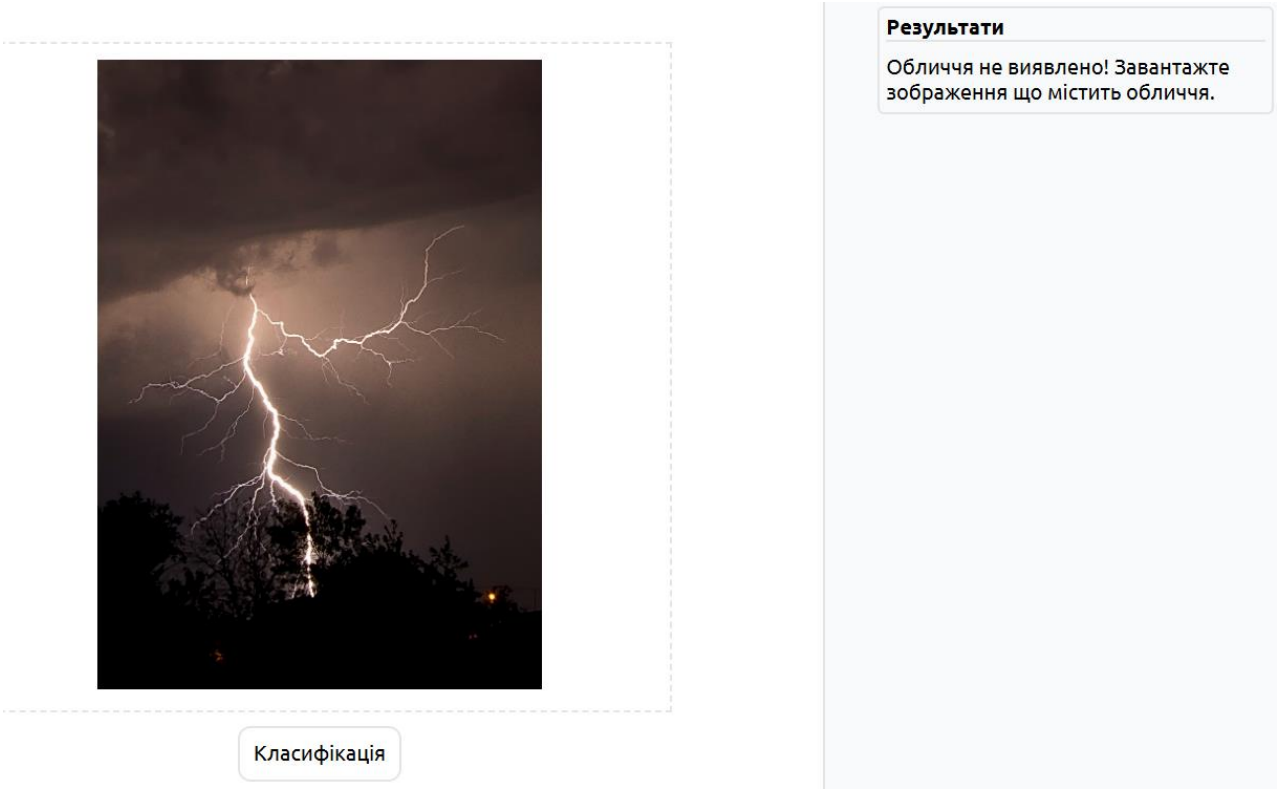


Рисунок 4.14 – Перевірка наявності обличчя на фото

Якщо ж blazeface виявляє обличчя, запускаються функції класифікації, після чого виводиться відповідний результат. Якщо фото модифіковане, блок виводу результату має наступний вигляд (Рисунок 4.15).

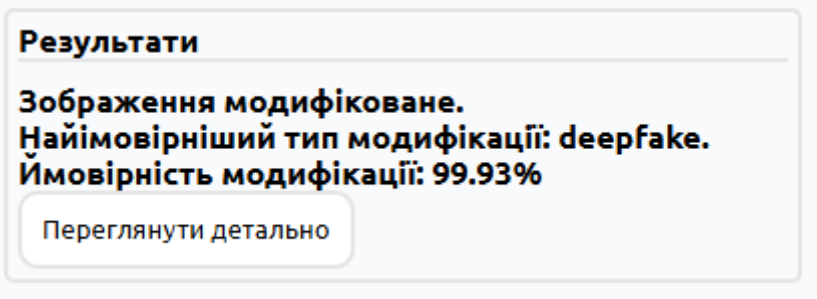


Рисунок 4.15 – Вивід результатів класифікації користувачу

Якщо користувач натисне на кнопку «Переглянути детально», відкриється наступне вікно, у якому результат відображається в відсотковому співвідношенні належності до кожного класу (Рисунок 4.16).

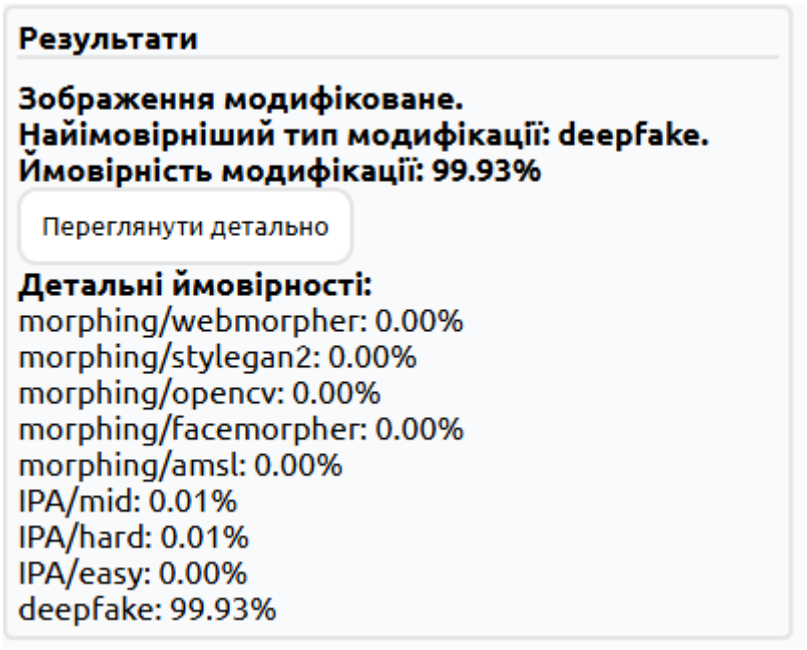


Рисунок 4.16 – Перегляд детальної інфомрації

Якщо ж фото не модифіковане, користувачу буде надано наступний результат (Рисунок 4.17).

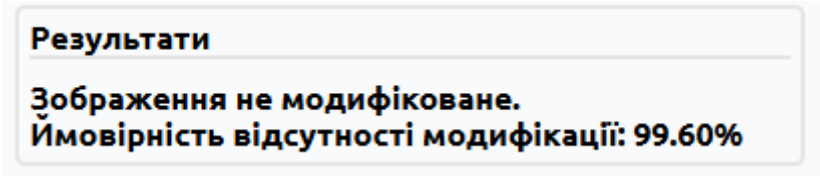


Рисунок 4.17 – Результат класифікації у випадку немодифікованого фото

Для тестування прикладної реалізації методу виявлення модифікованих зображень облич людей, було створено декілька тест-кейсів, які перевіряють основний функціонал методу.

Користувач має змогу натренувати модель за допомогою власних даних, тест-кейс даної функції наведено у Таблиці 4.1 (Рисунок 4.18 – 4.19).

Таблиця 4.1 – Тест-кейс A01

<table>
<tr><td>Тест-кейс ID: A01</td><td>Пріоритет: 1</td><td>Створено:07.11.2023, А.В. Похитун</td></tr>
<tr><td colspan="3">Назва: Перевірка функції тренування моделі
Вхідні дані:
1. Датасет із зображеннями
Передумова: Датасет має містити хоча б по одній фото для кожного класу</td></tr>
<tr><td colspan="2">Кроки</td><td>Очікуваний результат</td></tr>
<tr><td colspan="2">1. Відкрити інформаційну систему
2. Натиснути кнопку «Датасет»
3. Обрати папку із датасетом
4. Натиснути кнопку «Навчити модель»</td><td>Кнопка «Зберегти модель» стала активною Вивелась інформація про успішне навчання моделі</td></tr>
<tr><td colspan="3">Результат виконання тест-кейсу: пройдено успішно</td></tr>
</table>

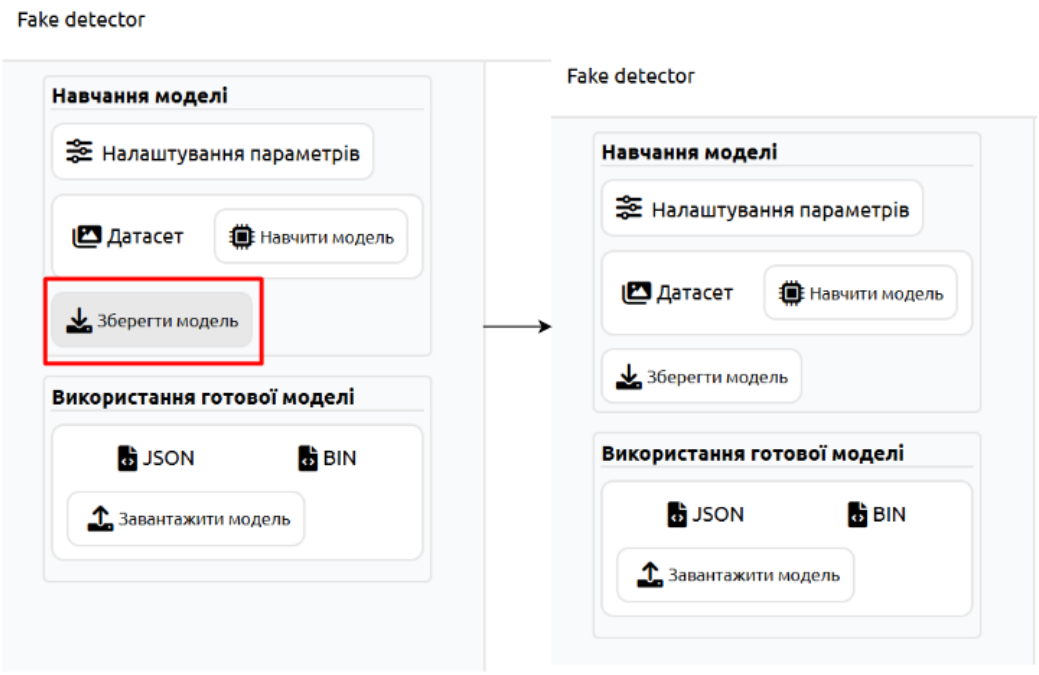


Рисунок 4.18 – Зміна статусу кнопки на «активна»

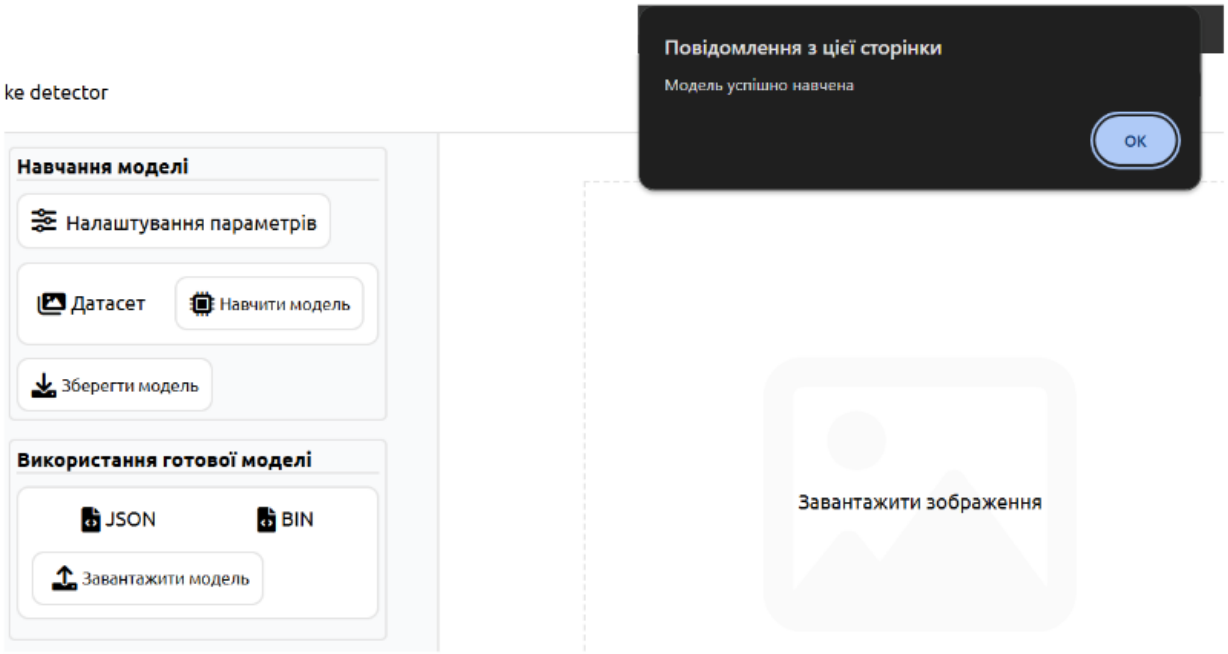


Рисунок 4.19 – Сповіщення про успішне навчання моделі

Після навчання моделі, користувач має змогу завантажити її для використання без повторного тренування. Тест-кейс цієї функції описано у Таблиці 4.2 (Рисунок 4.20).

Таблиця 4.2 – Тест-кейс A02

Тест-кейс ID: A02	**Пріоритет: 2**	**Створено:**07.11.2023, А.В. Похитун
Назва: Перевірка функції завантаження навченої моделі **Вхідні дані:** **Передумова:** Модель має бути попередньо навчена		
Кроки		**Очікуваний результат**
1. Відкрити інформаційну систему 2. Натиснути кнопку «Зберегти модель»		Завантаження фалів формату JSON та BIN
Результат виконання тест-кейсу: пройдено успішно		

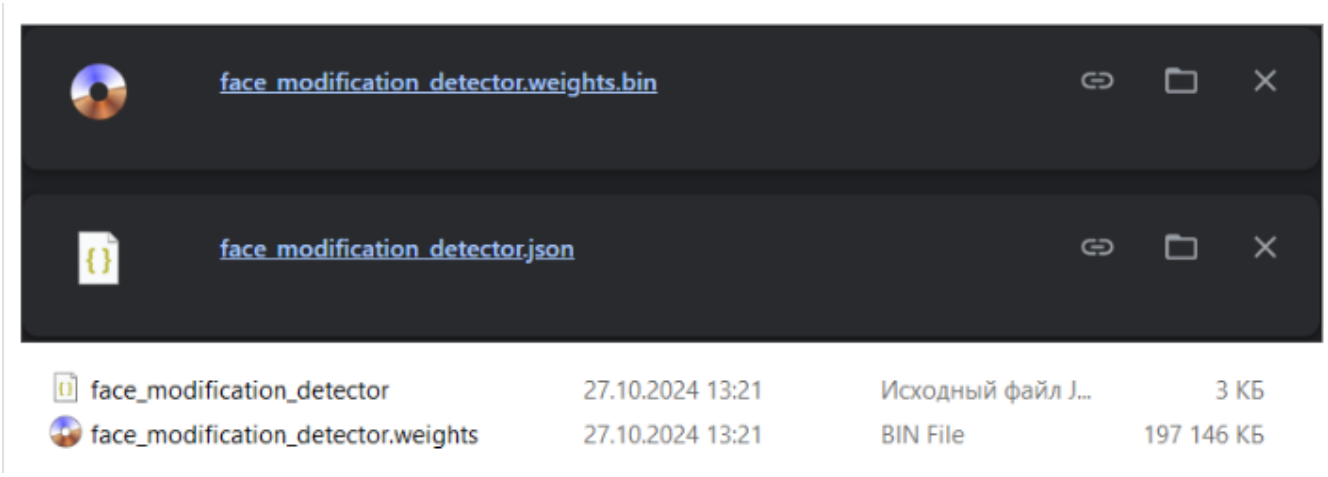


Рисунок 4.20 – Результат завантаження попередньо навченої моделі

Для того щоб користувач зміг використати попередньо навчену модель, реалізований функціонал використання готової моделі. Опис тест-кейсу даного функціоналу наведено у Таблиці 4.3 (Рисунок 4.21).

Таблиця 4.3 – Тест-кейс A03

<table>
<tr><td>Тест-кейс ID: A03</td><td>Пріоритет: 3</td><td>Створено:07.11.2023, А.В. Похитун</td></tr>
<tr><td colspan="3">Назва: Перевірка функції використання готової моделі
Вхідні дані:
1. JSON файл
2. BIN файл
Передумова: Модель має бути попередньо навчена та завантажена</td></tr>
<tr><td colspan="2">Кроки</td><td>Очікуваний результат</td></tr>
<tr><td colspan="2">1. Відкрити інформаційну систему
2. Натиснути кнопку «JSON», обрати відповідний файл
3. Натиснути кнопку «BIN», обрати відповідний файл
4. Натиснути кнопку «Завантажити модель»</td><td>Вивід результату про успішне завантаження готової моделі</td></tr>
<tr><td colspan="3">Результат виконання тест-кейсу: пройдено успішно</td></tr>
</table>

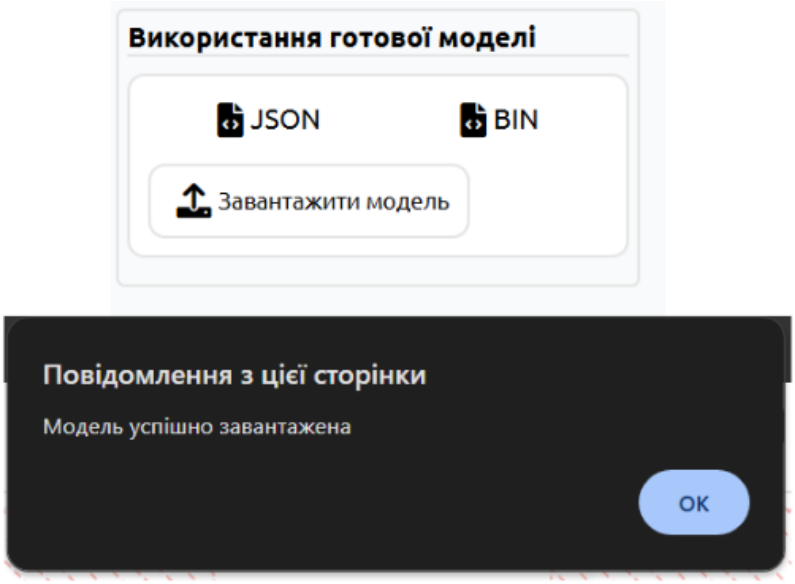


Рисунок 4.21 – Результат тестування функціоналу завантаження моделі

Отже, було детально розписано варіативність дій користувача при роботі із інформаційною системою, що використовує метод виявлення модифікованих зображень облич людей. Також була проведена робота із тестуванням певного функціоналу інформаційної системи виявлення модифікованих зображень облич людей. Було реалізовано 3 тест-кейса, які перевіряють працездатність функцій збереження, завантаження та навчання моделі за допомогою набору даних.

