

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ШТУЧНОГО ІНТЕЛЕКТУ**

КВАЛІФІКАЦІЙНА РОБОТА
на тему: «Система розпізнавання емоцій на фото та відео
за допомогою методів глибокого навчання»

на здобуття освітнього ступеня бакалавра
зі спеціальності 122 Комп'ютерні науки
освітньо-професійної програми Штучний інтелект

*Кваліфікаційна робота містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело*

_____ Владислав ШКЛЯРЕНКО
(підпис)

Виконав: Здобувач вищої освіти група ШІД-42
Владислав ШКЛЯРЕНКО

Керівник: Антон ШАНТИР
к.т.н., доцент

Рецензент:
науковий ступінь,
вчене звання

_____ Ім'я, ПРИЗВИЩЕ

Київ 2025

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
Навчально-науковий інститут інформаційних технологій

Кафедра Штучного інтелекту

Ступінь вищої освіти Бакалавр

Спеціальність 122 Комп'ютерні науки

Освітньо-професійна програма Штучний інтелект

ЗАТВЕРДЖУЮ

Завідувач кафедри Штучного інтелекту

_____ Ольга ЗІНЧЕНКО

«_____» _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Шкляренко Владиславу Володимировичу

1. Тема кваліфікаційної роботи: Система розпізнавання емоцій на фото та відео за допомогою методів глибокого навчання.

керівник кваліфікаційної роботи Антон Шантир к.т.н., доцент,
затверджені наказом Державного університету інформаційно-комунікаційних технологій від «24» лютого 2025 р. № 56

2. Строк подання кваліфікаційної роботи «23» травня 2025 р.

3. Вихідні дані до кваліфікаційної роботи: науково-технічна література, архітектура CNN, документація фреймворків TensorFlow/Keras та Flask.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

Дослідження та аналіз предметної області

Розробка моделі розпізнавання емоцій

Розробка вебдодатку

5. Перелік графічного матеріалу: презентація

6. Дата видачі завдання «25» лютого 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1	Аналіз актуальності та наявних рішень	25.02-25.03.25	Виконав
2	Вивчення та вибір методів глибокого навчання	26.03-10.04.25	Виконав
3	Розробка та тестування моделі розпізнавання емоцій	11.04-22.04.25	Виконав
4	Вибір технологічного стеку для розробки вебдодатку	23.04-25.04.25	Виконав
5	Розробка вебдодатку та інтеграція моделі	26.04-10.05.25	Виконав
6	Оформлення роботи: вступ, реферат, висновки	04.05-12.05.25	Виконав
7	Розробка демонстраційних матеріалів	12.05-23.05.25	Виконав

Здобувач вищої освіти

(підпис)

Владислав ШКЛЯРЕНКО

Керівник

кваліфікаційної роботи

(підпис)

Антон ШАНТИР

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття освітнього ступеня бакалавра: 57 стор., 14 рис., 20 джерел.

Мета роботи – розробка ефективної системи розпізнавання емоцій на фото та відео за допомогою методів глибокого навчання для практичного застосування у вебдодатку.

Об'єкт дослідження – процеси аналізу та розпізнавання емоцій за допомогою алгоритмів глибокого навчання на фото та відео.

Предмет дослідження – процеси інтеграції та використання нейронної мережі для розпізнавання емоцій у вебдодатку, включаючи обробку фото та відео, передобробку даних, візуалізацію результатів та організацію взаємодії між фронтендом та бекендом.

Короткий зміст роботи: У межах роботи досліджено сучасні підходи до побудови систем емоційного аналізу, проаналізовано актуальні моделі глибокого навчання та публічні набори даних. В якості навчального датасету обрано FER2013, що дозволяє класифікувати 7 базових емоцій людини. Реалізовано клієнт-серверну вебсистему, де фронтенд побудовано з використанням HTML, CSS та JavaScript, а серверна частина – на мові Python із застосуванням фреймворку Flask. Система забезпечує обробку як зображень, так і відеофайлів, виконує автоматичне виявлення обличчя за допомогою попередньо навченого детектора (SSD), а також передбачає візуалізацію результатів розпізнавання у режимі реального часу. Запропоноване рішення протестовано на прикладах користувацького контенту та адаптовано до можливого подальшого використання в освітніх, соціальних або медичних платформах.

КЛЮЧОВІ СЛОВА: CNN, OPENCV, ГЛИБОКЕ НАВЧАННЯ, РОЗПІЗНАВАННЯ ЕМОЦІЙ, ВЕБДОДАТОК.

ABSTRACT

Text part of the bachelor's qualification work: 57 pages, 14 pictures, 20 sources.

Purpose - to develop an effective system for recognizing emotions in photos and videos using deep learning methods for practical application in a web application.

Object of research – the processes of analyzing and recognizing emotions using deep learning algorithms in photos and videos.

Subject of research – the processes of integrating and using a neural network for emotion recognition in a web application, including photo and video processing, data preprocessing, visualization of results, and organization of interaction between the frontend and backend.

Summary of the work: The present work investigates modern approaches to building emotional analysis systems, analyzes current deep learning models and public datasets. The FER2013 dataset was chosen as a training dataset, which allows to classify 7 basic human emotions. A client-server web system is implemented, where the frontend is built using HTML, CSS, and JavaScript, and the server side is built in Python using the Flask framework. The system provides processing of both images and video files, performs automatic face detection using a pre-trained detector (SSD), and provides real-time visualization of recognition results. The proposed solution is tested on examples of user-generated content and adapted for possible further use in educational, social, or medical platforms.