4.4 Дослідження методу виявлення модифікованих зображень облич людей

4.4.1 Дослідження нейромережі для виявлення наявності модифікації

Під час розробки методу виявлення модифікованих зображень облич людей, слід звернути увагу, на те що архітектура, гіперпараметри, розмір датасету впливають на швидкість навчання та точність класифікації.

Під час проведення дослідження ефективності нейронної мережі для ідентифікації модифікованих зображень облич людей, виявлено що модель досягає бажаної точності за 8-10 епох (Таблиця 4.4, Рисунок 4.22).

Таблиця 4.4 – Точність класифікатора на різних епохах

Епохи	Acuraccy
1	0.10
2	0.24
3	0.36
4	0.50
5	0.72
6	0.87
7	0.90
8	0.96
9	0.96

Результати тестування з таблиці 4.4 наведені на діаграмі (Рисунок 4.22).

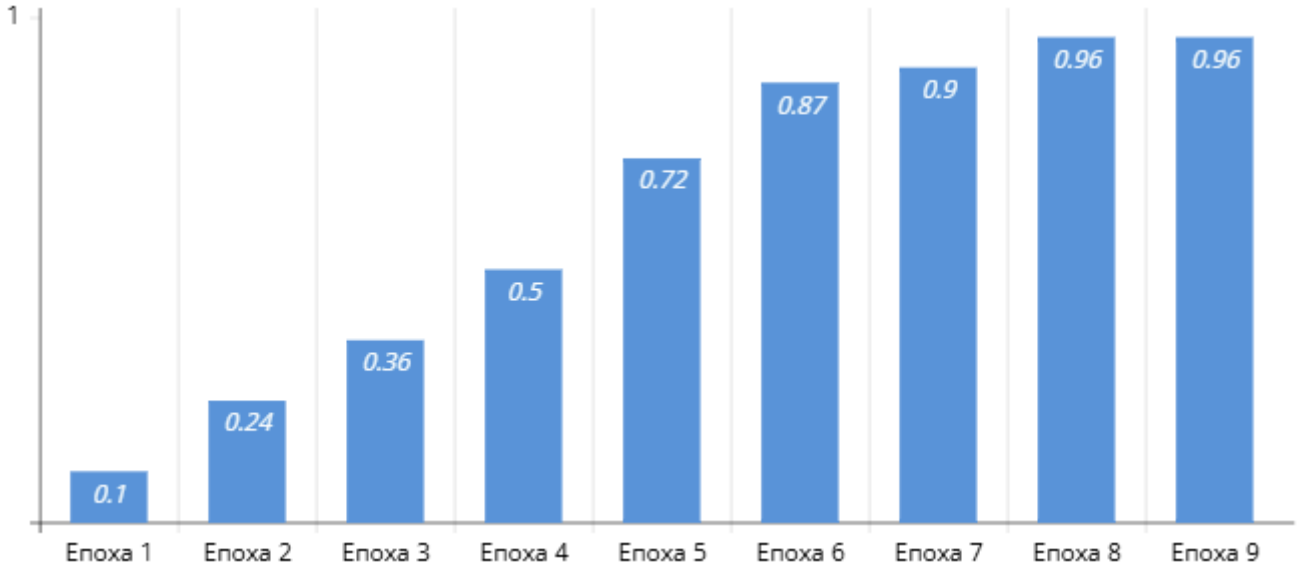


Рисунок 4.22 – Точність на різних стадіях навчання

Для проведення дослідження ефективності, модель було навчено на різній кількості зображень та проведено тестування (Рисунок 4.23). Результати описані у Таблиці 4.5.

Таблиця 4.5 – Ефективність виявлення наявності модифікацій

Кількість зображень	Час навчання (с)	Accuracy	Precision	Recall	F1 score
200	125.5	0.98	0.97	0.97	0.97
400	180.8	0.955	0.92	0.92	0.92
600	253.3	0.96	0.95	0.95	0.95
800	324.7	0.99	0.98	0.98	0.98
1000	377.6	0.95	0.95	0.95	0.95

Результати дослідження з таблиці 4.5 наведені на діаграмі (Рисунок 4.23).

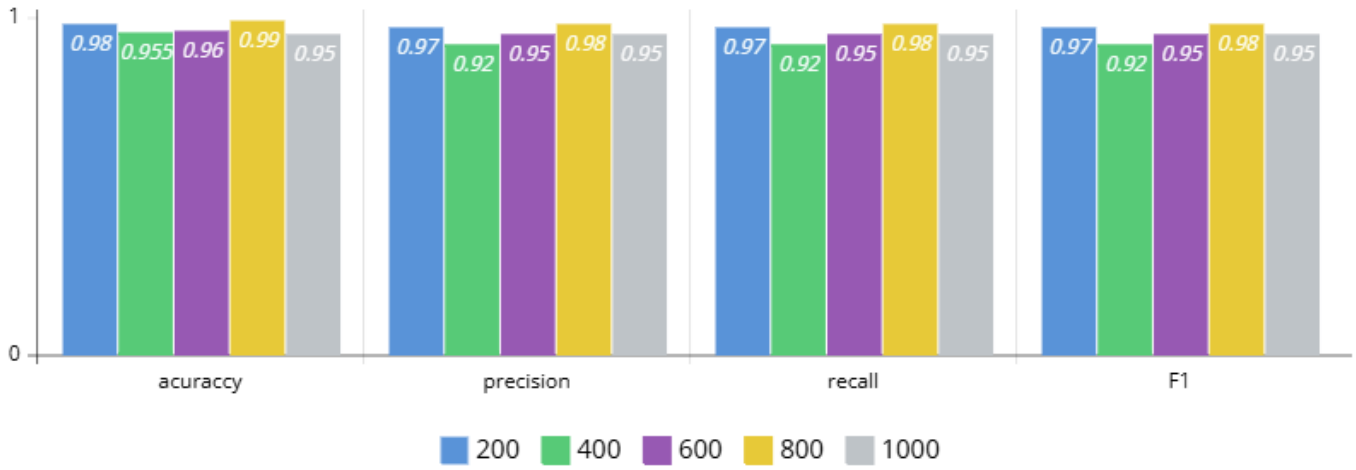


Рисунок 4.23 – Результати дослідження виявлення наявності модифікацій

Проаналізувавши таблицю 4.5, можна зробити висновок, що оптимальна кількість вхідних даних для навчання класифікатора методу виявлення модифікованих зображень – 800 зразків, при цьому швидкість навчання = 324.7 секунди, accuracy = 0.99, precision = 0.98, recall = 0.98, F1 = 0.98.

В ході дослідження ефективності навчання, проведено декілька тестувань із різними гіперпараметрами. Тестування проводились на відносно невеликій вибірці зображень (200 екземплярів), результати описані у таблиці 4.6.

Таблиця 4.6 – Ефективність виявлення наявності модифікацій із різними гіперпараметрами

Batch size	Epochs	Час навчання(с)	Accuracy	Precision	Recall	F1
20	16	187	0.96	0.96	0.96	0.96
16	10	152	0.98	0.97	0.97	0.97
12	8	136	0.95	0.962	0.962	0.962

Результати експерименту з таблиці 4.6 наведені на діаграмі (Рисунок 4.24).

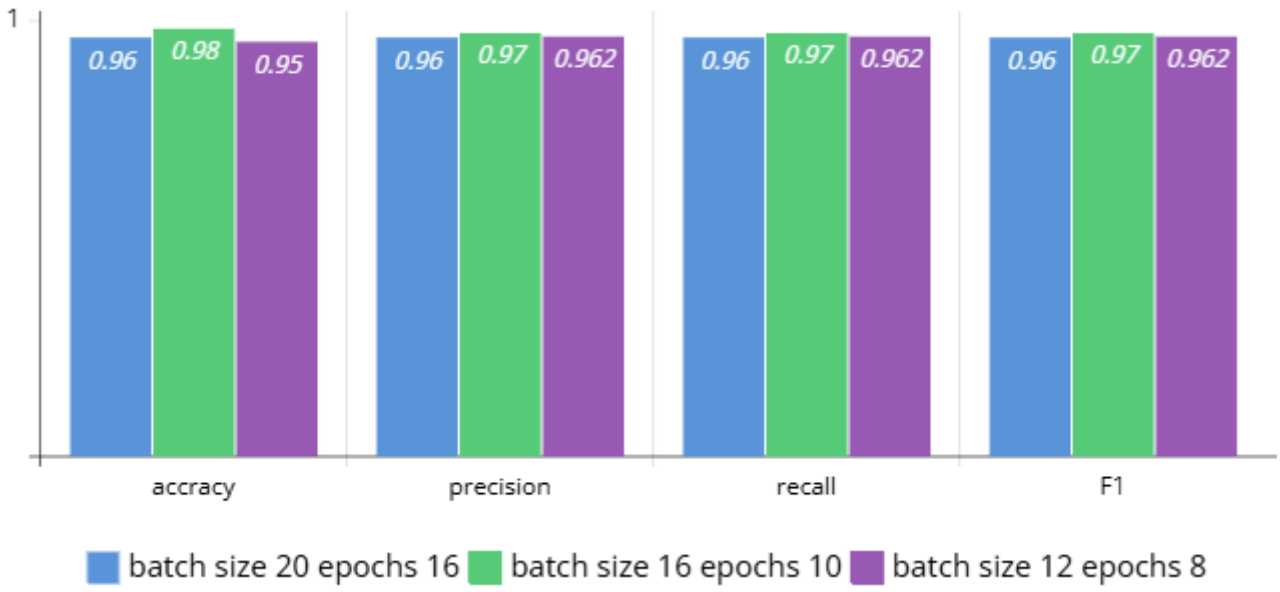


Рисунок 4.24 – Вплив гіперпараметрів на результат класифікатора

Як видно з рисунку 4.24 та з таблиці 4.6, найкращі результати навчання досягнуті при параметрах Batch size = 16 та Epochs = 10. Для таких гіперпараметрів час навчання нейромережі для виявлення модифікацій склав 152 секунди. При цьому отримані метрики Accuracy = 0.98, Precision = 0.97, Recall = 0.97, F1 = 0.97.

Отже, провівши дослідження, виявлено, що для розробки методу виявлення модифікованих зображень облич людей, найкраща точність була отримана при розмірі датасету 800 екземплярів та при вказанні наступних гіперпараметрів batch size = 16, epochs = 10.

4.4.2 Дослідження нейромережі для виявлення видів модифікації

Під час проведення дослідження ефективності методу виявлення модифікованих зображень облич людей, виявлено що в середньому модель досягає бажаною точності (0.95%) за 8-10 епох (Таблиця 4.7, Рисунок 4.25).

Таблиця 4.7 – Точність класифікатора на різних епохах

Епохи	Acuraccy
1	0.11
2	0.23
3	0.38
4	0.57
5	0.75
6	0.9
7	0.94
8	0.96

Результати тестування з таблиці 4.7 наведені на діаграмі (Рисунок 4.25).

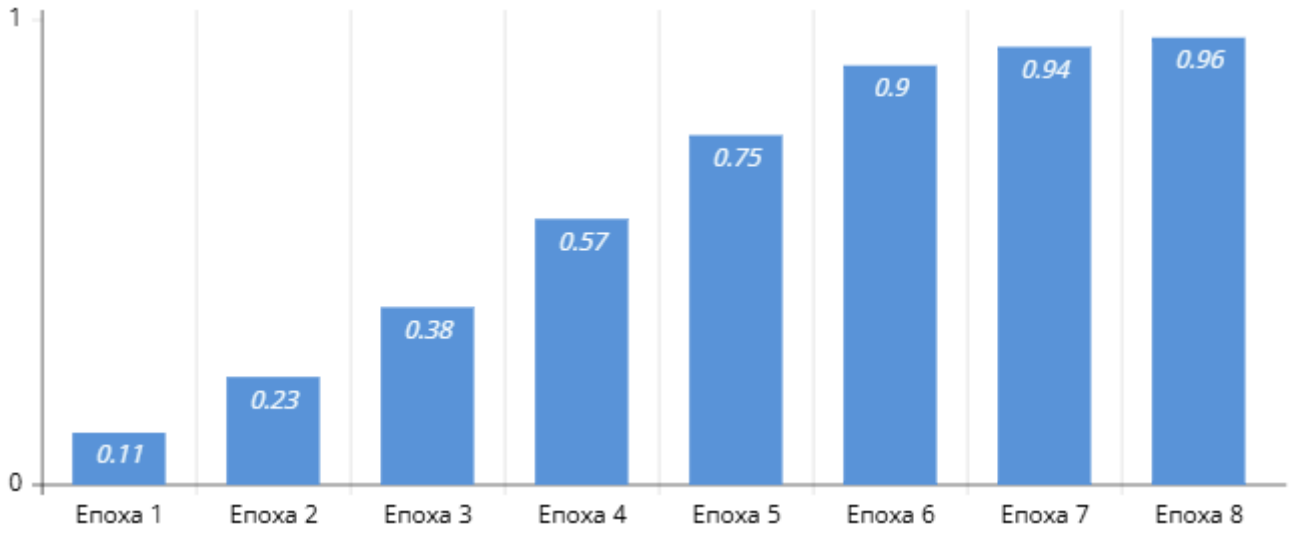


Рисунок 4.25 – Точність на різних стадіях навчання

Отже, проаналізувавши таблицю 4.7, можна зробити висновок, що для навчання класифікатора оптимально зазначити 10 епох.

В ході проведення дослідження, модель було навчено на різній кількості вхідних даних та проведено тестування (Рисунок 4.26). Результати описані у Таблиці 4.8.

Таблиця 4.8 – Ефективність виявлення типу модифікацій

Кількість зображень	Час навчання (с)	Accuracy	Precision	Recall	F1 score
200	145.2	0.97	0.95	0.95	0.95
400	208.3	0.945	0.90	0.90	0.90
600	268.3	0.95	0.94	0.94	0.94
800	354.7	0.98	0.97	0.97	0.97
1000	416.2	0.97	0.96	0.96	0.96

Результати із таблиці 4.8 зображено на рисунку 4.23.

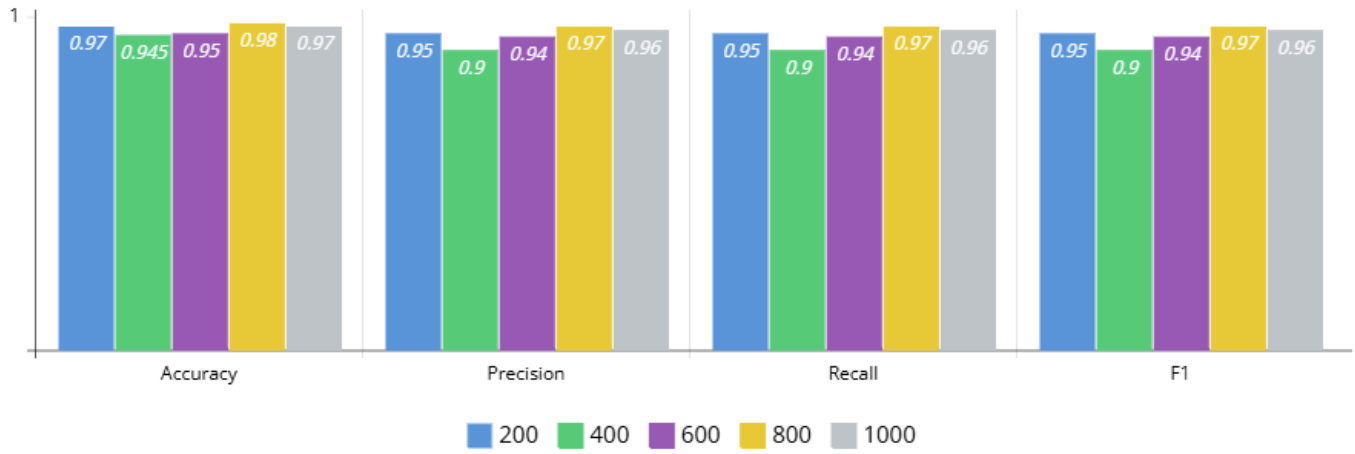


Рисунок 4.26 – Результати виявлення типу модифікації

Проаналізувавши Рисунок 4.23, можна зробити висновок, що найкращі результати у НМ навченій на 800 зразках зображень. Отримані наступні результати accuracy = 0.98, precision = 0.97, recall = 0.97, f1 = 0.97. Час навчання склав 324.7 секунди.

Аналогічно як і з НМ для виявлення наявності модифікації, проведено дослідження впливу гіперпараметрів на результативність (Таблиця 4.9).

Таблиця 4.9 – Дослідження впливу гіперпараметри на результативність НМ

Batch size	Epochs	Час навчання(с)	Accuracy	Precision	Recall	F1
20	16	237	0.95	0.94	0.94	0.94
16	10	172	0.98	0.97	0.97	0.97
12	8	156	0.94	0.943	0.943	0.943

Результати із таблиці 4.9, зображені на рисунку 4.27

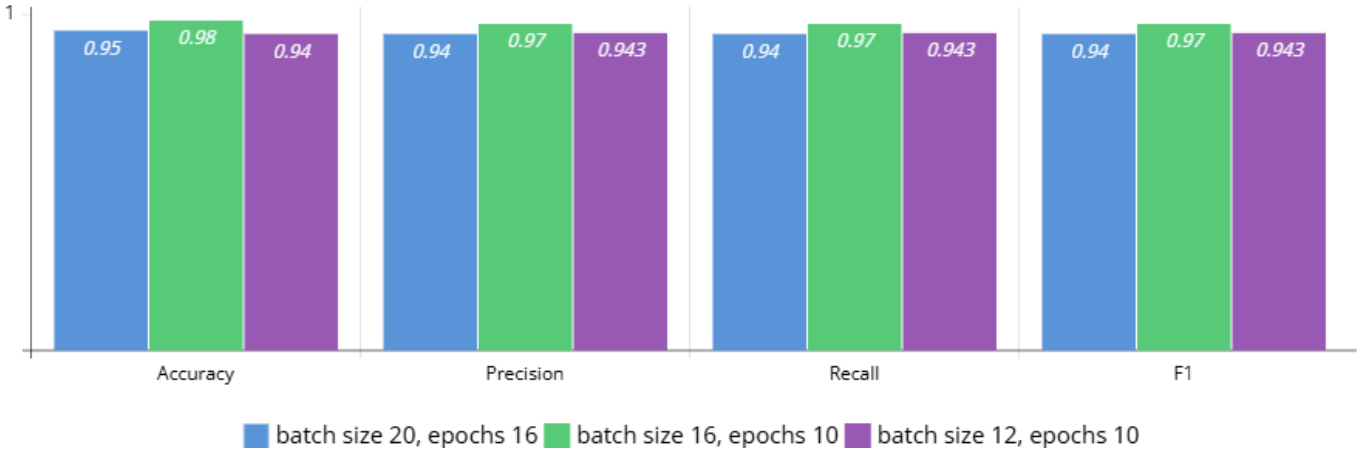


Рисунок 4.27 – Дослідження впливу гіперпараметрів на НМ

Провівши аналіз Рисунка 4.. можна зробити висновок, що оптимально вкзати наступні гіперпараметри: batch size = 16 та epochs = 10. Із даними гіперпараметрами НМ отримала найкращу результативність, а саме: accuracy = 0.98, precision = 0.97, recall = 0.97, f1 = 0.97.