KEYWORDS: CNN, OPENCV, DEEP LEARNING, EMOTION RECOGNITION, WEB APPLICATION.

ЗМІСТ

ВСТУП.....	9
1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ	11
1.1 Актуальність задачі розпізнавання емоцій.....	11
1.2 Теоретичні основи розпізнавання емоцій.....	12
1.3 Методи глибокого навчання у розпізнаванні образів	19
1.4 Огляд існуючих систем розпізнавання емоцій	22
1.5 Висновки до розділу 1	26
2 РОЗРОБКА, НАВЧАННЯ ТА ОЦІНКА МОДЕЛІ РОЗПІЗНАВАННЯ ЕМОЦІЙ	28
2.1 Підготовка даних.....	28
2.2 Структура нейронної мережі	30
2.3 Налаштування гіперпараметрів	31
2.4 Процес навчання моделі	33
2.5 Використані бібліотеки та інструменти.....	34
2.6 Оцінка точності моделі.....	35
2.8 Порівняння з іншими моделями	38
3 РОЗРОБКА ВЕБДОДАТКУ ТА ДЕМОНСТРАЦІЯ ГОТОВОГО ПРОЄКТУ	40
3.1. Постановка завдання.....	40
3.2. Вибір технологічного стеку	41
3.3. Архітектура вебдодатку.....	42
3.4. Опис реалізації фронтенду	43
3.5. Опис реалізації бекенду	45
3.7. Демонстрація роботи проєкту.....	47
3.9. Можливості для розширення	49
ВИСНОВКИ.....	51
ПЕРЕЛІК ПОСИЛАНЬ	53
ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ	55

ВСТУП

У сучасному світі багато речей, які донедавна сприймалися як фантастика, стрімко впроваджуються в реальне життя; прикладом є автоматизоване розпізнавання емоцій, що перестало бути недосяжною технологією та зайняло важливе місце в галузі комп'ютерного зору. Система розпізнавання емоцій на фото та відео – це технологія, яка аналізує вирази обличчя для визначення емоційного стану людини за допомогою методів штучного інтелекту і комп'ютерного зору. Така система належить до сфери афективних обчислень – напрямку інформатики, що зосереджений на розробці пристроїв і програм, здатних розпізнавати, інтерпретувати та реагувати на емоції людини. Сучасні технології дозволяють автоматично визначати широкий спектр емоцій як на статичних зображеннях, так і в режимі реального часу з відеопотоку. Це відкриває можливості для застосування систем розпізнавання емоцій у різних практичних сферах.

FER-системи (від англ. Facial Emotion Recognition) вже знаходять застосування у багатьох галузях. Зокрема, інтеграція таких технологій у користувацькі інтерфейси та служби підтримки клієнтів робить взаємодію людини з комп'ютером більш інтуїтивною та “емпатичною” – машина може реагувати на емоційний стан користувача. У сфері безпеки аналіз емоцій допомагає виявляти аномальну чи агресивну поведінку, а в медицині системи розпізнавання емоцій застосовують для моніторингу психоемоційного стану пацієнтів. Значний науковий та комерційний інтерес до таких систем підтверджується великою кількістю досліджень і проєктів у цій галузі. Водночас на сьогодні реалізовано порівняно небагато подібних систем, причому більшість із них потребують спеціалізованого обладнання та високих обчислювальних ресурсів. Додатковою перешкодою є чутливість існуючих рішень до неякісних, розмитих зображень, а також їх закритість (пропрієтарність), що ускладнює доопрацювання або вдосконалення таких систем. Таким чином, існує потреба у створенні більш доступної, гнучкої та ефективної системи автоматичного розпізнавання емоцій, що й обумовлює актуальність обраної теми.

Метою даного дослідження є розробка високоякісної та ефективної системи автоматизованого розпізнавання емоцій за виразом обличчя на фотографіях і відео, здатної працювати в реальному часі. Запропонована система повинна використовувати сучасні методи глибокого навчання (згорткові нейронні мережі) для точного визначення емоційного стану людини та забезпечувати зручний інтерфейс для практичного застосування результатів аналізу.

Для досягнення поставленої мети необхідно виконати такі завдання:

1. Проаналізувати сучасні методи та засоби розпізнавання емоцій на обличчі людини, вивчити існуючі системи та підходи в цій галузі.
2. Обґрунтувати вибір і підготувати набір даних FER-2013 для навчання моделі.
3. Розробити архітектуру згорткової нейронної мережі та реалізувати модель розпізнавання емоцій мовою Python із використанням бібліотек TensorFlow/Keras.
4. Навчити й оптимізувати модель на вибраному наборі даних FER-2013, оцінити точність класифікації емоцій та при необхідності вдосконалити модель.
5. Створити вебдодаток для емоційного аналізу зображень і відео – розробити інтерфейс користувача (HTML/CSS/JS) і серверну частину на Flask (Python).
6. Протестувати роботу розробленої системи на контрольному наборі зображень та в реальних умовах відеопотоку, проаналізувати результати та визначити точність і надійність розпізнавання емоцій.

1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Актуальність задачі розпізнавання емоцій

У сучасному суспільстві цифрові технології проникають у всі сфери життя: освіту, охорону здоров'я, бізнес, розваги. Однак, попри технічний прогрес, комп'ютери залишаються емоційно «глухими» — вони не розуміють настрій, не адаптують відповідь до емоційного стану людини. Саме це створює актуальність дослідження в напрямку розпізнавання емоцій, зокрема — за допомогою зображень обличчя.