Протестувавши НМ для виявлення наявності модифікацій та НМ для виявлення типу модифікацій, можна зробити висновок, що НМ мають хорошу точність приблизно 0.97 – 0.98. Проте, нейрона мережа для виявлення наявності модифікацій має трішки вищу скорість навчання, через шар Dropout, який випадково відключає половину нейронів (Таблиця 4.10).

Результати таблиці 4.10, зображені на Рисунку 4.28.

Таблиця 4.10 – Порівняння швидкості навчання

Кількість зображень	Час навчання	
	НМ ідентифікації модифікацій	НМ виявлення типу модифікацій
200	125.5	145.2
400	180.8	208.3
600	253.3	268.3
800	324.7	354.7
1000	377.6	416.2

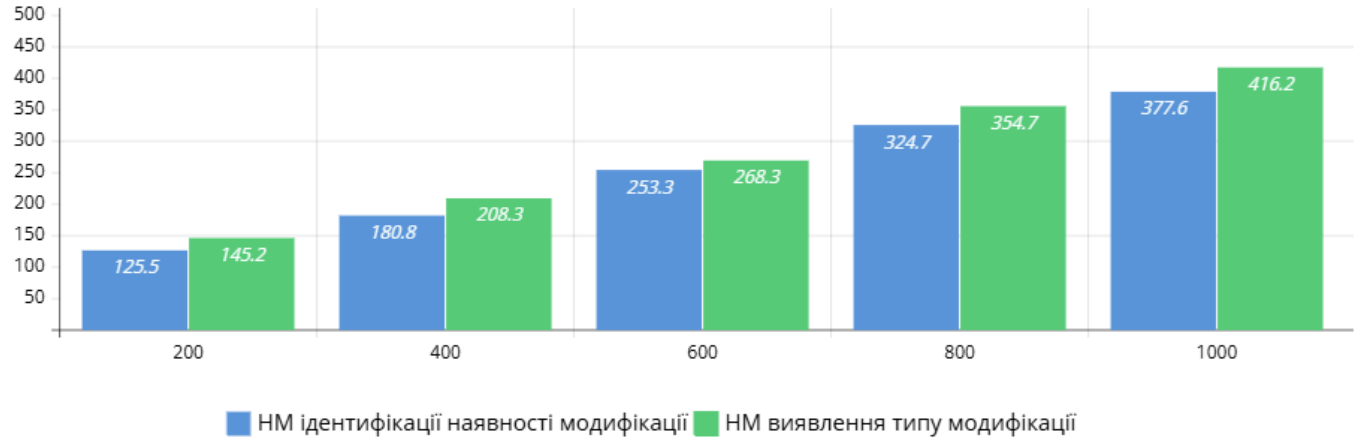


Рисунок 4.28 – Порівняння швидкостей навчання нейромереж

Проаналізувавши Рисунок 4.28, можна зробити висновок, що нейронна мережа для визначення наявності модифікації навчається швидше ніж НМ для визначення типу модифікації приблизно на 10%. Це досить важливий аспект, оскільки на великій кількості даних, різниця у часі навчання може бути значно більша.

Отже у даному розділі були протестовані дві НМ для ідентифікації та виявлення типу модифікації, які застосовуються у інфомраційній системі методу виявлення модифікованих зображень облич людей. Обрано оптимальні гіперпараметри та кількість вхідних даних при яких отримується найвища точність (приблизно 98%). За рахунок шару який відключає половину нейронів НМ для

ідентифікації модифікацій навчається приблизно на 10% швидше, проте точність досягнута практично однакова.

Висновки до розділу 4

В даному розділі було розглянуто структуру методу виявлення модифікованих зображень облич людей, описано їх значення та взаємодію. Також описані додаткові компоненти, які використовуються у ІС.

Описано функціональні можливості інформаційної системи та створено інструкцію користувача, на якій поетапно зображено варіативність дій та можливості системи.

Розписано особливості програмної реалізації, а саме процес навчання, донавчання нейронних мереж. Завдяки даному функціоналу, була вирішена проблема із нехваткою обчислювальних ресурсів.

Описано функціонал, який вирішує проблему із перенавчанням завдяки стоп-функції, яка зупиняє процес навчання досягнувши бажаної точності.

Створено тест-кейси, в яких перевірено працездатність основних функцій, а саме: навчання нейромережі; використання вже готової нейромережевої моделі; завантаження попередньо навченої моделі. Всі тест-кейси пройдено успішно.

Також було проведено декілька досліджень ефективності нейронних мереж, які застосовуються у інформаційній системі. Найкращі результати отримано при навчанні за допомогою датасету розмірністю 800 екземплярів зображень.

Для нейронної мережі визначення наявності модифікацій отримано наступні метрики: accuracy = 0.99, precision = 0.98, recall = 0,98, f1 = 0,98.

Під час дослідження ефективності нейронної мережі виявлення типу модифікацій отримані наступні результати: accuracy = 0.98, precision = 0.97, recall = 0,97, f1 = 0,97. Отриманні значення за метриками, показують спроможність розробленого методу виявлення модифікованих зображень облич людей до вирішення науково-практичної задачі.

Загальні висновки

Як результат кваліфікаційної роботи магістра, розроблено метод виявлення модифікованих зображень облич людей, та інформаційну систему, яка використовує даний метод. Прикладне дослідження встановило, що з використанням розробленого методу досягається підвищення точності виявлення модифікованих фотографій облич людей, що визначає успішне досягнення мети кваліфікаційної роботи магістра.

В ході виконання, досліджено предметну область виявлення модифікованих зображень облич людей та її актуальність. Розглянуто варіативність модифікацій та способи їх створення.

Проаналізовано та проведено тестування вже готового, аналогічного програмного забезпечення, виділено основні переваги та недоліки а також досліджено ефективність їх роботи. Розглянуто проблеми та їх рішення із якими зіткнулись розробники, під час розробки методів для виявлення модифікацій.

Досліджено особливість застосування згорткових нейронних мереж, які будуть обрані у якості нейромережевих моделей для виявлення модифікацій. Також розглянуто базову архітектуру нейромережі та створену власну, для розробки методу виявлення модифікованих зображень облич людей.

Під час виконання кваліфікаційної роботи, досліджено вже готові набори даних та розроблено власний, який містить приблизно 1000 зображень. Даний набір розподіллено на класи та структуровано. Всі фото створеного датасету були взяті із вже готових наборів даних, знайдених у відкритому доступі.

Розроблено схему методу виявлення модифікованих зображень облич людей та поетапно розписано кроки, які будуть виконані в процесі роботи даного методу від завантаження зображення до отримання результату.

В процесі виконання роботи було обрано вже навчену модель blazeface, яка виконує роль посередника та визначає чи містить зображення обличчя. Розглянуто варіативність результату роботи даної моделі.

Кінцевий результат кваліфікаційної роботи магістра – інформаційна система, яка використовує розроблений метод виявлення модифікованих зображень облич людей. Даний метод показав точність за метриками, приблизно 97% при навчанні на створеному датасеті. Для інформаційної системи, розроблено інструкцію користувача та проведено тестування найважливіших функцій.

Таким чином, в процесі виконання роботи створено структурований датасет та розроблено новий метод виявлення модифікованих зображень облич людей, що дозволяє виявляти не лише наявність модифікації зображення обличчя, а і спосіб її походження. Запропонований метод реалізовано у вигляді інформаційної системи, апробація якої підтвердила зростання точності ідентифікації модифікації облич на зображеннях, при цьому метрики становили: Accuracy 0.99, Precision 0.98, Recall 0.98, F1 0.98, що є підвищенням на понад 0.04 для кожної метрики.

За темою кваліфікаційної роботи магістра автором виконано три наукові публікації. Основні наукові й практичні результати роботи доповідались у доповіді «Method for Neural Network Detecting Changed Images of People's Faces Using CNN» на I Міжнародній науково-практичній конференції «New Horizons in Scientific Research: Challenges and Solutions» (Marseille, France) 21-23 жовтня 2024 року та у доповіді «Підхід до формування датасету для нейромережевого виявлення модифікованих фотографій облич людей» на XVI Всеукраїнській науково-практичній конференції «Актуальні проблеми комп’ютерних наук» (м.Хмельницький) 15-16 листопада 2024 року [49, 50].

Перелік посилань

1. Новинний сайт scoop.upworthy.com. URL: https://scoop.upworthy.com/model-starts-the-filter-drop-challenge-on-instagram-to-fight-against-retouched-beauty-ads

2. Rana M. S., Nobi M. N., Murali B., Sung A. H. Deepfake detection: A systematic literature review. IEEE access, 10, 2022 pp. 25494-25513.

3. Trymaverick. Сайт для створення фейкових фото та відео trymaveric.com. URL: https://www.trymaverick.com/blog-posts/are-deep-fakes-all-evil-when-can-they-be-used-for-good

4. Siluette. Освітня веб платформа siluette.com.ua. URL: https://siluette.com.ua/en/product/morfing/

5. Schardong G., Novello T. Paz, I. da Silva, V. Velho L., Gonçalves, N. Neural implicit morphing of face images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024. pp. 7321-7330.

6. Boopathi S., Pandey B. K., Pandey D. Advances in artificial intelligence for image processing: techniques, applications, and optimization. In Handbook of research on thrust technologies' effect on image processing, 2023. pp. 73-95.

7. IGI Global. Офіційний сайт Adobe. URL: https://www.adobe.com/ua/products/photoshop.html

8. Веб-сервіс для коригування фото fixthephoto.com. URL: https://fixthephoto.com

9. Taye M. M. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. Computation, 11(3), 2023, p.52.

10. IBM. What is a neural network? URL: https://www.ibm.com/topics/neural-networks

11. Unite. Наукова база даних unite.ai. URL: https://www.unite.ai/uk/generative-vs-discriminative-machine-learning-models

12. Datascientest. Convultaion neural network everything you need to Know. URL: https://datascientest.com/en/convolutional-neural-network-everything-you-need-to-know

13. Dou. Використання CNN для оброки зображень. URL - https://dou.ua/forums/topic/48368

14. Roboflow. What is a Convulation Neural Network?. URL: https://blog.roboflow.com/what-is-a-convolutional-neural-network

15. Medium. A practical Guide to ReLU. URL: https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7

16. Eremio. Pooling Layers. URL: https://www.dremio.com/wiki/pooling-layers

17. Paperswithcode. Max Pooling. URL: https://paperswithcode.com/method/max-pooling

18. Geeksforgeeks. CNN, intorduction to Pooling Layer. URL: https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/

19. Umadevi M., Krishna S., Kumar N., «Deep Fake Face Detection using Efficient Convolutional Neural Networks,» 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal, 2024, pp. 344-352, doi: 10.1109/ICIPCN63822.2024.00063.

20. Deepfake detection using deep learning methods: A systematic and comprehensive review / A. Heidari et al. WIREs Data Mining and Knowledge Discovery. 2023. URL: https://doi.org/10.1002/widm.1520 (date of access: 16.12.2024).

21. Chauhan R., Sethi M., Ahuja S., «Fake Faces Unveiled: A Comprehensive Study on Detecting Generated Facial Images», 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 475-482, doi: 10.1109/AUTOCOM60220.2024.10486084.

22. Singh K.R., Zaveri Mukesh, Raghuwanshi M.M. Illumination and Pose Invariant Face Recognition: A Technical Review, 2020, pp. 30-31.

23. Remone R., Dash S. Face Recognition and Face Detection Benefits and Challenges Section A-Research paper 2561 Eur. European Chemical Bulletin. 12. 2023, pp. 2561-2566.

24. Gunasekaran K., Jaiman N. . Now you see me: Robust approach to partial occlusions. In 2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning, PRML, 2023, pp. 168-175.

25. Zeng D., Veldhuis R., Spreeuwers L. A survey of face recognition techniques under occlusion. IET biometrics, 10(6), 2021, pp. 581-606.

26. S. Hemathilaka, A. Apono. A Comprehensive study on occlusion invariant face recognition under face mask. 10.48550/arXiv.2201.09089. 2022, pp. 8-9.

27. Salagean, Gabriela & Leba, Monica Face Recognition: A Literature Review. 10.31410/ITEMA.2023.55, 2023, pp. 55-60.

28. Kudubayeva S., Razakhova B.. Modern Problems of Face Recognition Systems and Ways of Solving Them. Revue d'Intelligence Artificielle. 37. 2023, pp. 209-214.

29. Imageedited. Головна сторінка imageedited. URL: http://imageedited.com/

30. Fakeimagedetector. Головна сторінка. URL: https://www.fakeimagedetector.com/

31. Databricks. What is dataset? URL: https://www.databricks.com/glossary/what-is-dataset

32. Idiap. FRLL-morphs. URL: https://www.idiap.ch/en/scientific-research/data/frll-morphs

33. Kaggle. Deepfake_faces. URL: https://www.kaggle.com/datasets/dagnelies/deepfake-faces

34. Github. BlazeFace. URL: https://github.com/hollance/BlazeFace-PyTorch

35. Tensorflow. Офіційний сайт tensorflow. URL: https://www.tensorflow.org/?hl=en

36. Jeremyjordan. Normalizing your data. URL: https://www.jeremyjordan.me/batch-normalization/

37. Medium. Image augmentation for convukation neural networks. URL: https://odsc.medium.com/image-augmentation-for-convolutional-neural-networks-18319e1291c

38. Towardsdatascience. Deciding optimal kernel size for CNN. URL: https://towardsdatascience.com/deciding-optimal-filter-size-for-cnns-d6f7b56f9363

39. Geeksforgeeks. CNN. Introduction to padding. URL: https://www.geeksforgeeks.org/cnn-introduction-to-padding/

40. Geeksforgeeks. How to choose Batch Size and Number of Epochs when fitting a model. URL: https://www.geeksforgeeks.org/how-to-choose-batch-size-and-number-of-epochs-when-fitting-a-model/

41. Towardsdatascience. Various optimization algorithms for training neural network. URL: https://towardsdatascience.com/optimizers-for-training-neural-network-59450d71caf6

42. Digitalocean. Evaluation deep learning models, accuracy. URL: https://www.digitalocean.com/community/tutorials/deep-learning-metrics-precision-recall-accuracy

43. Educative. Precision in neural networks. URL: https://www.educative.io/answers/precision-vs-recall-vs-accuracy-in-neural-networks

44. Machinelearningmastery. How to calculate recall. URL: https://machinelearningmastery.com/how-to-calculate-precision-recall-f1-and-more-for-deep-learning-models/

45. V7labs. F1 score in machine learning: into calculation. URL: https://www.v7labs.com/blog/f1-score-guide

46. W3schoolsua. HTML Підручкни. URL: https://w3schoolsua.github.io/html/index.html#gsc.tab=0

47. CSS Підручник. URL: https://w3schoolsua.github.io/css/index.html#gsc.tab=0

48. Visualstudio. Офіційний сайт Visual studio code. URL: https://code.visualstudio.com/

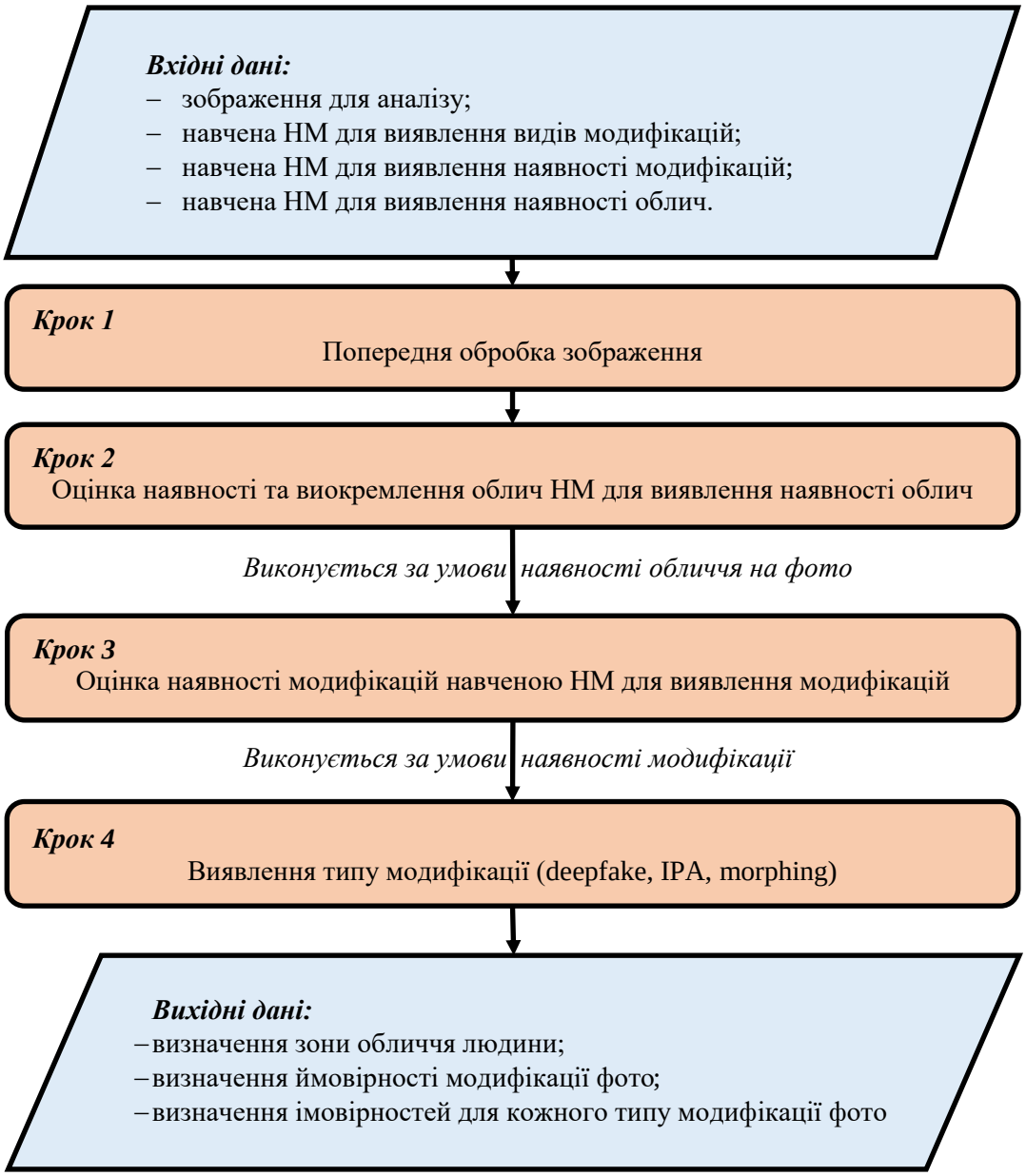
49. Pokhytun A., Mazurets O., Molchanova M., Tyschenko O. Method for Neural Network Detecting Changed Images of People's Faces Using CNN. New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International scientific and practical conference. October 21-23, 2024. Marseille, France. 2024. Pp. 35-40.