Емоційний інтелект і його роль

Людські емоції — важлива частина невербального спілкування. Вони формують контекст, передають наміри, допомагають інтерпретувати зміст сказаного. Емоційний інтелект (EQ) є здатністю розпізнавати, аналізувати та реагувати на емоційний стан інших. У системах штучного інтелекту ця функція реалізується через технології так званого "Affective Computing" (емоційних обчислень), де комп'ютерні алгоритми вчаться інтерпретувати сигнали, пов'язані з емоціями.

Такі сигнали включають:

- вираз обличчя,
- інтонацію голосу,
- текст (аналіз емоційної окраски слів),
- фізіологічні параметри (пульс, температура, електропровідність шкіри).
- Реальні приклади застосувань
- Сьогодні емоційний ШІ вже знаходить застосування в багатьох сферах:
- Медицина: Системи, що аналізують настрій пацієнта під час онлайн-консультацій, використовуються для попередження депресії та вигорання.
- Освіта: Онлайн-платформи відслідковують увагу та емоційний стан учня для адаптації матеріалу.
- Маркетинг: Системи типу Affectiva аналізують реакції людей на рекламу та продукти.

- Безпека: Технології використовуються для виявлення тривожної поведінки у відеоспостереженні.

- Автомобільна індустрія: Автомобілі нового покоління фіксують стан водія (сонливість, стрес) для запобігання аваріям.

Технічні та етичні виклики

Розпізнавання емоцій — це складне завдання з огляду на велику варіативність у виразах облич, культурні відмінності в прояві емоцій, вплив зовнішніх факторів (освітлення, ракурс тощо).

Окрім технічних труднощів, виникають і етичні питання:

- Наскільки правомірним є аналіз емоцій без згоди користувача?
- Чи не стане така система інструментом маніпуляції?
- Як забезпечити конфіденційність емоційних даних?

У зв'язку з цим деякі країни, зокрема ЄС, вводять обмеження на використання Emotion AI у сферах освіти, поліції та роботі з дітьми.

Аналіз емоцій має потенціал зробити цифрові технології людянішими, зручнішими та безпечнішими. Проте для ефективного впровадження потрібен баланс між технологічними можливостями та етичними стандартами.

1.2 Теоретичні основи розпізнавання емоцій

Автоматичне розпізнавання емоцій (англ. Emotion Recognition) – це складова афективних обчислень, що має на меті ідентифікацію емоційного стану людини на основі аналізу даних різних типів. Зокрема, розпізнавання емоцій за виразом обличчя (Facial Emotion Recognition, FER) розглядається як задача комп'ютерного зору, в якій виявляються та класифікуються емоційні вирази людини. Такі системи знаходять застосування в освіті, охороні здоров'я, безпеці, маркетингу та інших сферах, де важлива автоматична інтерпретація невербальних сигналів. Окрім міміки, емоції можуть передаватися через голос (інтонації, тембр, ритм мовлення) та через текст (слова, лексичні та семантичні ознаки). Кожен із цих каналів несе частину інформації про емоційний стан і може аналізуватися окремо або у поєднанні (мультимодальні системи). В цьому підрозділі основну

увагу приділено технічним основам розпізнавання емоцій за обличчям, але також коротко розглянуто підходи до аналізу емоцій в голосі та тексті.

Моделі представлення емоцій: дискретна та дименсійна

Побудова автоматичних систем розпізнавання емоцій спирається на фундаментальні моделі представлення емоцій в психології. Однією з найвідоміших є дискретна модель Пола Екмана, згідно з якою існує невелика кількість базових емоцій, спільних для всіх культур. Класично виділяють шість базових емоцій: радість, сум, гнів, страх, здивування та огида (деякі варіації моделі додають сьому емоцію – відразу або презирство, а також нейтральний стан). Екман показав, що мімічні прояви цих емоцій є універсальними – вирази обличчя для базових емоцій розпізнаються людьми незалежно від культурного середовища. Крім того, на основі комбінації базових емоцій або варіювання їх інтенсивності можуть виникати похідні емоційні стани (наприклад, поєднання «здивування» і «радість» дає емоцію приємного подиву). Для опису мімічних проявів Екманом та Фрізенем був запропонований Facial Action Coding System (FACS) – система кодування рухів обличчя, яка містить 46 так званих одиниць дії (Action Units, AU) для різних рухів мімічних м'язів. Кожна емоція в моделі Екмана може бути охарактеризована певним набором активованих AU; наприклад, для емоції «щастя» характерна активація м'язів, що піднімають кутики губ (AU12), тощо. FACS знайшла застосування в ранніх системах автоматичного розпізнавання емоцій, де алгоритми намагалися спочатку визначити одиниці дії з зображення обличчя, а вже на їх основі – емоцію. Альтернативний підхід до опису емоцій – дименсійна модель, яка розглядає емоційний стан як точку в безперервному просторі кількох основних вимірів. Найбільш поширена двовимірна модель валентності–активації (або valence–arousal). В цій моделі валентність відображає полярність емоції (наскільки вона позитивна чи негативна за забарвленням), а активація (arousal) – інтенсивність або збудженість стану (висока – емоція з сильним збудженням, низька – спокійні, мляві стани). Таким чином, будь-яку емоцію можна розмістити на площині цих двох координат. Наприклад, радість має позитивну валентність і високий рівень активації, тоді як

спокій або умиротворення – позитивну валентність, але низьку активацію; гнів – негативна валентність і висока активація, а смуток – негативна валентність при низькій активації. Дименсійний підхід дозволяє кількісно оцінювати емоційні стани і особливо корисний при аналізі безперервних змін емоції у часі. У сфері комп’ютерного розпізнавання емоцій існують системи, які на виході дають не категорію емоції, а координати (або класи) валентності та активації. Проте таких розробок менше, і більшість практичних реалізацій все ж оперує дискретними класами емоцій. Варто зазначити, що сучасні нейропсихологічні дослідження ставлять під сумнів абсолютну універсальність виразів обличчя. Згідно з мета-аналізом Лізи Фельдман Барретт, позбавленим контексту зображення обличчя може бути недостатньо для точного визначення внутрішнього емоційного стану людини. Тим не менш, більшість існуючих систем автоматичного розпізнавання емоцій базуються на припущенні про наявність визначеного набору універсальних емоційних категорій, що відбиваються на обличчі і можуть бути розпізнані алгоритмами. Саме ці припущення лежать в основі технічних підходів, розглянутих далі.

Розпізнавання емоцій за виразом обличчя

Обличчя людини є основним джерелом інформації про її емоції. Система розпізнавання емоцій за зображенням обличчя зазвичай працює в декілька етапів. Спочатку здійснюється детекція обличчя на зображенні (виявлення області, де знаходиться обличчя). Далі з цієї області виділяються інформативні ознаки (рис), на основі яких алгоритм приймає рішення про емоцію – цей крок часто називають екстракцією ознак. Завершальний етап – власне класифікація емоції, тобто віднесення виразу обличчя до однієї з передбачених категорій (наприклад, до однієї з базових емоцій). Така послідовність дій є стандартною для більшості сучасних FER-систем. На рис. 1.1 наведено схему типової архітектури глибокої нейронної мережі для розпізнавання емоцій на основі зображення обличчя.

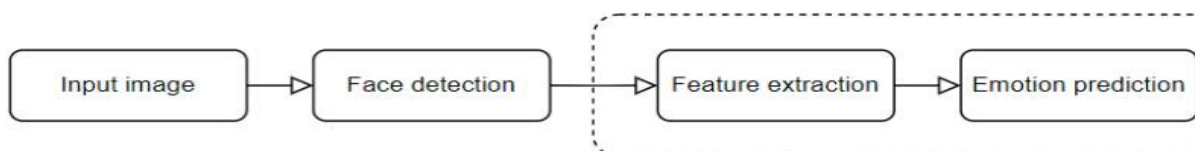


Рис. 1.1 Структурна схема процесу розпізнавання емоцій за зображенням

На ранніх етапах розвитку FER використовувалися традиційні методи комп'ютерного бачення та машинного навчання. Вони полягали у ручному конструюванні ознак: наприклад, вимірювалися геометричні характеристики міміки (положення та форма ключових точок обличчя – кути губ, брів, очей тощо) або апперцептивні характеристики (текстурні ознаки певних областей обличчя, отримані, наприклад, фільтрами Габора чи локальними бінарними шаблонами). Отримані таким чином ознаки подавалися на вхід класифікатору (наприклад, SVM або Random Forest), що визначав емоцію. Недоліком таких підходів була залежність від якості ручних ознак і складність врахування всіх варіацій виразів. Прорив у точності розпізнавання був досягнутий із запровадженням методів глибокого навчання, зокрема глибоких нейронних мереж типу згорткових. Convolutional Neural Networks (CNN), або згорткові нейронні мережі, виявилися особливо ефективними для задач аналізу зображень, оскільки автоматично навчаються виділяти релевантні ознаки з сирих піксельних даних. Типова архітектура CNN для класифікації зображень складається з кількох блоків згорткових шарів і шарів підвибірки (пулінгу), які поступово виділяють все більш складні візуальні ознаки, а наприкінці мережі розташовано кілька повнозв'язних шарів, що інтегрують ці ознаки і видають результат через функцію SoftMax для вибору класу емоції. На відміну від класичних методів, тут немає необхідності в явному ручному конструюванні ознак – мережа сама «вчиться» оптимальним характеристикам на основі великого обсягу тренувальних прикладів. Це дозволяє охопити тонкі деталі міміки, які важко запрограмувати вручну. Упродовж останнього десятиліття розроблено багато архітектур глибоких CNN, успішно застосованих до розпізнавання емоцій за обличчям. Серед них – VGG16/VGG19 (глибокі конволюційні мережі з послідовним шаруванням згорток), різновиди ResNet (Residual Network) з механізмом прямого пропускання шарів, архітектури родини Inception (GoogLeNet) та їхні модифікації на зразок Inception-ResNet, легкі мережі типу Xception, а також новіші моделі на кшталт EfficientNet. Наприклад, ResNet запропонувала інноваційний підхід із використанням резидуальних блоків – прямих з'єднань, що оминають кілька шарів і додають вихід попереднього