50. Похитун А.В., Мазурець О.В., Молчанова М.О., Бармак О.В. Підхід до формування датасету для нейромережевого виявлення модифікованих фотографій облич людей. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп’ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 428-433.

ДОДАТКИ

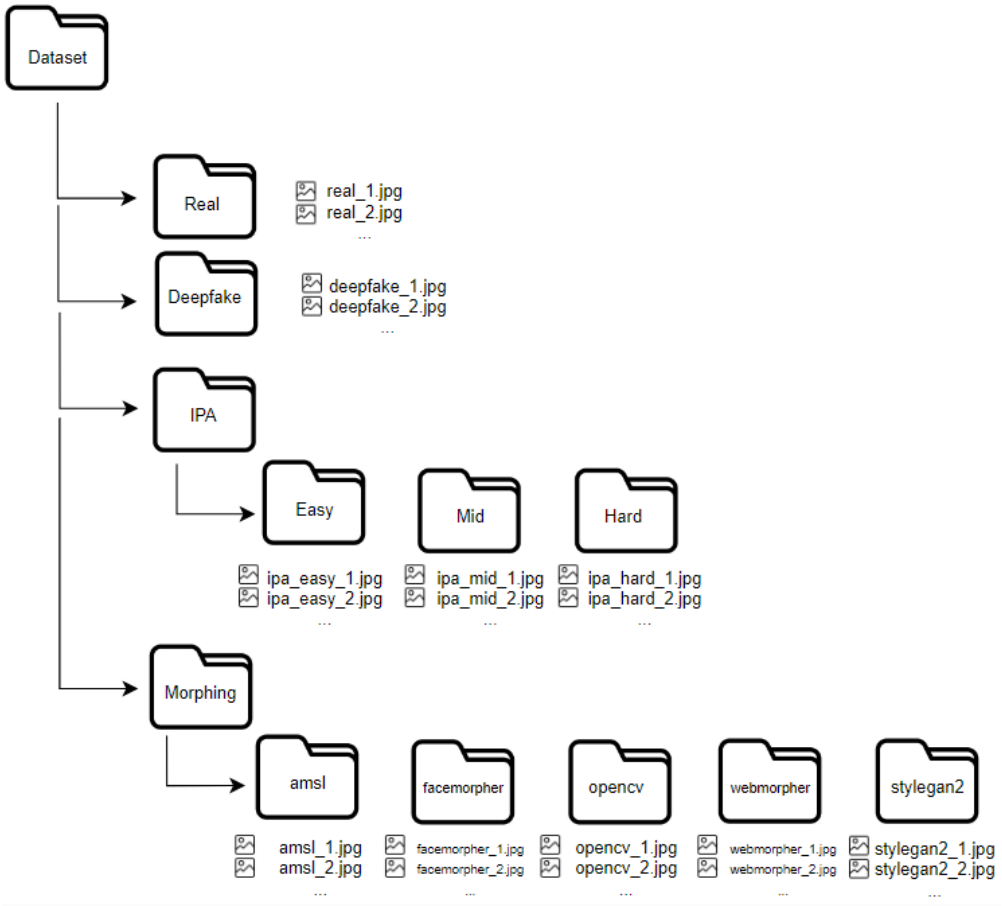
Додаток А

Схема методу виявлення модифікованих зображень облич людей



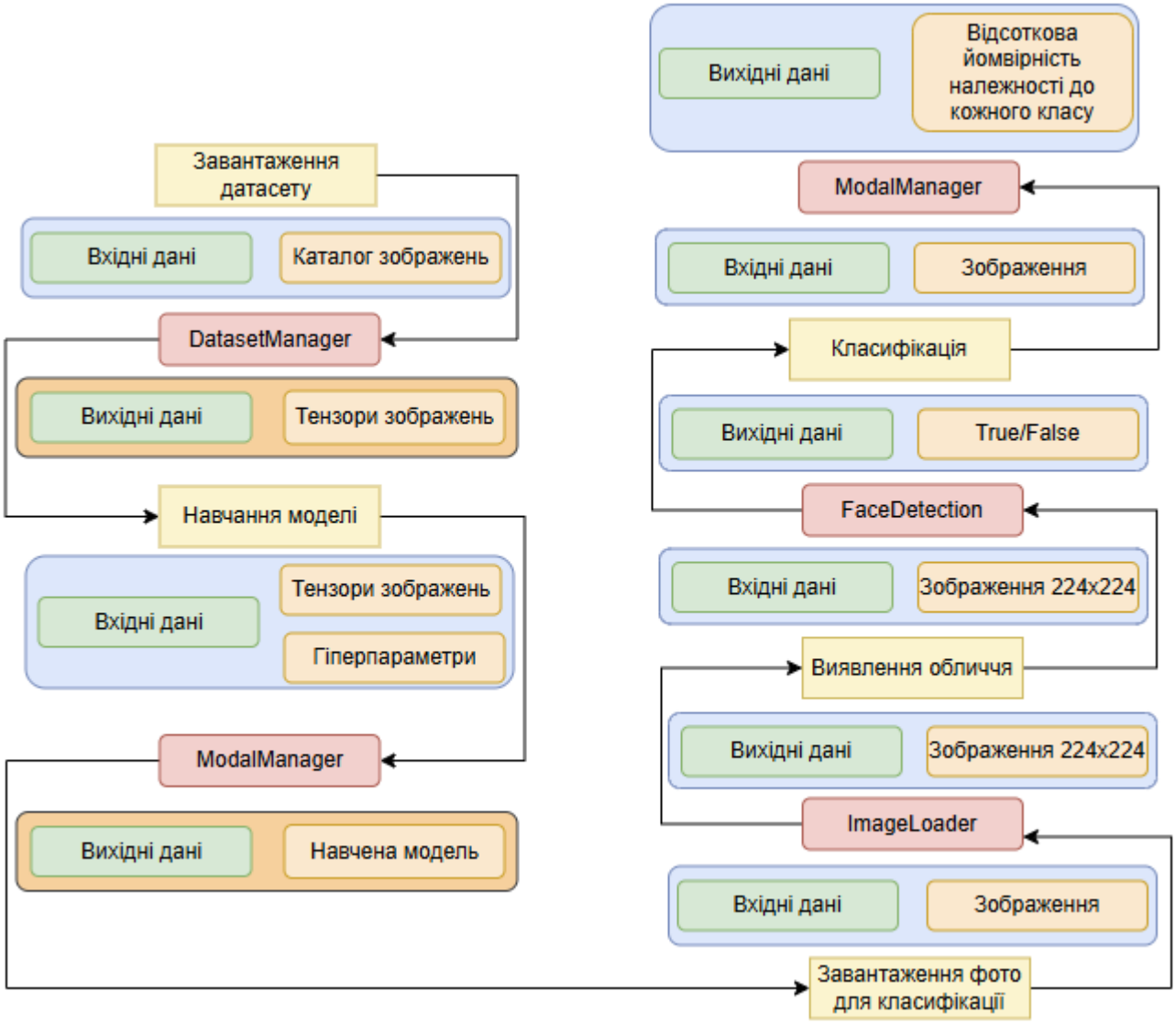
Додаток Б

Схема структури датасету для виявлення модифікованих зображень облич людей



Додаток В

Схема взаємодії компонентів методу



Додаток Г

Світлини наукових публікацій, виконаних при роботі над кваліфікаційною роботою магістра

Наукові публікації:

1. Pokhytun A., Mazurets O., Molchanova M., Tyschenko O. Method for Neural Network Detecting Changed Images of People's Faces Using CNN. New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International scientific and practical conference. October 21-23, 2024. Marseille, France. 2024. Pp. 35-40.
2. Похитун А.В., Мазурець О.В., Молчанова М.О., Бармак О.В. Підхід до формування датасету для нейромережевого виявлення модифікованих фотографій облич людей. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп’ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 428-433.
3. Похитун А.В., Мазурець О.В., Дидо Р.А., Молчанова М.О. Програмна архітектура для нейромережевого виявлення модифікованих фотографій облич людей. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2025. №1 (Довідка з редакції).


eoss-conf.com
ISSUE
№7
EUROPEAN OPEN
SCIENCE SPACE
COLLECTION OF SCIENTIFIC PAPERS
1st INTERNATIONAL
SCIENTIFIC
AND PRACTICAL
CONFERENCE
NEW HORIZONS
IN SCIENTIFIC RESEARCH:
CHALLENGES
AND SOLUTIONS
OCTOBER 21-23, 2024, MARSEILLE, FRANCE

EUROPEAN OPEN SCIENCE SPACE

CONTENT

що сприяє ідентифікації ключових вузлів та потенційних джерел загроз. Автоматизація збору даних та аналізу інформаційних потоків стає критично важливим для попередження дезінформації та координаційних атак. Результати дослідження дозволяють сформувати рекомендації щодо покращення моніторингу й захисту інформаційного середовища, спрямовані на підвищення рівня кібербезпеки.

Список використаних джерел

1. О. Борсукова. (2022). Які повідомлення у соцмережах шкідливі під час війни? Українська правда. https://life.pravda.com.ua/society/2022/03/09/247739/
2. О. Мороз. (2022). Як відрізнити шкідливу інформацію від корисної? Громадське радіо. https://hromadske.radio/podcasts/freeourfamily/1061123
3. Ay N., Polani D. (2008). Information flows in causal networks. Advances in Complex Systems, 11(01), 17–41. https://doi.org/10.1142/s0219525908001465
4. Lyon W., Lesica G. Information Flow Through a Network. Neo4j. https://neo4j.com/graphgists/information-flow-through-a-network/
5. Telethon's Documentation. Telethon. https://docs.telethon.dev/en/stable/.
6. C.Brew. (2024). From content to harm: how harmful information contributes to civilian harm. ICRC Humanitarian Law & Policy blog. https://blogs.icrc.org/law-and-policy/2024/02/27/from-content-to-harm-how-harmful-information-contributes-to-civilian-harm/

METHOD FOR NEURAL NETWORK DETECTING CHANGED IMAGES OF PEOPLE'S FACES USING CNN

Pokhytun Andrii
Postgraduate student
Mazurets Oleksandr
Ph.D in Engineering Science, Associate Professor
Molchanova Maryna
Teacher
Tyschenko Olena
Teacher
Computer Science Department
Khmelnytskyi National University, Ukraine

In today's world, changing your appearance based on a photo is not a difficult task. One of the most influential factors is the emergence of generative models that are able to create fairly realistic images of people's faces, or change them. Generative models are types of machine learning that can generate new data similar to those on which the model was trained [1]. It can be not only images, but also photos, texts, videos, etc.

Proceedings of the 1st International Scientific and Practical Conference
"New Horizons in Scientific Research: Challenges and Solutions"
October 21-23, 2024
Marseille, France
EUROPEAN OPEN SCIENCE SPACE

Among the large number of types of modifications, we can single out a few that are most popular [2]:

– digital filters and cosmetic changes;
– deep fakes;
– morphing;
– image processing algorithm.

Digital filters are one of the most common types of photo modifications in modern society. This type of photo modification gained its popularity due to the increase in the number of users of social networks. With the help of filters, users could change their appearance in just a few clicks on the screen.

Deep fakes are a technology that allows you to change the face in a photo or video using a neural network. Algorithms used in this technology allow not only to create new images, but also can superimpose images on images, which allows replacing only the face of a person in the entire photo.

Morphing is a technology that allows you to create fake images using a smooth transition from one photo to another. Morphing can be used by attackers to forge identity documents.

Image processing algorithms are algorithms that include simpler methods of changing images using software [3]. One of the most famous examples of software is the well-known Adobe Photoshop. Adobe Photoshop is a powerful graphic editor that is usually used to work with bitmap images.

Image processing algorithms are also often used in cinematography. With the help of various software, you can change facial features, skin texture, skin color, hair color and much more [4, 5].

In the modern world, there are quite a lot of opportunities for adjusting, editing and processing images [6, 7], so it is important to be able to distinguish fake images from real ones, because modified photos can be used not only for fun, but also for forging documents, harassment, which can have quite negative consequences. consequences

The purpose of the work is to develop a method for detecting modified images of people's faces using neural network technologies, which is designed to transform input data in the form of a dataset of modified images of people's faces and working images for classification into output data with a conclusion or modified photo, as well as the types of modifications used.

A method for detecting altered human face images using neural network technologies is designed to process the input data, which is represented by a training set of face images for the training stage, a test set of images for verification, and a working face image for classification. The method includes the steps of training and evaluating the neural network using the training set of face images, as well as the process of detecting changes in the working face image.

The input data of the modified human face image detection method is a training sample of face images and a working image for classification (Figure 1).

Neural network training takes place with the help of a prepared dataset, which includes different types of modifications and different levels of complexity of modifications.

Neural network testing takes place by sampling images of various types and complexity of modifications.

Detecting the type of modifications is that the method returns not only the result or a modified photo or not, but that it returns a specific type of modifications.

The process of classifying a working image of faces consists in the fact that the user loads an image of a face into a previously trained model, after which he receives a certain result.

The output data of the method of detecting modified images of human faces is whether the result is a modified photo or not, and also, if the photo is modified, the type of modification.

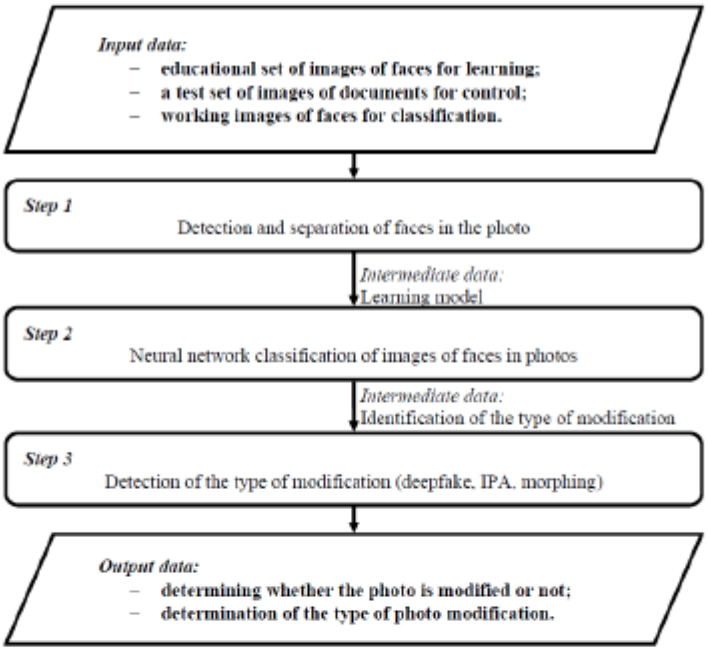


Figure 1. Scheme of the method of detection of modified images of human faces

So, it was described how the method should work for detecting modifications of images of people's faces by means of neural networks. The method is designed to transform the input data in the form of a dataset of modified human face images and work images for classification into output data with the conclusion whether the photo is modified, as well as the types of modifications used.

Convolutional neural networks are the basis for image processing and analysis in current computer vision tasks. Their popularity is explained by the ability of ANNs to extract significant features from images without the need to do it manually. The

tensorflow.js library already has several ready-made basic functions that only need to be modified for a specific task (Figure 2).

The first, the input layer (input) is the layer responsible for the size of the input image, where 128x128 is the size of the image, and 3 is the number of channels (RGB).

The input layer is followed by several convolutional and several pooling layers. Convolutional layers apply 32x32 filters to the input image to extract local features. Each filter generates a feature map that stores spatial information.

After each convolutional layer, it is desirable to apply an activation function in order to add nonlinearity to the model [8].

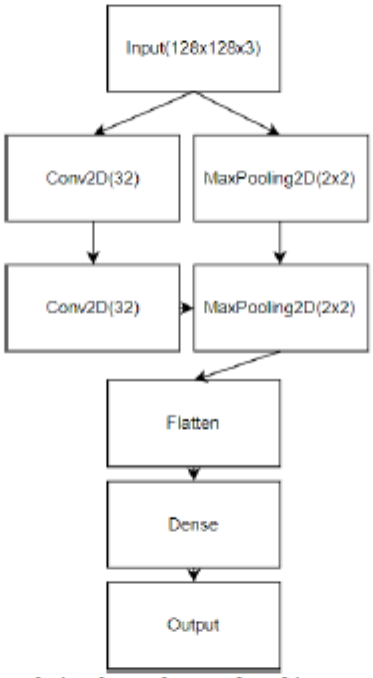


Figure 2. Basic convolutional neural network architecture using tensorflow.js

Pooling layers (maxPooling(2x2)) are applied after the convolutional layers. These layers reduce the feature map by half, which allows for dimensionality reduction and removal of insignificant features, increasing computational efficiency.

The flatten function is a function that converts a multidimensional output into a vector for further use.

After converting the multidimensional array into a vector, the dense function is applied, which processes and extracts features from the image. And pass the result to the output function that outputs the result.

Therefore, a method for detecting altered human face images using neural network technologies was proposed to process the input data, which is represented by a training set of face images for the training stage, a test set of images for verification, and a working face image for classification. The method includes the steps of training and evaluating the neural network using the training set of face images, as well as the process of detecting changes in the working face image. The created method for detecting modified images of human faces is designed to transform the input data in the form of a dataset of modified images of human faces and working images for classification into output data with the conclusion whether the photo is modified, as well as the types of modifications used.

References

1. Zharnovskyi O., Sobko O., Klimenko V. Intelligent System for Neural Network Detection of Fake Document Images for Automated Personality Identification. Proceedings of IV International Scientific and Practical Conference «Innovative research and perspectives of the development of science and technology». January 29-31, 2024. Stockholm, Sweden. 2024. Pp. 337-343.
2. A site for creating fake photos and videos trymaveric.com: URL: https://www.trymaverick.com/blog-posts/are-deep-fakes-all-evil-when-can-they-be-used-for-good
3. Novak Y., Mazurets O. Practical Application of Method of Automated Personal Identification by Fingerprints Using Convolution Neural Networks. Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». December 27-29, 2023. Stockholm, Sweden, International Scientific Unity. 2023. Pp. 136-140.
4. Mazurets O., Sobko O., Vit R., Pasternak V. Practical Approach for Detection by Deep Learning of Target Objects of Subject Area Based on Semantic Connectivity Indicators in Audio Database. Proceedings of XXIV International Scientific and Practical Conference «Modern Scientific Challenges are the Driving Force of the Development of Scientific Research». May 22-24, 2024. Bruges, Belgium. International Scientific Unity. 2024. Pp. 91-96.
5. Mazurets O., Zalutska O., Tyschenko O., Bohdanova A. An Approach to Using MobileNet CNN-model for Gesture Recognition. Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions». May 15-17, 2024. Munich, Germany. 2024. Pp. 59-64.
6. Mazurets O. V., Klimenko V. I., Molchanova M. O., Sultanov A. V. Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones. Global Science: Prospects and Innovations. Proceedings of the 10th International scientific and practical conference. Cognum Publishing House. Liverpool, United Kingdom. 2024. Pp. 198-207. Веб-сервіс для коригування фото fixthephoto.com: URL: https://fixthephoto.com
7. Mazurets O., Molchanova M., Klimenko V., Klopotivskyi D. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services.

Proceedings of XXII International Scientific and Practical Conference «Modern Scientific Research: Theoretical and Practical Aspects». May 8-10, 2024. Oslo, Norway. 2024. Pp. 64-68.
8. Molchanova M., Mazurets O., Klimenko V., Kuflevsky Ev. Object-oriented model for neural network damage detection of mail packages. Proceedings of XIV International Scientific and Practical Conference «Solving Scientific Problems Using Innovative Concepts». March 13-15, 2024. Copenhagen, Denmark. 2024. Pp. 58-62.

ДОСЛІДЖЕННЯ МЕТОДІВ КЛАСТЕРИЗАЦІЇ ПРИ СТВОРЕННІ ОНЛАЙН-БІБЛІОТЕК

Щербакова Галина
д.т.н, професор
Манікаєва Ольга
к.т.н, старший викладач
Нікова Діана
здобувач вищої освіти магістерського рівня
Фурман Данило
здобувач вищої освіти магістерського рівня
Кафедра інформаційних систем
Національний університет «Одеська політехніка», Україна

В останні роки стало актуальним створення та використання інформаційних систем на основі фільтрування. Таке фільтрування спрямоване на виявлення серед великих обсягів даних – даних саме для певного користувача, тому що найкращим чином відповідають його потребам. З таких систем популярністю користуються рекомендаційні системи. Тому і прикладів таких систем можна навести багато [1]. Серед них можна назвати і системи книгарень та онлайн-бібліотек, які спрямовані на надання користувачу рекомендацій щодо вибору книг, методичних матеріалів, підручників.

В системі користувач має залишити інформацію про вподобання тим чи іншим чином, система фільтрує та зберігає її. Такі системи основані на припущенні, що люди зі схожими інтересами можуть мати схожі вподобання. Аналізуючи дані про читачів та їх вподобання (історію переглядів, рейтинги книг), ці системи мають визначити кількісні характеристики схожості між користувачами та/або обраною літературою. Ці дані використовуються алгоритмами машинного навчання для передбачення успішних рекомендацій читачам [2].

Тобто важливим етапом обробки даних в таких системах є саме визначення складу груп читачів, схожих по певним параметрам та/або вподобанням. Саме для цього використовуються алгоритми кластеризації чи навчання без вчителя (Unsupervised learning). Таких алгоритмів існує значна кількість. Вони діляться

Міністерство освіти і науки України
Хмельницький національний університет

ЗБІРНИК НАУКОВИХ ПРАЦЬ

за матеріалами XVI Всеукраїнської науково-практичної конференції
«Актуальні проблеми комп'ютерних наук АПКН-2024»

15-16 листопада 2024

Хмельницький 2024

УДК 004.8

Похитун А.В., Мазурець О.В., Молчанова М.О., Бармак О.В.

Хмельницький національний університет

ПІДХІД ДО ФОРМУВАННЯ ДАТАСЕТУ ДЛЯ НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ МОДИФІКОВАНИХ ФОТОГРАФІЙ ОБЛИЧ ЛЮДЕЙ

У роботі розроблено та програмно реалізовано метод виявлення модифікованих зображень облич людей. Метод розроблений для перетворення вхідних даних у вигляді датасету у вихідні дані, а саме тип, клас, складність та ймовірність модифікації. Також було проаналізовано готові датасети та створено власний датасет, що містить 4 класи, кожен із класів містить 200 зображень різної розмірності, а зведення до одного розміру реалізується програмно.

The method of detecting modified images of people's faces was developed and implemented in software. The method is designed to transform input data in the form of a dataset into output data, namely type, class, complexity, and modification probability. Ready-made datasets were also analyzed and a custom dataset containing 4 classes was created, each of the classes contains 200 images of different dimensions, and reduction to one size is implemented programmatically.

У сучасному світі змінити зовнішній вигляд по фото не є важкою задачею. Одним із найбільш впливових факторів є поява генеративних моделей які здатні створювати досить реалістичні зображення облич людей, або змінювати їх.

Нейронна мережа – це застосунок, або обчислювальна модель машинного навчання, яка приймає рішення подібно до людського мозку та складається з великої кількості взаємопов'язаних елементів – нейронів [1]. Нейронні мережі використовуються для розв'язання різноманітних завдань, зокрема в обробці зображень, розпізнавання мови, тексту та у багатьох інших сферах [2, 3].

Генеративні моделі – це види машинного навчання, що можуть генерувати нові данні подібні до тих, на яких була навчена модель [4]. Це можуть бути не лише зображення а й фото, тексти, відео тощо.

Серед великої кількості видів модифікацій, можна виділити декілька, які користуються найбільшою популярністю

- цифрові фільтри та косметичні зміни;
- глибинні фейки;
- морфінг;
- алгоритм обробки зображень.

Отож, існує досить багато можливостей для коригування, редагування та обробки зображень, тому важливо вміти розрізняти фейкові зображення від

реальних оскільки модифіковані фото можуть використовуватись не лише для забави а й для підробки документів, цькування, що може мати досить негативні наслідки.

Мета роботи полягає у формуванні датасету та створенні методу для виявлення модифікованих зображень облич людей засобами нейромережевої класифікації.

Для розробки методу виявлення модифікованих зображень облич людей, критично важливо мати добре структуровані вхідні дані. В глобальній мережі доступна досить велика кількість датасетів, проте не всі із них підходять для цієї задачі. Під час розробки методу виявлення модифікованих зображень облич людей, у відкритому доступі, було знайдено декілька датасетів з яких було сформовано один, структурований набір даних із різними класами.

Зокрема для створення набору даних, були використані, вже готові, наступні датасети:

- FRLL-morphs (200 зображень);
- deepfake_faces (200 зображень);
- набір даних розроблений департаментом комп'ютерних наук університету Йонсей (200 зображень).

FRLL-morphs – набір структурованих даних, сформованих на основі даних взятих із Face Research London Lab [5]. Однією із ключових переваг даного датасету є те, що модифіковані зображення створювались різними методами (Рисунок 1).

Рисунок 1 – Приклад модифікованих зображень різними алгоритмами [5]

На Рисунку 2 видно, що алгоритми AMSL та StyleGan2 досить важко розпізнати модифікацію, а ПЗ FaceMorpher та алгоритм OpenCV мають характерне розмиття на фоні.

Deepfake_faces - набір даних, розмірністю 224 на 224 пікселі, що є переваго, оскільки саме на таку розмірність існує досить багато навчених моделей [6]. На деяких зразках даного датасету присутні шуми (Рисунок 2).

Рисунок 2 – Приклад зображень датасету depfake_faces [6]

Набір даних розроблений департаментом комп'ютерних наук університет Йонсей містить декілька рівнів модифікацій, від простіших – відразу помітних, до більш складних [7], модифікації на яких важко помітити (Рисунок 3).

Рисунок 3 – Приклад зображень, створених департаментом комп'ютерних наук університету Йонсей [7]

На основі вищеописаних датасетів, було створено власний датасет із різними та рівнями модифікацій (Рисунок 4).

Отже, було проаналізовано готові набори даних, та створено власний датасет. Створений набір даних містить наступні класи, підкласи:

- deepfake;
- real;
- ipa (easy, mid, hard);
- morphing (amsl, facemorfer, opencv, webmorpher, stylefan2).

Кожен із класів містить 200 зображень різної розмірності. Зведення до одного розміру реалізується програмно.

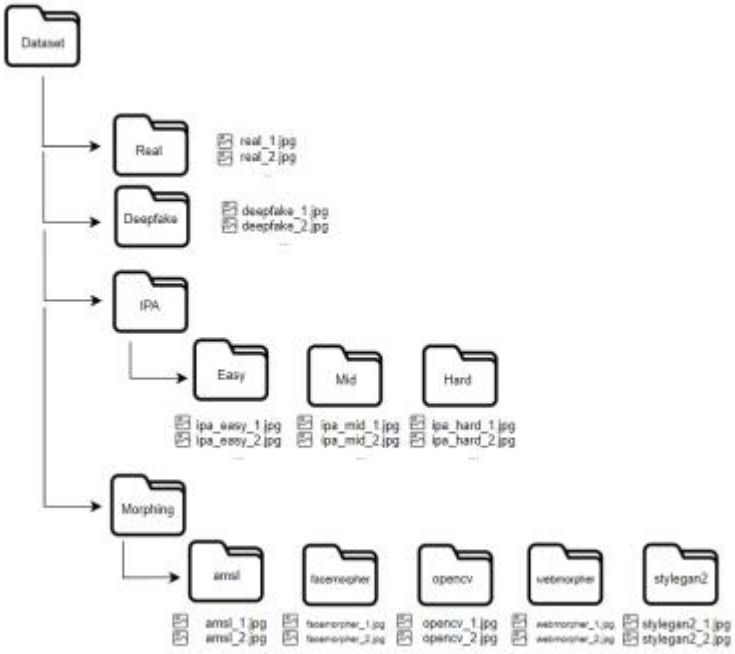


Рисунок 4 – Структура створеного датасету

Метод виявлення модифікованих зображень облич людей призначений для перетворення вхідних даних у вигляді зображення, у вихідні дані у вигляді результату класифікації, а саме тип, складність (якщо це дозволяє датасет) та алгоритм за допомогою якого було модифіковане фото (Рисунок 5).

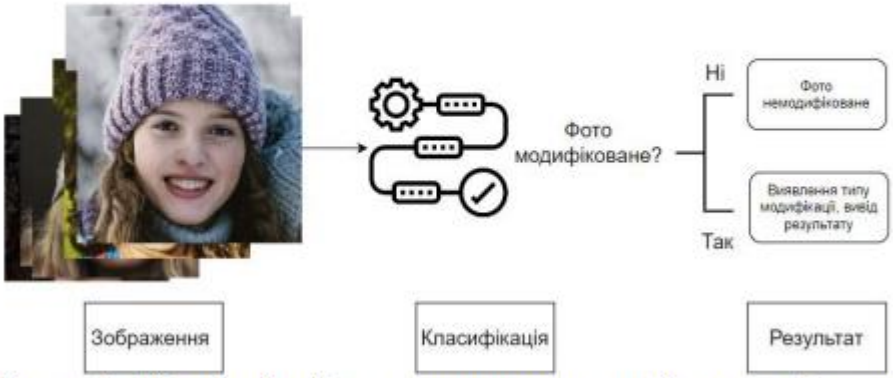


Рисунок 5 – Візуалізація роботи методу визначення модифікованих зображень

Для вхідних даних було створено датасет із приблизно 1500 зображеннями різних класів, а також тестова вибірка зображень для оцінки коректності (Рисунок 6).

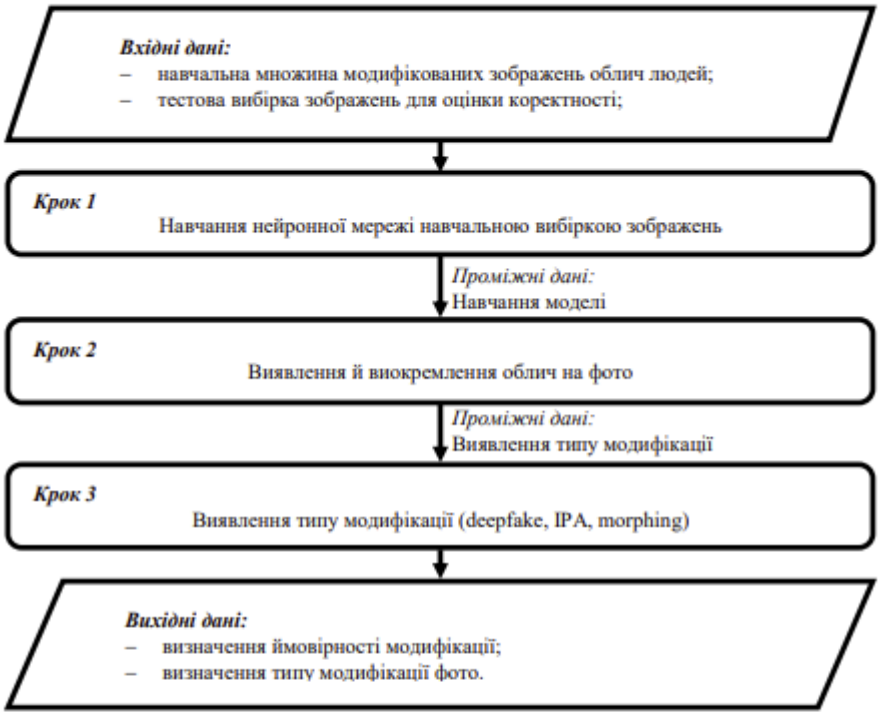


Рисунок 6 – Схема методу виявлення модифікованих зображень облич людей

Процес виявлення модифікованого зображення обличчя можна розділити на декілька етапів: виявлення облич на фото, навчання нейронної мережі за допомогою датасету та класифікація зображення [8, 9].

Першим етапом є навчання нейромережевої моделі, яке відбувається за допомогою навчальної множини зображень (датасету) [10]. Також є можливість завантажити попередньо навчену модель, щоб не витрачати час на повторне навчання [11].

Другий етап – виявлення та виокремлення облич на фото. На даному етапі відбувається перевірка, чи взагалі присутнє обличчя на зображенні і чи є сенс в подальшій перевірці на модифікації.

Завершальний етап, виявлення типу модифікації, складності та алгоритму, за допомогою якого фото було модифіковане (якщо це дозволяє датасет).

Вихідними даними методу виявлення модифікованих зображень облич, є результат у вигляді ймовірності належності зображення до конкретної модифікації.

Отже, було проаналізовано готові набори даних, та створено власний датасет. Створений набір даних містить 4 класи, кожен із класів містить 200 зображень різної розмірності. Зведення до одного розміру реалізується програмно. Також було розроблено метод виявлення модифікованих зображень облич людей. Метод розроблений для перетворення вхідних даних у вигляді датасету у вихідні дані, а саме тип, клас, складність та ймовірність модифікації.

Перелік посилань

1. Mazurets O., Uspenska K., Vit R., Tyschenko O. Intelligent System for Determining the Object Attributes Values by Neural Networks Means by Graphic Images in Databases. Current Trends in the Development of Scientific Research in Today's Conditions. Proceedings of XXV International scientific and practical conference. May 29-31, 2024. International Scientific Unity. Florence, Italy. 2024. Pp. 86-91.
2. Kharysh I., Sobko O., Mazurets O. Designing CNN Neural Network Model for Detecting Fractures of Lower Extremities by X-ray Images. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. October 23-25, 2024. Dubrovnik, Croatia. 2024. Pp. 91-96.
3. Mazurets O. V., Klimenko V. I., Molchanova M. O., Sultanov A. V. Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones. Global Science: Prospects and Innovations. Proceedings of the 10th International scientific and practical conference. Cognum Publishing House. Liverpool, United Kingdom. 2024. Pp. 198-207.
4. Mazurets O., Zalutska O., Tyschenko O., Bohdanova A. An Approach to Using MobileNet CNN-model for Gesture Recognition. Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions». May 15-17, 2024. Munich, Germany. 2024. Pp. 59-64.
5. FRLL-morphs: URL - https://www.idiap.ch/en/scientific-research/data/frll-morphs
6. Deepfake_faces: URL - https://www.kaggle.com/datasets/dagnelies/deepfake-faces
7. BlazeFace: URL - https://github.com/hollance/BlazeFace-PyTorch
8. Mazurets O., Molchanova M., Klimenko V., Klopotivskyi D. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services. Proceedings of XXII International Scientific and Practical Conference «Modern Scientific Research: Theoretical and Practical Aspects». May 8-10, 2024. Oslo, Norway. 2024. Pp. 64-68.
9. Novak Y., Mazurets O. Practical Application of Method of Automated Personal Identification by Fingerprints Using Convolution Neural Networks. Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». December 27-29, 2023. Stockholm, Sweden, International Scientific Unity. 2023. Pp. 136-140.
10. Мазурець О.В., Петровський С.С., Дидо Р.А. Нейромережева модель для ідентифікації особистості за зображенням обличчя у реальному часі Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції. 31 жовтня – 1 листопада 2024 р. Одеса, ОНТУ. 2024. С.655-658.
11. Pokhytun A., Mazurets O., Molchanova M., Tyschenko O. Method for Neural Network Detecting Changed Images of People's Faces Using CNN. New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International scientific and practical conference. October 21-23, 2024. Marseille, France. 2024. Pp. 35-40.

Довідка: ВХНУ ТН 6-12/2024

Видання: Herald of Khmelnytskyi National University. Technical Sciences (Вісник Хмельницького національного університету. Технічні науки)

Категорія фаховості видання: затверджено як наукове фахове видання України, у якому можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії, категорії «Б» (наказ МОН №1643 від 28.12.2019, наказ МОН №409 від 17.03.2020).

Напрям – технічні науки за спеціальностями – 101, 121, 122, 123, 124, 125, 141, 151, 161, 172, 181, 182 (28.12.2019), спеціальності – 131, 132, 133 (17.03.2020).