блоку до наступного . Це розв’язує проблему згасання градієнта при тренуванні дуже глибоких мереж і дозволяє успішно тренувати моделі із сотнями шарів. Архітектура EfficientNet, у свою чергу, використовує збалансоване масштабування глибини, ширини і роздільної здатності моделі, що дає змогу досягати високої точності при відносно меншій кількості параметрів мережі. Нерідко ці готові моделі, попередньо натреновані на великій базі зображень (наприклад, ImageNet), застосовуються до задачі FER шляхом трансферного навчання – або як екстрактори ознак (із замороженими вагами, де мережа використовується для отримання векторів ознак обличчя, а класифікатор навчається окремо), або шляхом тонкого донавчання (fine-tuning) моделі на спеціалізованому датасеті з емоціями . Використання трансферного навчання суттєво прискорює розробку та підвищує якість, оскільки великі нейронні мережі вже «знають» багато базових ознак і патернів зображень , і для адаптації до емоційної класифікації потрібно менше даних та обчислювальних ресурсів. Під час перетворення зображення обличчя в рішення про емоцію важливо сфокусуватися на тих особливостях, які найбільше відображають емоційний стан. Дослідження показують, що інформативними є насамперед динамічні риси обличчя – рухомі області навколо очей, брів, рота, щік . Саме деформація цих областей (мімічні зміни) відповідає проявам емоцій. Натомість статичні риси (форма обличчя, риси, які не змінюються – наприклад, відстань між очима, структура шкіри тощо) менш важливі для визначення конкретної емоції . З іншого боку, автоматичне розпізнавання емоцій стикається з низкою технічних викликів. Серед них – варіативність пози і кута повороту голови відносно камери, різне освітлення та тіні, що впливають на візуальне сприйняття виразу, можливі перешкоди (окуляри, маски, часткове закриття обличчя руками тощо), а також індивідуальні відмінності у міміці різних людей . Для підвищення надійності систем FER розробляються методи попередньої обробки зображень (нормалізація яскравості, видалення шумів, вирівнювання положення обличчя по ключових точках), а також алгоритми, стійкі до зазначених факторів (наприклад, використання аугментації даних при навчанні, або застосування спеціальних

шарів нейронних мереж, інваріантних до зсувів і поворотів). Попри ці складнощі, сучасні системи розпізнавання емоцій за обличчям вже демонструють високі результати на тестових наборах даних, наближаючись до точності, порівнянної з людиною, в умовах контрольованих зображень. В «диких» умовах (необмежене середовище, різні люди, спонтанні емоції) задача лишається відкритою для подальших досліджень, однак поєднання вдосконалених моделей глибокого навчання та великих масивів даних дає поступове підвищення ефективності розпізнавання.

Розпізнавання емоцій за голосом

Вокальні характеристики мовлення є важливим носієм емоційної інформації. Розпізнавання емоцій за голосом (Speech Emotion Recognition, SER) аналізує акустичні сигнали мовлення з метою визначення емоційного стану мовця. Люди інтуїтивно розпізнають емоції за інтонацією, тоном, гучністю, темпом мовлення – наприклад, підвищений тон і швидкий темп можуть свідчити про збудження або гнів, тоді як сповільнена, тиха мова – про смуток тощо. Автоматичні системи прагнуть моделювати ці здібності. З технічної точки зору, процес SER зазвичай містить три основні стадії: попередню обробку аудіосигналу, виділення ознак та класифікацію. На першому етапі аудіо проходить очистку (фільтрацію шумів, нормалізацію гучності) і сегментацію (виділення фрагментів, що відповідають окремим висловлюванням або словам). Далі з обробленого сигналу обчислюються ознаки, які кількісно описують голос: це можуть бути спектральні ознаки (наприклад, коефіцієнти мел-чепстрального перетворення – MFCC), просодичні ознаки (статистика висоти тону – pitch, енергії сигналу, тембру) тощо. Поширеним підходом є перетворення звукового сигналу у візуальне представлення – спектрограму або мелспектрограму, яка відображає частотний склад звуку в часі. Отримана спектрограма може подаватися на вхід згорткової нейронної мережі аналогічно до звичайного зображення. Паралельно можуть використовуватися й традиційні числові ознаки (наприклад, середнє та діапазон частоти основного тону, швидкість мовлення та ін.). На останньому етапі класифікації застосовуються або класичні алгоритми (наприклад, SVM, якщо

ознаки вибрані вручну), або – що нині домінує – глибокі нейронні мережі. Архітектури глибокого навчання для SER часто комбінують згорткові та рекурентні мережі: CNN можуть використовуватися для автоматичного виділення ознак із спектрограми, після чого рекурентна мережа (наприклад, LSTM) аналізує часову динаміку цих ознак, враховуючи послідовність мовлення. Такий підхід дозволяє вловити як статичні спектральні характеристики голосу, так і їх зміни у часі (інтонаційні контури, паузи тощо). Як і у випадку з обличчям, глибоке навчання значно підвищило точність визначення емоцій з мовлення у порівнянні з ранніми методами. В дослідженнях повідомляється про точність класифікації емоцій за голосом на рівні 70–90% для окремих датасетів, що містять виразні емоційні висловлювання. Проте ці показники можуть знижуватися в умовах реального мовлення, де емоції проявляються тонше. Для підвищення надійності SER активно досліджуються методи ансамблю моделей (комбінація кількох нейронних мереж), а також мультимодальні підходи – одночасний аналіз голосу і зображення обличчя, що дає більш повну картину емоційного стану.