Назва статті: ПРОГРАМНА АРХІТЕКТУРА ДЛЯ НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ МОДИФІКОВАНИХ ФОТОГРАФІЙ ОБЛИЧ ЛЮДЕЙ

Автори: Похитун А.В., Мазурець О.В., Дидо Р.А., Молчанова М.О.
Хмельницький національний університет

Номер, у який прийнято статтю: №1 за 2025 рік.

06.12.2024

Начальника відділу
інтелектуальної власності та трансферу технологій Ю.В.Кравчик

ВІДДІЛ КАДРІВ

І.С.Мартинюк

УДК 004.8

ПОХИТУН АНДРІЙ
Хмельницький національний університет
e-mail: pokhytun.andrii@gmail.com
МАЗУРЕЦЬ ОЛЕКСАНДР
Хмельницький національний університет
https://orcid.org/0000-0002-8900-0650
e-mail: exe.chong@gmail.com
ДІДО РОСТИСЛАВ
Хмельницький національний університет
e-mail: rostyslav728@gmail.com
МОЛЧАНОВА МАРИНА
Хмельницький національний університет
https://orcid.org/0000-0001-9810-936X
e-mail: m.o.molchanova@gmail.com

ПРОГРАМНА АРХІТЕКТУРА ДЛЯ НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ МОДИФІКОВАНИХ ФОТОГРАФІЙ ОБЛИЧ ЛЮДЕЙ

У статті розглянуто сучасний стан наукового напрямку, що стосується нейромережевого виявлення модифікованих фотографій облич людей. Аналізуючи актуальні методи, автори запропонували новий підхід, який дозволяє не лише виявляти наявність модифікації обличчя, але й визначати спосіб її походження. Основна ідея запропонованого методу полягає в перетворенні вхідних даних у вигляді фотозображення за допомогою трьох окремих нейромережевих моделей. Перша модель відповідає за виявлення облич на фотографії, друга визначає наявність модифікацій, а третя класифікує види модифікацій, їхню складність і алгоритми, за допомогою яких були створені ці модифікації. Вихідні дані представлені у вигляді результату класифікації, що включає тип, складність та алгоритм модифікації.

У межах дослідження було створено програмну архітектуру, що автоматизує процес виявлення модифікованих зображень облич людей за допомогою нейромережевих методів. Ця архітектура сприяє створенню більш безпечних вебсередовищ, дозволяючи автоматично ідентифікувати та класифікувати модифіковані фотографії.

Для оцінки ефективності запропонованих нейромережевих моделей було проведено експерименти, в яких аналізувався вплив параметрів навчання на значення метрик виявлення модифікацій облич. Тестування проводилося на вибірці з 200 зображень, які не входили у навчальний набір. Результати експерименту показали, що найкращі результати були досягнуті при параметрах Batch size 16 та Epochs 10, де час навчання нейромережі склав 152 секунди. Метрики: Accuracy – 0.98, Precision – 0.97, Recall – 0.97, F1 – 0.97.

Подібні дослідження були проведені для оцінки нейромережевої моделі, яка відповідає за виявлення видів модифікацій. Результати також підтвердили ефективність при тих самих параметрах Batch size 16 та Epochs 10. Час навчання у цьому випадку склав 172 секунди, а метрики: Accuracy – 0.98, Precision – 0.97, Recall – 0.97, F1 – 0.97.

Висновки дослідження показують, що запропонований метод є ефективним для виявлення та класифікації модифікованих зображень облич. Подальші експерименти з архітектурою нейромереж можуть покращити результати. Такий підхід сприятиме підвищенню рівня інформаційної безпеки та захисту приватності у цифровому просторі.

Ключові слова: нейромережа, виявлення модифікованих фотографій облич людей.

POKHYTUN ANDRII, MAZURETS OLEKSANDR, DYDO ROSTYSLAV, MOLCHANOVA MARYNA

Khmelnytskyi National University

Software Architecture for Neural Network Detection of Human Faces Modified Photos

The article reviews the current state of the scientific direction related to neural network detection of modified photos of human faces. Analyzing current methods, the authors proposed a new approach that allows not only to detect the presence of facial modification, but also to determine the method of its origin. The main idea of the proposed method is to transform the input data in the form of a photo image using three separate neural network models. The first model is responsible for detecting faces in a photograph, the second determines the presence of modifications, and the third classifies the types of modifications, their complexity and the algorithms by which these modifications were created. The output data is presented in the form of a classification result, which includes the type, complexity and algorithm of the modification.

As part of the research, a software architecture was created that automates the process of detecting modified images of human faces using neural network methods. This architecture contributes to the creation of safer web environments, allowing for the automatic identification and classification of modified photographs.

To assess the effectiveness of the proposed neural network models, experiments were conducted to analyze the impact of training parameters on the values of the metrics for detecting facial modifications. Testing was performed on a sample of 200 images that were not included in the training set. The results of the experiment showed that the best results were achieved with the parameters Batch size 16 and Epochs 10, where the neural network training time was 152 seconds. Metrics: Accuracy – 0.98, Precision – 0.97, Recall – 0.97, F1 – 0.97.

Similar studies were conducted to evaluate the neural network model responsible for detecting types of modifications. The results also confirmed the effectiveness with the same parameters Batch size 16 and Epochs 10. The training time in this case was 172 seconds, and the metrics: Accuracy – 0.98, Precision – 0.97, Recall – 0.97, F1 – 0.97.

The findings of the study show that the proposed method is effective for detecting and classifying modified facial images. Further experiments with neural network architecture can improve the results. This approach will contribute to increasing the level of information security and privacy protection in the digital space.

Keywords: neural network, detection of modified photos of people's faces.

Аналіз предметної області

Сучасний світ відзначається швидким розвитком цифрових технологій, особливо в галузі обробки зображень. Разом із цим зростанням виникли проблеми, пов'язані з поширенням фальшивої інформації та необхідністю її виявлення. Зокрема, методи створення модифікованих зображень облич, такі як морфінг, deep fake та різні алгоритми генерації штучних зображень, застосовуються не тільки для розваг, але й у злочинних цілях. Ці технології дозволяють створювати реалістичні, але фальшиві зображення, що можуть використовуватися для шахрайства, шантажу та дезінформації. Тому розробка ефективних методів виявлення модифікованих зображень є важливим завданням для забезпечення інформаційної безпеки та захисту приватності.

Водночас, розвиток методів штучного інтелекту відкриває нові можливості для автоматизації процесів виявлення модифікованих зображень. Сучасні алгоритми глибокого навчання, такі як конволюційні нейронні мережі, генеративно-змагальні мережі та автокодери, здатні аналізувати великі обсяги даних і виявляти найменші ознаки підробки. Це значно підвищує ефективність і точність виявлення фальшивих зображень [1].

Відповідно, автоматизація процесу виявлення модифікованих зображень облич людей нейромережевими методами сприятиме створенню безпечних вебсередовищ.

Сучасний стан досліджень

Розвиток цифрових технологій призвів до створення програм, які можуть становити загрозу для демократії, національної безпеки та конфіденційності, зокрема через технології deepfake. Контент типу

"deepfake", особливо у вигляді зображень, поширюється з небаченою швидкістю. Такий фальшивий контент створюється за допомогою передових алгоритмів глибокого навчання, таких як генеративно-змагальні мережі (GAN), автокодери та варіаційні автокодери. Це явище сприяє поширенню дезінформації, що суттєво впливає на суспільство, знижуючи рівень довіри до контенту в соціальних мережах. Тому ця тема привертає значну увагу науковців, які працюють над розробкою ефективних методів виявлення та протидії таким загрозам. Наприклад, дослідження [2] присвячене аналізу різних методів виявлення deep fake, які навчаються на малих вибірках даних. Запропонована робота демонструє ефективну модель CNN та три попередньо навчені моделі CNN, які використовують метод переносу навчання на великому наборі даних з Kaggle, що містить 140 тисяч зображень облич. Запропонована модель CNN досягла точності 96%, в той час як DenseNet121 – 97%.

Дослідження [3] надає огляд літератури щодо методів виявлення deep fake за допомогою DL-алгоритмів, категоризуючи їх за застосуваннями: відео, зображення, аудіо та гібридні мультимедійні методи. Метою є допомогти читачам краще зрозуміти, як генеруються та виявляються deep fake, останні досягнення в цій сфері, слабкі місця існуючих методів безпеки та напрямки для подальших досліджень. Результати показують, що найбільш поширеним методом у публікаціях є використання згорткових нейронних мереж.

У дослідженні [4] було застосовано унікальну активну судово-експертну стратегію на основі архітектури Compact Ensemble-дискримінаторів з використанням глибоких умовних генеративних суперечливих мереж (CED-DCGAN) для виявлення deep fake в реальному часі під час відеоконференцій. DCGAN зосереджується на виявленні deep fake у відео, яке розбивається по кадрам, аналізуючи характеристики, оскільки технології створення переконливих підробок швидко розвиваються.

Аналіз сучасних надбань вчених, пов'язаних із виявленням модифікованих фотографій облич людей, зосереджується на кількох ключових напрямках, які показують як розвиток технологій так і нові виклики та проблеми в даній галузі. Процес розпізнавання обличчя стикається із багатьма проблемами, такими як варіації поз обличчя, зміна освітлення, більшість алгоритмів значно піддаються впливу цих змін [5, 6].

Мета дослідження полягає в створенні програмної архітектури для нейромережевого виявлення модифікованих фотографій облич людей.

Основна частина

Метод нейромережевого виявлення модифікованих фотографій облич людей відрізняється від існуючих тим, що дозволяє виявляти не лише наявність модифікації обличчя, а і спосіб її походження й призначений для перетворення вхідних даних у вигляді фотозображення, навченої нейромережевої моделі для виявлення видів модифікацій, навченої нейромережевої моделі для виявлення наявності модифікацій та навченої нейромережевої моделі для виявлення наявності облич у вихідні дані у вигляді результату класифікації, а саме тип, складність та алгоритм за допомогою якого було модифіковане фото. Схема та кроки методу наведені на рис. 1.

Процес виявлення модифікованого зображення обличчя можна розділити на декілька кроків: попередня обробка фотографії, оцінка виявлення та виокремлення обличчя на фотографії, оцінка наявності модифікації, виявлення типу модифікації [7].

Перший крок – попередня обробка фотографій. На цьому етапі фото, завантажене користувачем, проходить попередню обробку, включаючи зміну розмірності до 224x224, нормалізацію та конвертацію у тензор [8, 9]. Другий крок – виявлення та виокремлення облич на фото. На цьому етапі перевіряється, чи присутнє обличчя на фотографії, що визначає доцільність подальшої перевірки на модифікації. Третій крок – оцінка наявності модифікації на фото. На цьому етапі визначається, чи піддавалось фото будь-яким модифікаціям. Завершальний крок – виявлення типу модифікації. Це включає аналіз складності та алгоритму, за допомогою якого фото було модифіковане.

Вихідними даними методу виявлення модифікованих зображень облич є результат у вигляді ймовірності належності зображення до конкретної модифікації, а також оцінка наявності модифікації.

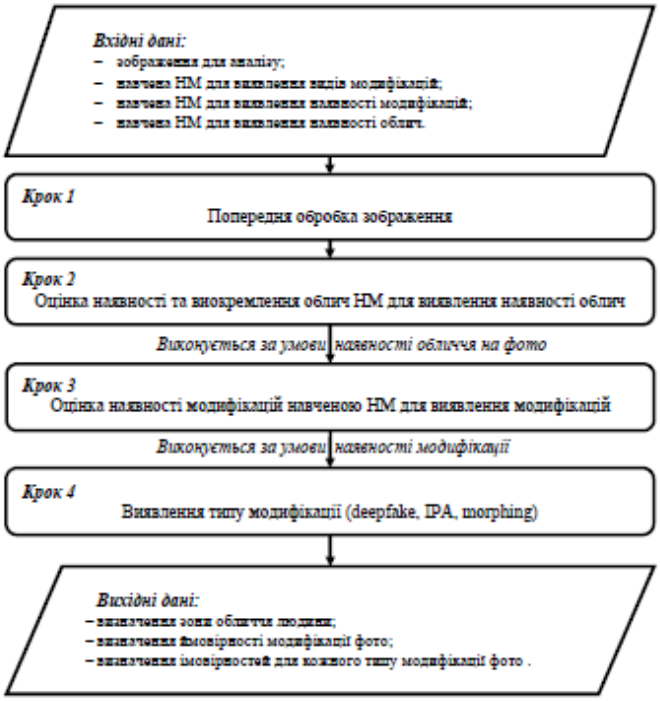


Рис. 1. Схема та кроки методу виявлення модифікованих зображень облич людей

Архітектура нейромережі для виявлення наявності модифікації, що є частиною вхідних даних методу, наведена на рис. 2.

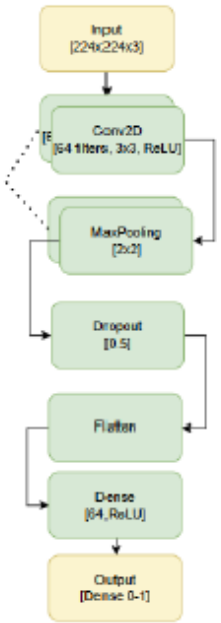


Рис. 2. Архітектура нейромережі для виявлення наявності модифікації

Вхідний шар приймає вхідні дані у вигляді зображення розмірністю 224x224x3, для можливості роботи із кольоровими зображеннями.

Згорткові шари містять 64 фільтри розміром 3x3, та функцію активації ReLU. Після операції згортки, на вихід отримується тензор розмірністю 224x224x64, оскільки кількість фільтрів замінює кількість каналів. Функція активації ReLU додає нелінійність у модель, що допомагає прибрати негативні значення [10].

Шари пулінгу отримують на вхід тензор розмірністю 224x224x64, після чого відбувається зменшення розмірності вдвічі, завдяки ядру 2x2. Тобто на вихід передається тензор розмірністю 112x112x64. Шар dropout виконує функції випадкового відключення, щоб запобігти перенавчанню моделі. Тобто на даному етапі відбувається відключення 50% нейронів (обрано випадково). Шар flatten перетворю тривимірний тензор у одновимірний вектор, оскільки вихідний шар dense, приймають на вхід лише одновимірні вектори. Вихідний шар dense, повертає один нейрон 0-1, для отримання результату класифікації, тобто даний шар визначає чи модифіковане зображення.

Архітектура нейромережі для виявлення наявності модифікації, що є також частиною вхідних даних методу, наведена на рис. 3.

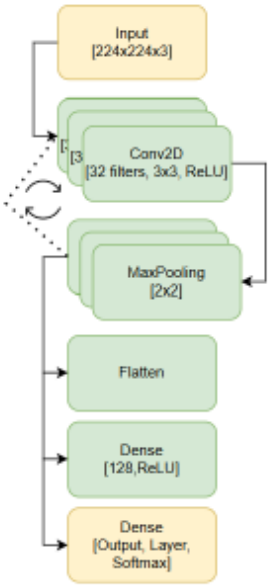


Рис. 3. Архітектура нейромережі для виявлення для виявлення типу модифікації

Вхідний шар (Input) отримує зображення розміром 224x224 пікселі, із трьома каналами, оскільки зображення кольорові. Згорткові шари (Conv2D), містять 32 фільтри розмірністю 3x3, що обробляють зображення. За допомогою даних шарів, модель навчається виявляти базові ознаки(краї, текстури). Шари пулінгу (MaxPooling), використовують філбтри 2x2, які зменшують розмірність вхідного зображення вдвічі, отримуючи найважливіші ознаки, що знижує обчислювальну складність нейромережевої моделі. Наступний шар перетворює багатомірний масив, отриманий із попередніх шарів, у одновимірний вектор даних, для того щоб підготувати дані для обробки у повнозв'язних шарах. Повнозв'язний шар (Dense), обробляє отриманий вектор, що дає змогу виявляти складніші ознаки. Шар містить 128 нейронів, кожен із яких запускає функцію активації.

Вихідний шар (Output), створює ймовірності для кожного із можливих класів модифікацій зображення

та повертає той, де сума ймовірностей найбільша. Вихідний шар містить 9 виході відповідно до кожного класу, а саме deepfake, ipa (easy), ipa (mid), ipa (hard), morphing (amsl), morphing (facemorfer), morphing (opencv), morphing (webmorpher), morphing (stylefan2).

Описані архітектури імплементовані у програмну архітектуру для нейромережевого виявлення модифікованих фотографій облич людей, ланцюжок послідовних дій від завантаження датасету до отримання результатів класифікації якої наведено на рис. 4. Початок процесу відбувається після того як користувач завантажує вхідний набір даних – датсет [11]. Після чого відбувається перетворення зображень у тензор.

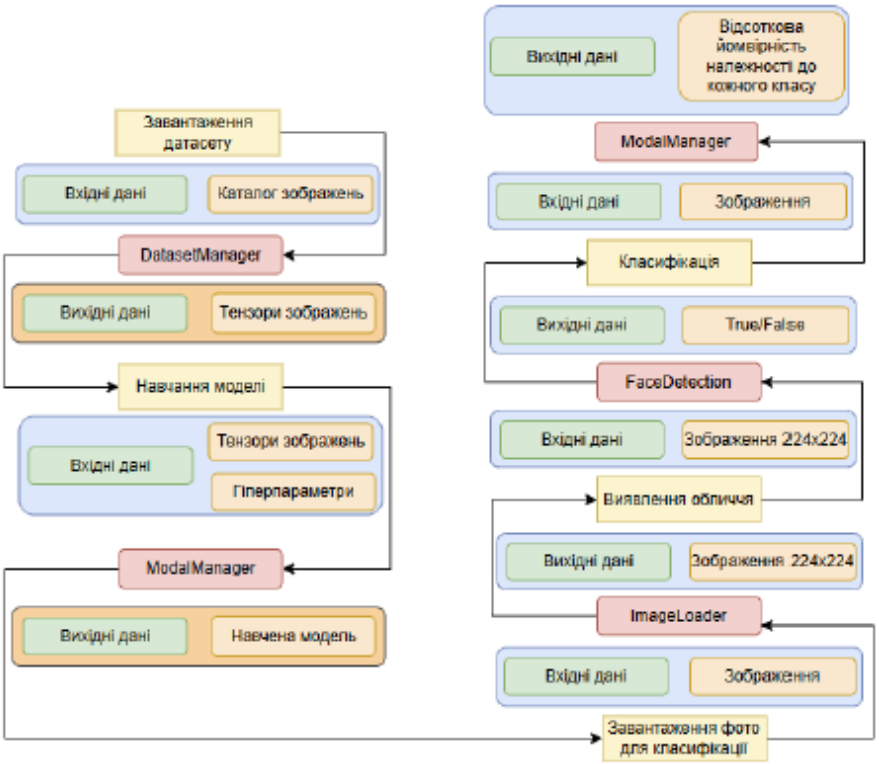


Рис. 4. Програмна архітектура системи

Наступним етапом є навчання моделі, куди подаються тензори зображень та гіперпараметри за замовчуванням або ж ті, які вказав користувач. В якості вихідних даних отримана навчена нейромережева модель.