Розпізнавання емоцій у тексті

Ще одним напрямом є розпізнавання емоцій за текстовими даними. Тут завдання полягає в тому, щоб автоматично визначити емоційне забарвлення уривку тексту – наприклад, повідомлення у соцмережі, відгуку користувача або репліки в чаті. Текст не передає міміку чи інтонацію, але емоції можуть бути виражені лексично (вибір слів), синтаксично (побудова речень), а також за допомогою символів (наприклад, знаків пунктуації, емодзі). Ранній підхід до цієї задачі був тісно пов'язаний з аналізом тональності (sentiment analysis), де визначалося лише позитивне, негативне чи нейтральне забарвлення тексту. Нині завдання сформульовано ширше: визначити конкретну емоцію або набір емоцій, що виражені у тексті (гнів, радість, страх, здивування тощо). Для цього застосовуються методи обробки природної мови (NLP) у поєднанні з машинним навчанням. Сучасні дослідження демонструють значний прогрес у цій галузі завдяки інтеграції глибоких нейронних мереж і особливо трансформерних моделей (наприклад, BERT, RoBERTa, GPT тощо), які здатні враховувати

контекст слів і виявляти навіть тонкі емоційні відтінки в тексті. Такі моделі навчаються на великих корпусах текстових даних і можуть класифікувати уривки тексту за декількома емоційними класами з високою точністю. Окрім безпосереднього визначення емоцій, додатково використовуються техніки аналізу полярності (визначення позитивності/ негативності висловлювання) та облік емотиконів/емодзі, що дозволяє підвищити точність і врахувати приховані сигнали емоцій. Наприклад, наявність у тексті «!!!» може вказувати на сильне збудження (емоція здивування або гніву), слова в рідкісно високому регістрі можуть свідчити про гнів тощо. Врахування подібних факторів покращує якість автоматичного аналізу. Застосування технологій розпізнавання емоцій у тексті вже знаходить місце в аналізі відгуків клієнтів (щоб виявити емоційне задоволення чи фрустрацію), моніторингу соціальних мереж (виявлення масових настроїв, наприклад, панічних чи агресивних тенденцій) та в діалогових системах (для створення більш чуйних чат-ботів, які реагують на емоційний стан користувача). Втім, обробка тексту з метою визначення емоцій залишається складною через багатозначність мови, іронію, сарказм та культурні особливості мовців. Тому активно досліджуються підходи, що поєднують аналіз тексту з іншими модальностями (наприклад, врахування тону голосу при аналізі розшифровки мовлення), а також розширені моделі розуміння контексту (онтології емоцій, моделювання знань про ситуацію) для більш точного виявлення прихованих емоційних смислів.

1.3 Методи глибокого навчання у розпізнаванні образів

Основи згорткових нейронних мереж (CNN): архітектура, принцип роботи

Методи глибокого навчання, зокрема згорткові нейронні мережі (Convolutional Neural Networks, CNN), відіграють ключову роль у сучасному комп'ютерному зорі, особливо у задачах розпізнавання образів. На відміну від класичних методів комп'ютерного бачення, що потребували ручного проектування ознак, CNN автоматично навчаються витягувати релевантні ознаки

із зображення, використовуючи послідовність згорткових та інших спеціалізованих шарів. На рис. 1.2 зображено базову структуру CNN

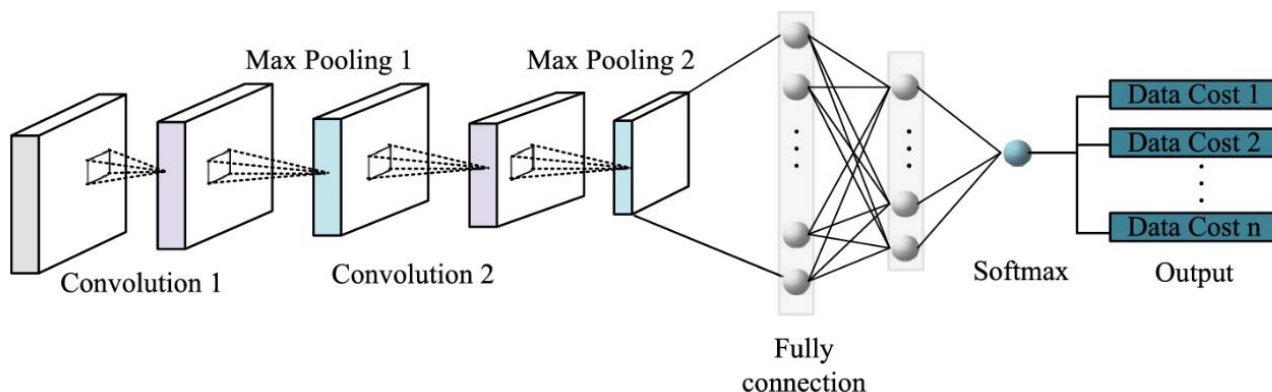


Рис. 1.2 Структура простої моделі CNN

Типова структура CNN включає:

- *Згорткові шари (Conv2D)*: ці шари відповідають за виявлення локальних ознак у зображенні, таких як краї, текстури, форми. Вони застосовують набір фільтрів (ядра згортки), які проходять по всьому зображенню та створюють карти ознак. Кожен фільтр навчається автоматично і спеціалізується на певному типі ознак. Ці шари зберігають просторову інформацію, що особливо важливо для завдань класифікації зображень.

- *Функції активації (ReLU)*: після згортки застосовується функція активації, яка вводить нелінійність у модель. Найчастіше використовується ReLU (Rectified Linear Unit), яка встановлює всі від'ємні значення в нуль, а позитивні залишає без змін. Це дозволяє ефективно передавати градієнти при навчанні та робить мережу здатною моделювати складні залежності у даних.

- *Шари підвибірки (MaxPooling)*: ці шари зменшують розмірність просторового представлення ознак, виділяючи найбільш значущі значення. Наприклад, у фільтрі 2×2 вибирається максимум з кожної області. Це дозволяє зменшити кількість параметрів, знизити ризик перенавчання і зробити мережу більш стійкою до змін позиції об'єкта на зображенні.

- *Нормалізація (BatchNormalization)*: цей шар нормалізує значення активацій у межах одного пакету (batch), що прискорює та стабілізує навчання. Він зменшує внутрішнє ковзання (internal covariate shift) — проблему, коли розподіл вхідних

даних змінюється під час навчання. Завдяки цьому можна використовувати вищу швидкість навчання і швидше досягати збіжності.

- *Регуляризація (Dropout)*: використовується як метод запобігання перенавчанню. Під час тренування випадковим чином деактивується певна частина нейронів, що змушує модель не залежати від окремих шляхів проходження інформації. Це покращує здатність моделі до узагальнення.

- *Повнозв'язні шари (Dense)*: після згорткових і підвибіркових шарів ознаки згладжуються у вектор і подаються до повнозв'язних шарів. Тут кожен нейрон з'єднаний з усіма попередніми, і на цьому етапі модель приймає остаточне рішення про клас об'єкта. Ці шари часто використовуються разом із функціями активації та регуляризацією.

- *Вихідний шар з Softmax*: цей шар перетворює вихід моделі на ймовірності, що відображають ступінь впевненості моделі у приналежності зображення до кожного з можливих класів. Вибирається той клас, для якого ймовірність найбільша.

Завдяки таким особливостям CNN стали стандартом де-факто для задач класифікації зображень, включаючи розпізнавання емоцій на обличчях.

Популярні CNN-архітектури

З моменту появи перших CNN було створено низку архітектур, які суттєво вплинули на розвиток глибокого навчання:

- *LeNet-5 (1998)*: одна з перших CNN, застосована до задачі розпізнавання рукописних цифр (MNIST); проста та компактна архітектура.

- *AlexNet (2012)*: переможець конкурсу ImageNet, яка популяризувала використання GPU для тренування, ReLU та Dropout.

- *VGGNet (2014)*: відзначалася простотою — лише фільтри 3×3 у глибокій мережі до 19 шарів; недолік — велика кількість параметрів.

- *GoogLeNet / Inception (2014)*: запровадила концепцію інцепшин-блоків з фільтрами різного розміру у паралельних гілках.

- *ResNet (2015): революціонізувала тренування глибоких мереж завдяки залишковим зв'язкам (residual connections), які дозволили навчати мережі з понад 100 шарами.*

- *Xception (2017): застосовує глибинно-роздільні згортки (depthwise separable convolutions), що зменшують кількість параметрів без втрати точності.*

Ці архітектури відрізняються за глибиною, обчислювальною складністю та точністю. Наприклад, ResNet-50 має 25 млн параметрів, тоді як VGG-16 — понад 138 млн, при тому, що ResNet зазвичай показує вищу точність.

Методи підвищення ефективності CNN

Сучасні підходи до навчання CNN включають:

- Аугментацію даних: штучне збільшення обсягу даних шляхом обертання, масштабування, віддзеркалення тощо.

- Регуляризацію: Dropout, L2-нормалізація, BatchNormalization.

- Оптимізатори: Adam, SGD із momentum, RMSProp.

- Ранню зупинку (EarlyStopping): завершення навчання при зниженні продуктивності на валідаційних даних.

- Планування learning rate: адаптація швидкості навчання під час тренування (ReduceLROnPlateau тощо).

Завдяки поєднанню ефективних архітектур і сучасних методів навчання, CNN залишаються найкращим вибором для задач класифікації зображень, зокрема для розпізнавання емоцій на обличчях.

1.4 Огляд існуючих систем розпізнавання емоцій

Розпізнавання емоцій за виразами обличчя є одним із ключових напрямів «affective computing». За останні роки з'явилося безліч відкритих і комерційних рішень для автоматичного аналізу емоційного стану користувача. Більшість систем націлена на класифікацію базових емоцій (радість, сум, гнів, здивування, страх, огиду, спокій тощо) і тренувана на публічних наборах даних (FER2013, JAFFE, CK+ тощо). Проте ці класичні набори містять здебільшого «показові» вирази обличчя (стандартні постановочні емоції), тому на них досягається висока

точність (близько 90%). Водночас у реальних умовах (спонтанні вирази, неконтрольоване освітлення, різні ракурси) точність суттєво знижується . Нижче розглянуто окремі відомі системи з точки зору алгоритмів, точності, типових кейсів, обмежень та доступності.

OpenFace 3.0

OpenFace 3.0 – це відкритий набір інструментів (open-source) для комплексного аналізу обличчя (виявлення ключових точок, Action Units, оцінка погляду і емоцій) . Система побудована на спільній мультимоделі з глибокими згортковими мережами, навченими на великих датасетах (наприклад, AffectNet). У версії 3.0 реалізована класифікація 8 емоцій. Вона демонструє приблизно 60% точності на тесті AffectNet (що є конкурентним показником порівняно з SOTA-моделями) . OpenFace особливо добре працює з нефронтальними обличчями (важливо для «in-the-wild» даних) . Застосування: реальний час, мультимодальні системи, HCI, робототехніка . Мінімальні обмеження пов'язані з обчислювальними ресурсами (модель ~29.4М параметрів і оптимізована під CPU) і тим, що на практиці потребує якісного кадру обличчя. OpenFace – повністю відкрите програмне забезпечення (GitHub, MIT-ліцензія), що робить його популярним у наукових дослідженнях.