Після того, як модель готова до класифікації, користувач завантажує зображення, після чого відбувається зміна розміру фото (якщо потрібно) до 224x224.

Після приведення до потрібного розміру, виконується посередник (blazeface) для виявлення та виокремлення обличчя на зображені. Якщо результат позитивний (зображення містить обличчя) запускається процес класифікації та виведення результатів у вигляді відсоткової ймовірності належності до всіх класів. Якщо ж обличчя не виявлено користувачу отримує сповіщення, що обличчя не знайдено та прохання завантажити друге фото.

Приклад роботи створеного програмного забезпечення для нейромережевого виявлення модифікованих фотографій облич людей наведено на рис. 5.

Рис. 5. Приклад роботи програмного системи

Отже, наведено програмну архітектуру системи для нейромережевого виявлення модифікованих фотографій облич людей, яка є програмною реалізацією запропонованого методу, що дозволяє автоматизувати процес виявлення ймовірності модифікацій із зазначенням її виду.

Експерименти та дискусія

В ході дослідження ефективності нейромереж було досліджено вплив параметрів навчання на отримані значення метрик. Тестування проводились на відносно невеликій вибірці зображень (200 екземплярів), результати наведені у таблиці 1.

Таблиця 1

Вплив параметрів навчання на виявлення наявності модифікацій

Batch size	Epochs	Час навчання (с)	Accuracy	Precision	Recall	F1
20	16	187	0.96	0.96	0.96	0.96
16	10	152	0.98	0.97	0.97	0.97
12	8	136	0.95	0.962	0.962	0.962

Результати експерименту з таблиці 1 відображені на рис.6.

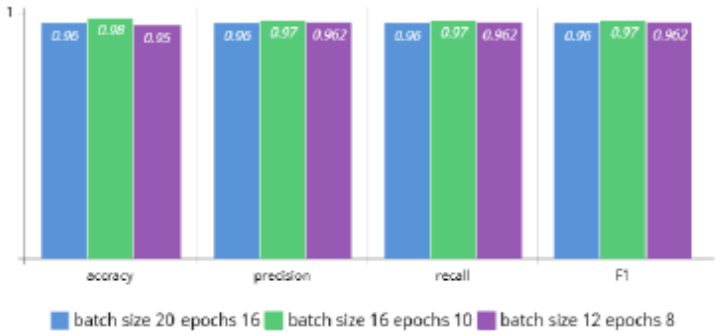


Рис. 6. Вплив параметрів навчання на виявлення наявності модифікацій

Як видно з рис.6 та з таблиці 1, найкращі результати навчання досягнуті при параметрах Batch size 16 та Epochs 10. Для таких параметрів час навчання нейромережі для виявлення модифікацій склав 152 секунди. При цьому отримані метрики Accuracy = 0.98, Precision = 0.97, Recall = 0.97, F1 = 0.97.

Аналогічно як і з нейромережею для виявлення наявності модифікації, проведено дослідження впливу

параметрів навчання на виявлення видів модифікації. Отримані результати наведено в таблиці 2.

Таблиця 2

Вплив параметрів навчання на виявлення видів модифікацій

Batch size	Epochs	Час навчання (с)	Accuracy	Precision	Recall	F1
20	16	237	0.95	0.94	0.94	0.94
16	10	172	0.98	0.97	0.97	0.97
12	8	156	0.94	0.943	0.943	0.943

Результати з таблиці 2, відображені на рис. 7.

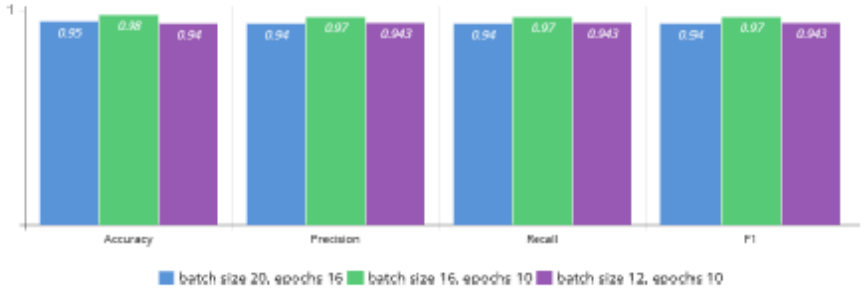


Рис. 7. Вплив параметрів навчання на виявлення видів модифікацій

Провівши аналіз отриманих результатів, наведених на рис.7, можна зробити висновок, що оптимально вказувати наступні параметри навчання: batch size 16 та epochs 10. Із даними параметрами навчання нейромережа для виявлення видів модифікацій отримала такі результати: accuracy = 0.98, precision = 0.97, recall = 0.97, f1 = 0.97.

Однак, результати розробленого методу можна покращити шляхом експериментів над архітектурами нейромереж, на що і будуть спрямовані подальші дослідження.

Висновки

У статті розглянуто актуальний стан наукового напрямку нейромережевого виявлення модифікованих фотографій облич людей. На базі опрацьованого матеріалу запропоновано метод нейромережевого виявлення модифікованих фотографій облич людей, що відрізняється від існуючих тим, що дозволяє виявляти не лише наявність модифікації обличчя, а й спосіб її походження.

Розроблений метод призначений для перетворення вхідних даних у вигляді фотозображення за допомогою навченої нейромережевої моделі для виявлення видів модифікацій, навченої нейромережевої моделі для виявлення наявності модифікацій, та навченої нейромережевої моделі для виявлення наявності облич. Вихідні дані представлені у вигляді результату класифікації, що включає тип, складність та алгоритм, за допомогою якого було модифіковане фото.

У межах проведеного дослідження було створено програмну архітектуру, яка дозволяє автоматизувати процес виявлення модифікованих зображень облич людей нейромережевими методами, що в свою чергу сприятиме створенню безпечних вебсередовищ.

В ході дослідження ефективності запропонованих нейромереж було проаналізовано вплив параметрів навчання на значення метрик виявлення модифікацій облич. Тестування проводилося на вибірці зображень 200 екземплярів. Результати експерименту показали, що найкращі результати були досягнуті при параметрах Batch size 16 та Epochs 10, де час навчання нейромережі склав 152 секунди. Метрики показали таку ефективність: Accuracy становила 0.98, Precision – 0.97, Recall – 0.97, та F1 – 0.97.

Подібні дослідження були проведені для оцінки нейромережі для виявлення видів модифікацій, де результати також підтвердили ефективність при тих самих параметрах Batch size 16 та Epochs 10. Час навчання у цьому випадку склав 172 секунди, а метрики були такими: Accuracy = 0.98, Precision = 0.97, Recall = 0.97, F1 = 0.97. Це свідчить про те, що обрані параметри навчання є оптимальними для даної задачі. Однак, результати можуть бути покращені шляхом подальших експериментів з архітектурою нейромереж, що стане предметом майбутніх досліджень.

Література

1. Pokhytun A. Method for Neural Network Detecting Changed Images of People's Faces Using CNN / A. Pokhytun, O. Mazurets, M. Molchanova, O. Tyschenko // New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International scientific and practical conference. October 21-23, 2024. Marseille, France. – 2024. – Pp. 35-40.

2. Umadevi M.Deep Fake Face Detection using Efficient Convolutional Neural Networks / M. Umadevi, S. Krishna, N. Kumar // 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal, 2024. – Pp. 344-352

3. Heidari A. Deepfake detection using deep learning methods: A systematic and comprehensive review / A. Heidari, N. Jafari Navimipour, H. Dag, M. Unal // WIREs Data Mining and Knowledge Discovery. – 2023. – [Електронний ресурс] – Режим доступу: https://doi.org/10.1002/widm.1520 (дата доступу: 16.12.2024).

4. Chauhan R. Fake Faces Unveiled: A Comprehensive Study on Detecting Generated Facial Images / R. Chauhan, M. Sethi, S. Ahuja // 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024. – Pp. 475-482.

5. Bohdanova A. Gesture recognition using a neural network in real time / A. Bohdanova, O. Mazurets, O. Sobko // Black Sea Science 2023: Proceedings of the International Competition of Student Scientific Works. Odesa National University of Technology. Odesa, ONUT, 2023. – Pp. 556-566.

6. Novak Y. Practical Application of Method of Automated Personal Identification by Fingerprints Using Convolution Neural Networks / Y. Novak, O. Mazurets // Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». December 27-29, 2023. Stockholm, Sweden, International Scientific Unity. – 2023. – Pp. 136-140..

7. Molchanova M. Object-oriented model for neural network damage detection of mail packages / M. Molchanova, O. Mazurets, V. Klimenko, Ev. Kuflevsky // Proceedings of XIV International Scientific and Practical Conference «Solving Scientific Problems Using Innovative Concepts». March 13-15, 2024. Copenhagen, Denmark. – 2024. – Pp. 58-62.

8. Mazurets O. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services / O. Mazurets, M. Molchanova, V. Klimenko, D. Klopotivskyi // Proceedings of XXII International Scientific and Practical Conference «Modern Scientific Research: Theoretical and Practical Aspects». May 8-10, 2024. Oslo, Norway. – 2024. – Pp. 64-68.

9. Mazurets O. An Approach to Using MobileNet CNN-model for Gesture Recognition / O. Mazurets, O. Zalutska, O. Tyschenko, A. Bohdanova // Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions». May 15-17, 2024. Munich, Germany. – 2024. – Pp. 59-64.

10. Mazurets O. V. Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones / O. Mazurets, V. Klimenko, M. Molchanova, A. Sultanov // Global Science: Prospects and Innovations. Proceedings of the 10th International scientific and practical conference. Cognum Publishing House. Liverpool, United Kingdom. – 2024. – Pp. 198-207.

11. Похитун А.В. Підхід до формування датасету для нейромережевого виявлення модифікованих фотографій облич людей / А.В. Похитун, О.В. Мазурець, М.О. Молчанова, О.В. Бармак // Збірник наукових

праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. – С. 428-433.

References

1. Pokhytun A. Method for Neural Network Detecting Changed Images of Peoples Faces Using CNN / A. Pokhytun, O. Mazurets, M. Molchanova, O. Tyschenko // New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International scientific and practical conference. October 21-23, 2024. Marseille, France. – 2024. – Pp. 35-40.

2. Umadevi M.Deep Fake Face Detection using Efficient Convolutional Neural Networks / M. Umadevi, S. Krishna, N. Kumar // 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), Dhulikhel, Nepal, 2024. – Pp. 344-352

3. Heidari A. Deepfake detection using deep learning methods: A systematic and comprehensive review / A. Heidari, N. Jafari Navimipour, H. Dag, M. Unal // WIREs Data Mining and Knowledge Discovery. – 2023. – [Elektronnyi resurs] – Rezhym dostupu: https://doi.org/10.1002/widm.1520 (data dostupu: 16.12.2024).

4. Chauhan R. Fake Faces Unveiled: A Comprehensive Study on Detecting Generated Facial Images / R. Chauhan, M. Sethi, S. Ahuja // 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024. – Pp. 475-482.

5. Bohdanova A. Gesture recognition using a neural network in real time / A. Bohdanova, O. Mazurets, O. Sobko // Black Sea Science 2023: Proceedings of the International Competition of Student Scientific Works. Odesa National University of Technology. Odesa, ONUT, 2023. – Pp. 556-566.

6. Novak Y. Practical Application of Method of Automated Personal Identification by Fingerprints Using Convolution Neural Networks / Y. Novak, O. Mazurets // Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». December 27-29, 2023. Stockholm, Sweden, International Scientific Unity. – 2023. – Pp. 136-140..

7. Molchanova M. Object-oriented model for neural network damage detection of mail packages / M. Molchanova, O. Mazurets, V. Klimenko, Ev. Kuflevsky // Proceedings of XIV International Scientific and Practical Conference «Solving Scientific Problems Using Innovative Concepts». March 13-15, 2024. Copenhagen, Denmark. – 2024. – Pp. 58-62.

8. Mazurets O. Datalogic Model for Image Recognition by Convolutional Neural Network Using Cloud Services / O. Mazurets, M. Molchanova, V. Klimenko, D. Klopotivskyi // Proceedings of XXII International Scientific and Practical Conference «Modern Scientific Research: Theoretical and Practical Aspects». May 8-10, 2024. Oslo, Norway. – 2024. – Pp. 64-68.

9. Mazurets O. An Approach to Using MobileNet CNN-model for Gesture Recognition / O. Mazurets, O. Zalutska, O. Tyschenko, A. Bohdanova // Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions». May 15-17, 2024. Munich, Germany. – 2024. – Pp. 59-64.

10. Mazurets O. V. Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones / O. Mazurets, V. Klimenko, M. Molchanova, A. Sultanov // Global Science: Prospects and Innovations. Proceedings of the 10th International scientific and practical conference. Cognum Publishing House. Liverpool, United Kingdom. – 2024. – Pp. 198-207.

11. Pokhytun A.V. Pidkhid do formuvannia datasetu dlia neiromerezhevoho vyiavlennia modyfikovanykh fotohrafii oblych liudei / A.V. Pokhytun, O.V. Mazurets, M.O. Molchanova, O.V. Barmak // Zbirnyk naukovykh prats za materialamy XVI Vseukrainskoi naukovo-praktychnoi konferentsii «Aktualni problemy kompiuternykh nauk APKN-2024». 15-16 lystopada 2024. Khmelnytskyi, 2024. – S. 428-433.

Додаток Д

Презентаційний матеріал

Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

Кваліфікаційна робота на тему метод виявлення модифікованих зображень облич людей нейромережевими засобами

Виконав: А. В. Похитун
Керівник: О. В. Бармак

Актуальність кваліфікаційної роботи

Сучасний світ характеризується досить стрімким розвитком цифрових технологій, пов'язаних із обробкою зображень. Проте із стрімким розвитком цифрових технологій почали з'являтись проблеми із неправдивою, підробленою інформацією, та способами її виявлення. Зокрема, методи створення модифікованих зображень облич, такі як морфінг, дипфейк, та різні алгоритми генерації штучних зображень, використовуються не тільки у розважальних цілях, а й у злочинних.

Розробка методу для виявлення модифікованих облич людей може значно покращити стан у багатьох сферах життя та допоможе боротись із неправдивими новинами, сприятиме підвищенню довіри до соціальних мереж та може стати потужним інструментом у боротьбі із рядом інших проблем.

Мета та задачі кваліфікаційної роботи

Мета кваліфікаційної роботи полягає у підвищенні точності виявлення модифікованих фотографій облич людей. Для досягнення мети, обрано наступні пункти для дослідження:

- провести аналіз подібного програмного забезпечення;
- дослідити доступні в мережі набори даних, та підготувати власний датасет для навчання згорткової нейронної мережі;
- провести аналіз та обрати метод виявлення та виокремлення обличчя на фото;
- розробити метод виявлення модифікованих зображень облич людей;
- провести тестування та аналіз поданого методу виявлення модифікованих зображень облич людей.

Об’єкт та предмет дослідження

Об’єкт дослідження – процес виявлення модифікованих облич людей на зображенні.

Предмет дослідження – моделі, методи та засоби для виявлення модифікованих облич людей на зображенні за допомогою нейромережевої класифікації

Аналіз предметної області

Автоматизація процесів виявлення модифікованих облич людей є важливою задачею інформаційних технологій, яка дозволить підвищити швидкість та надійність ідентифікації.

У сучасному світі змінити зовнішній вигляд по фото не є важкою задачею. Одним із найбільш впливових факторів є поява генеративних моделей які здатні створювати досить реалістичні зображення облич людей, або змінювати їх.

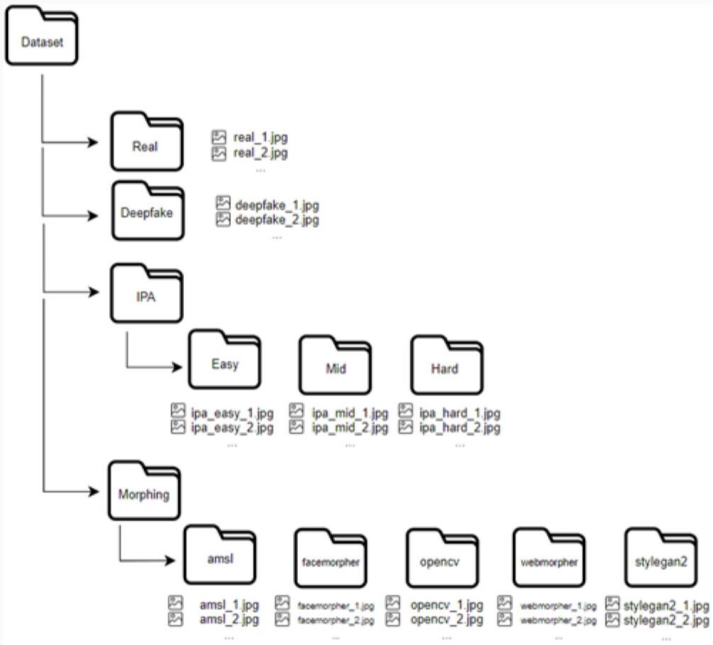
Серед великої кількості видів модифікацій, можна виділити декілька, які користуються найбільшою популярністю

- цифрові фільтри та косметичні зміни;
- глибинні фейки;
- морфінг;
- алгоритм обробки зображень.

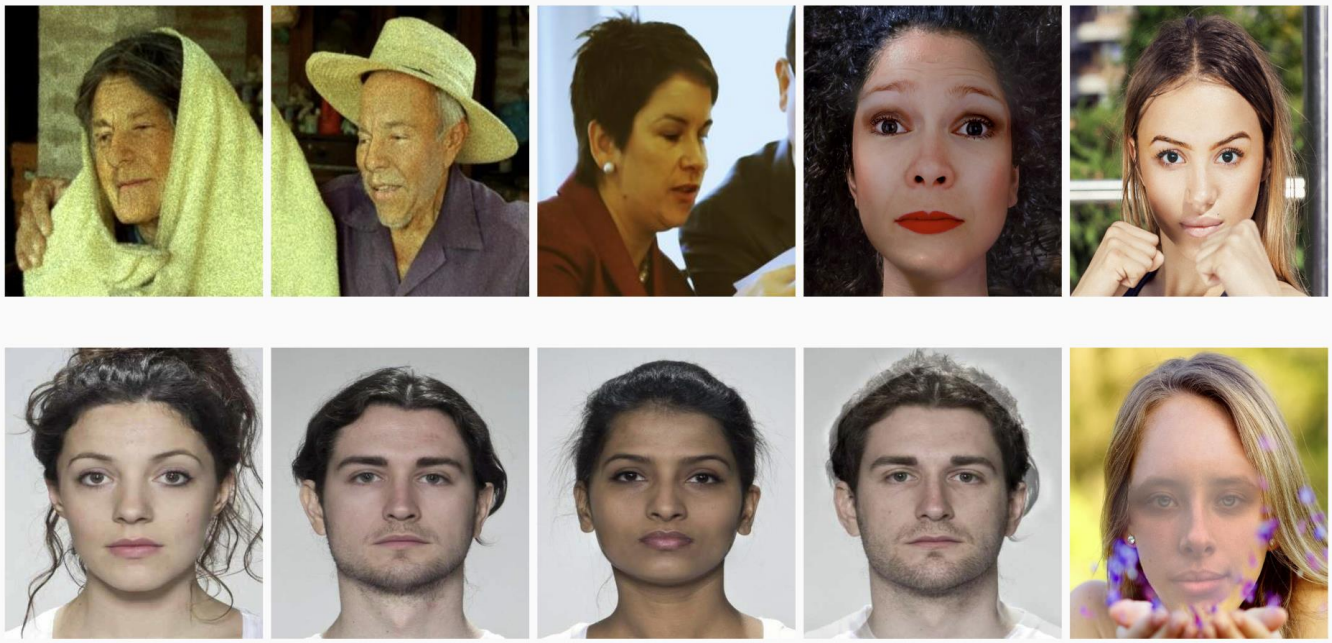
Підготовка зразків робочих даних. Формування датасету

Створений набір даних містить наступні класи (підкласи):

- deepfake;
- real;
- ipa (easy,mid,hard);
- morphing (amsl, facemorfer, opencv, webmorpher, stylefan2).



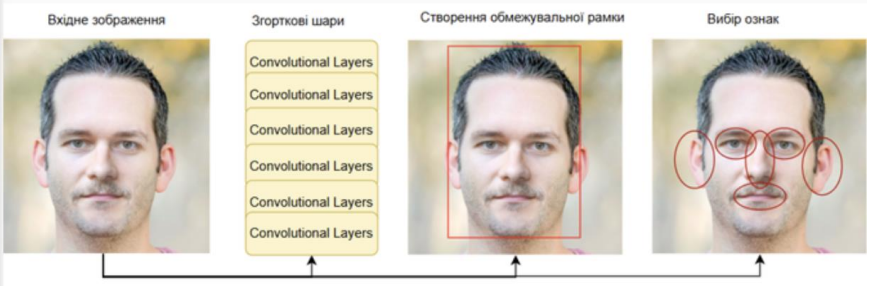
Приклади зображень із створеного датасету

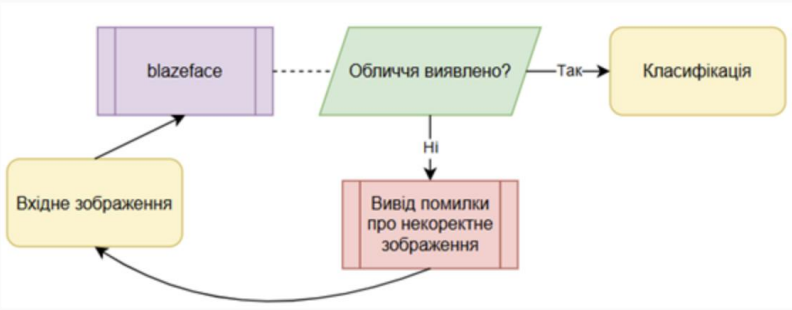


BlazeFace

Перш ніж виконувати процес виявлення типу модифікації, потрібно переконатись що на фото присутнє обличчя. В ході розробки методу виявлення модифікованих зображень облич людей було прийнято рішення використовувати готову модель – BlazeFace.

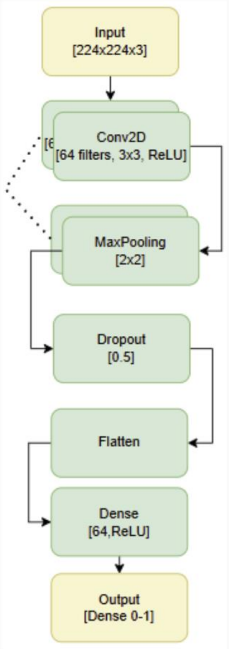
Як результат, blazeface повертає масив, де кожен елемент містить координати певної ключової точки(око, ніс, тощо). Якщо на фото буде виявлено більше ніж одне обличчя, на виході буде отримано список об'єктів із координатами. В випадку якщо обличчя не виявлено – модель вертає пустий масив.

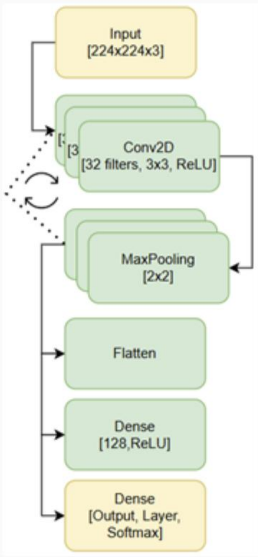




Архітектури нейронних мереж

Правильно побудована архітектура є важливим етапом у розробці методу виявлення модифікованих зображень, оскільки структура архітектури впливає на швидкість, точність та обчислювальну ефективність моделі





Ефективність НМ для ідентифікації модифікації

Час навчання = 324.7 секунди

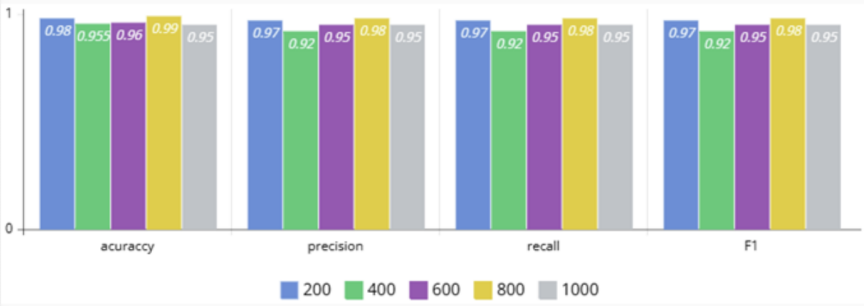
accuracy = 0.99

precision = 0.98

recall = 0.98

F1 = 0.98

Кількість зображень	Час навчання (с)	Accuracy	Precision	Recall	F1 score
200	125.5	0.98	0.97	0.97	0.97
400	180.8	0.955	0.92	0.92	0.92
600	253.3	0.96	0.95	0.95	0.95
800	324.7	0.99	0.98	0.98	0.98
1000	377.6	0.95	0.95	0.95	0.95



Ефективність НМ виявлення типу модифікації

Час навчання = 324.7 секунди

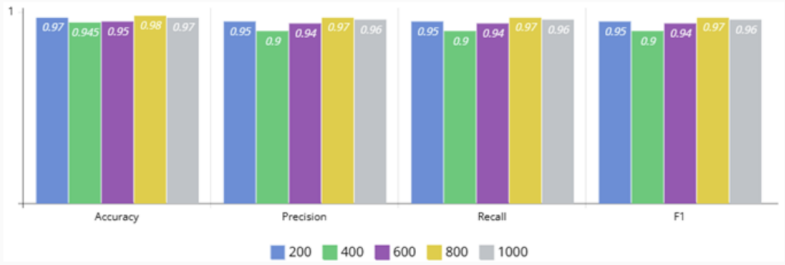
accuracy = 0.98

precision = 0.97

recall = 0.97

f1 = 0.97

Кількість зображень	Час навчання (с)	Accuracy	Precision	Recall	F1 score
200	145.2	0.97	0.95	0.95	0.95
400	208.3	0.945	0.90	0.90	0.90
600	268.3	0.95	0.94	0.94	0.94
800	354.7	0.98	0.97	0.97	0.97
1000	416.2	0.97	0.96	0.96	0.96



Висновок

Як результат кваліфікаційної роботи магістра, розроблено метод виявлення модифікованих зображень облич людей, та інформаційну систему, яка використовує даний метод.

В ході виконання, досліджено предметну область виявлення модифікованих зображень облич людей та її актуальність. Розглянуто варіативність модифікацій та способи їх створення.

В процесі виконання роботи було обрано вже навчену модель blazeface, яка виконує роль посередника та визначає чи містить зображення обличчя. Розглянуто варіативність результату роботи даної моделі.

Кінцевий результат кваліфікаційної роботи магістра – інформаційна система, яка використовує розроблений метод виявлення модифікованих зображень облич людей. Даний метод показав точність за метриками, приблизно 97% при навчанні на створеному датасеті. Для інформаційної системи, розроблено інструкцію користувача та проведено тестування найважливіших функцій.

Sun Dec 15 21:38:48 EET 2024, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism v-15.258 Educational

Максимальне співпадіння з одним документом 1.0%

Словники перевірки: en_US, ru_RU, ua_UA. **Помилок в документах: 11%**

	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
ID: 159734 Назва: КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА на тему Метод виявлення модифікованих фотографій облич людей нейромережевими засобами Додано в БД: 2024-12-15 Автора: Андрій ПОХИТУН Керівники: Олександр БАРМАК Консультанти: Опоненти:	71598	1066	2724 (4%)	40 (4%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Андрій ПОХИТУН

Співавтор:

Назва: Метод виявлення модифікованих фотографій облич людей нейромережевими засобами

Науковий керівник: Олександр БАРМАК, д.т.н., проф.

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1:9.6%

Коефіцієнт подібності 2:4.9%

Мікропробіли: 0

Заміна букв: 0

Інтервали: 0

Білі знаки: 92

Дата створення звіту: 2024-12-15 21:31:44.0

Після аналізу Звіту подібності констатую наступне:

☑ **Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.**

☐ **Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.**

☐ **Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.**

Обгрунтування:

Дата 15.12.2024 експерт Петровський Р.Р

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ
КАФЕДРИ КОМП'ЮТЕРНИХ НАУК
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ МАГІСТРА ДО ЗАХИСТУ
ЗА РЕЗУЛЬТАТАМИ АНАЛІЗУ ЗВІТУ ПОДІБНОСТІ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: *Метод виявлення модифікованих фотографій облич людей нейромережевими засобами*

Автор: *Андрій ПОХИТУН*

Спеціальність: *122 – Комп'ютерні науки*

Освітня програма: *освітньо-професійна*

Науковий керівник: *зав. каф. КН, д.т.н., професор Олександр БАРМАК*

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	–
3	Виявлені запозичення не є плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того як буде відкоригована та допрацьована і успішно пройде повторну перевірку на академічний плагіат.	–
4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	–

Підтвердження:

Запозичення, виявлені в роботі, є законними і не є плагіатом, оскільки:

1) за програмою Anti-Plagiarism виявлені 1%.
2) за програмою StrikePlagiarism КП1 9,6%, КЦ 4,9%,

які містять матеріали огляду предметної області; інші схожості є фрагментарними – містять поширені конструкції, загальновідомі терміни, скорочення та визначення, що, з урахуванням наведених обґрунтувань, відповідає характеру наукового дослідження і свідчить на користь кваліфікаційної роботи. КЦ 4,9% пояснюється переважним чином збігом із власними публікаціями автора за темою кваліфікаційної роботи магістра. Запозичення, виявлені в роботі є законними і не є плагіатом.

Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Керівник роботи _____ Олександр БАРМАК

Гарант ОП _____ Руслан БАГРІЙ

Завідувач кафедри КН _____ Олександр БАРМАК

ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу магістра

гр. КНм-23-1 Андрія Похитуна за темою: *Метод виявлення модифікованих фотографій облич людей нейромережевими засобами*

1. Актуальність обраної теми

У сучасному світі змінити зовнішній вигляд по фото не є важкою задачею. Одним із найбільш впливових факторів є поява генеративних моделей, які здатні створювати досить реалістичні зображення облич людей, або змінювати їх. Технології модифікації обличчя використовуються не лише в розважальних цілях, а й можуть бути застосовані у злочинних намірах. Це є особливо важливим в умовах інформаційних конфліктів, коли фальшиві фото та відео можуть мати серйозний вплив на моральний стан суспільства. Автоматизація процесів виявлення модифікованих облич людей є важливою задачею інформаційних технологій, яка дозволить підвищити швидкість та надійність ідентифікації.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Кваліфікаційна робота магістра Андрія Похитуна, пов'язана з створенням методу виявлення модифікованих фотографій облич людей нейромережевими засобами, повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи.

3. Професійні та особистісні якості магістранта

Під час виконання своєї кваліфікаційної роботи на ступінь магістра, Андрій Похитун виявив високі професійні якості, проявивши себе як дисциплінований та кваліфікований студент. Здобувач з великою увагою та завзяттям підійшов до кожного завдання, забезпечуючи їх виконання на високому рівні якості, з дотриманням встановлених термінів.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Результати, отримані в результаті виконання кваліфікаційної роботи бакалавра, є результатом самостійної діяльності студента. Отримані положення наукової новизни та інновації, описані в роботі, дозволили покращити існуючі методи в галузі виявлення модифікованих фотографій облич людей нейромережевими засобами.

5. Наукова новизна та оригінальність запропонованих підходів

Результати кваліфікаційної роботи магістра включають інноваційний метод для виявлення модифікованих фотографій облич людей нейромережевими засобами. В результаті створено структурований датасет та розроблено новий метод виявлення модифікованих зображень облич людей, що дозволяє виявляти не лише наявність модифікації зображення обличчя, а і спосіб її походження.

6. Ступінь оволодіння методами дослідження

Магістр Андрій Похитун виявив високий ступінь оволодіння необхідними методами дослідження.

7. Повнота та якість розкриття теми роботи

Тема роботи в повній мірі обґрунтована й розкрита, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання у роботі виконані, а також проведено аналіз результатів прикладного застосування запропонованих засобів методу виявлення модифікованих фотографій облич людей нейромережевими засобами.

8. Логічність, послідовність, аргументованість, літературна грамотність викладу матеріалу

Структура роботи й послідовність викладення логічні та відповідні поставленій меті. Викладення матеріалу грамотне та виявляє високий ступінь відповідності стилю.

9. Можливість практичного застосування кваліфікаційної роботи, окремих її частин

Було розроблено інформаційну систему для виявлення модифікованих фотографій облич людей нейромережевими засобами, яка дозволяє оцінити ступінь і тип змін на фото. Система складається з датасету та трьох підсистем: «Навчання та донавчання нейронної мережі (НМ)», «Використання попередньо навченої НМ» і «Класифікація зображення», яка визначає наявність і тип модифікацій на завантаженому користувачем фото. Ці підсистеми реалізують основний функціонал методу виявлення змінених облич, достатній для дослі.

10. Висновок про можливість допуску кваліфікаційної роботи до захисту, на яку оцінку заслуговує робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «добре».

Науковий керівник _____ зав. кафедри КН, д.т.н, проф. Олександр БАРМАК

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ
кафедра комп'ютерних наук

ВІДГУК ОПОНЕНТА

на кваліфікаційну роботу магістра

гр. КНм-23-1 Андрія Похитуна за темою: *Метод виявлення модифікованих фотографій облич людей нейромережевими засобами*

1. Актуальність обраної теми

Сучасний світ відзначається швидким розвитком цифрових технологій, зокрема у сфері обробки зображень. Однак разом із цим виникають нові проблеми, пов'язані з поширенням фальшивої та підробленої інформації, а також методами її виявлення. Зокрема, технології створення змінених зображень облич, такі як морфінг, дипфейк та різні алгоритми генерації штучних зображень, використовуються не тільки в розважальних цілях, але й у злочинних схемах. Це питання набуває особливої актуальності в умовах війни, оскільки Інтернет стає майданчиком для боротьби, де інформаційні атаки, включаючи фальшиві фото та відео, можуть серйозно вплинути на моральний стан населення. Таким чином, тема «Метод виявлення модифікованих фотографій облич людей нейромережевими засобами» є досить актуальною, тому кваліфікаційна робота магістра має значну наукову та практичну цінність.

2. Відповідність роботи предметній області спеціальності 122 Комп'ютерні науки та загальним вимогам до наукових робіт

Обрана тема «Метод виявлення модифікованих фотографій облич людей нейромережевими засобами», в межах якої виконані поставлені задачі, повною мірою відповідає предметній області спеціальності 122 «Комп'ютерні науки» та вимогам до кваліфікаційної роботи магістра.

3. Повнота розкриття мети та завдань дослідження

В роботі автор повністю розкриває мету дослідження та поставлені в межах теми завдання.

4. Наявність наукової новизни

У результаті виконання магістерської роботи був розроблений новий метод для виявлення модифікацій облич на зображеннях, який дозволяє не лише визначити факт зміни, а й з'ясувати метод її виконання. Метод реалізовано в інформаційній системі, яка продемонструвала покращення результатів ідентифікації модифікацій з точністю: Accuracy 0.99, Precision 0.98, Recall 0.98, F1 0.98, що на понад 0.04 більше за показники у відомих дослідженнях.

5. Зміст кожного розділу роботи

Робота містить чотири розділи. У першому розділі виконано аналіз сучасного стану області виявлення модифікованих облич людей. Другий розділ присвячено розробці методу виявлення модифікованих фотографій облич людей нейромережевими засобами. У третьому розділі виконано та проєктування інформаційної системи виявлення модифікованих зображень облич людей. У четвертому розділі виконано дослідження методу виявлення модифікованих зображень облич людей.

6. Ступінь розкриття теми роботи

Тема кваліфікаційної роботи повною мірою розкрита та обґрунтована, проведено аналіз актуальності та відомих досліджень в межах обраної теми, поставлені завдання, які у роботі виконані, та проведено аналіз результатів прикладного застосування запропонованих методу і засобів.

7. Якість оформлення кваліфікаційної роботи

Оформлення роботи відповідає необхідним нормам та вимогам, які ставляться до оформлення кваліфікаційних робіт.

8. Недоліки кваліфікаційної роботи

Перелік скорочень, поданий в кваліфікаційній роботі наведено не повною мірою, рекомендовано доповнити термінологією, що зустрічається в роботі.

9. Загальний висновок (допускається чи не допускається до захисту), якої оцінки заслуговує кваліфікаційна робота

Враховуючи високий рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «відмінно».

Опонент (прізвище, ім'я, по батькові, посада, місце роботи)

Тиагур Єлизавета Геннадіївна, д.т.н., доцент доцент каф. комп'ютерної інженерії та інформаційних систем

«16» 12 2024 р

підпис