DeepFace (Serengil et al.)

DeepFace – відкрита Python-бібліотека для розпізнавання осіб та аналізу атрибутів (вік, стать, емоції, раса) . Вона є «обгорткою» над декількома передтренованими нейромережами (VGGFace, FaceNet, OpenFace, DeepID, ArcFace тощо) . DeepFace вміє класифікувати емоції за зображенням обличчя, але точність залежить від обраної моделі та даних. Цей фреймворк підходить для розробників і дослідників, які хочуть швидко додати аналіз емоцій; він безплатний і відкритий (MIT-ліцензія). Проте його обмеження ті ж, що й у самих нейромереж: погана якість зображення чи екстремальні ракурси знижують точність, а додаткова обробка виконується поза межами бібліотеки.

FaceReader (Noldus)

FaceReader – комерційна система (SDK/додаток) від компанії Noldus для автоматизованого аналізу міміки та емоцій за системою FACS. Вона розпізнає 6 базових емоцій (радість, сум, гнів, здивування, страх, огиду), а також «байдужість» та зневагу. Заявлена точність – ~99% для основних виразів (на стандартизованих зображеннях зі спеціальних баз). Незалежні дослідження показують, що FaceReader має близько 88–90% точності (наприклад, щастя розпізнається ~96%, а гнів – ~76%) . Система використовує машинне навчання (ймовірно CNN/ SVM) та обробляє відео або потік камерного відео. Удалі кейси застосування: наукові дослідження (психологія, маркетинг), оцінка користувацьких інтерфейсів, HCI, освіта (моніторинг емоцій студентів), охорона здоров'я (аналіз виразу пацієнтів) . До обмежень належать висока вартість ліцензії, необхідність добре освітленого кадру з чітким обличчям (повороти голови й затінення знижують якість) і фокус на постановочних виразах (реальна точність на спонтанних виразах може бути нижчою) .

Affectiva Affdex SDK

Affectiva (тепер частина Smart Eye) пропонує SDK та API Affdex для аналізу емоцій на основі відеопотоку або фото. Використовується глибоке навчання: Affdex аналізує FACS-фічі та оцінює ймовірності базових емоцій . На стандартизованих наборах Affdex показує доволі високу точність (~88–90%), проте, як і інші системи, демонструє зниження успішності при природних «несценарних» виразах . Застосування широке: маркетингові дослідження, UX/HCI, освіта (наприклад, Affdex використовується для оцінки настрою учнів), а також автомобільні системи. Останні знаходять широке застосування: Affectiva Automotive AI відстежує емоції і когнітивні стани водія/пасажирів, адаптує інформаційно-розважальну систему та умови салону (наприклад, змінює маршрут чи підсвічування, враховуючи занепокоєння пасажирів) . Основні обмеження: комерційна ліцензія (API/SDK платні), закритий код алгоритмів, чутливість до культурних відмінностей і зовнішнього середовища. Варто зауважити, що всі

подібні сервіси попереджають, що результати – це лише оцінка емоцій за виглядом і не визначають внутрішній стан людини .

Microsoft Azure Face API (Emotion)

Microsoft Azure Face API (раніше окремий Emotion API) – хмарний сервіс розпізнавання облич і атрибутів, що включає й класифікацію емоцій. Модель базується на кастомних глибоких CNN і була натренована, серед іншого, на розширеній версії FER2013 (FER+), де зібрано численні мітки на одних і тих же зображеннях . Сервіс повертає ймовірності 8 категорій (зокрема радість, гнів, сум, страх тощо). За даними зовнішнього тестування, Azure API майже безпомилково визначає «щасливі» і «нейтральні» вирази (до 100% у деяких тестах), але має менше успіху із «страхом» (лише ~6% розпізнавань) та «гнівом» (~30% у наведеному прикладі) . Алгоритми доволі ефективні для чітких зображень, але сильніше чутливі до кута обличчя та демонстративності виразу. Застосовують Azure Face API в широких областях: від аналізу користувацької поведінки і HCI до освіти і охорони здоров'я. Обмеження: потребує з'єднання з хмарою Azure та оплати за запити (згідно з тарифами Azure), а також залежність від наявності сервісу (Microsoft об'єднала Emotion API в Face API). До того ж, як і інші алгоритми, він більше тренований на постановочних зображеннях, тому для «вулиці» можуть знадобитися додаткові налаштування.

Інші рішення

Існує багато інших сучасних інструментів (Google Cloud Vision, IBM Watson Visual Recognition, VisageSDK, Kairos тощо), що мають модулі розпізнавання емоцій. Наприклад, Google Cloud Vision може оцінювати ймовірність «joy/sorrow/anger/surprise», хоча їх набір емоцій трохи відрізняється. Багато систем (як платних, так і відкритих) перевіряються на тих самих публічних датасетах (FER2013, CK+, JAFFE), звідси їх точність зазвичай 80–90% на них . Загальні проблеми галузі включають культурно-зумовлені відмінності виразів, вплив освітлення і ракурсу, а також питання етики приватності (не варто «читати думки» користувача лише за обличчям) . Нижче подано узагальнену порівняльну таблицю основних характеристик цих систем.

1.5 Висновки до розділу 1

У розділі були висвітлені основні аспекти проблеми: актуальність розпізнавання емоцій на основі виразу обличчя, її теоретичні засади, сучасні методи глибокого навчання (зокрема CNN) та наявні системи. Як зауважують дослідники, розпізнавання емоцій є фундаментальним для соціальної взаємодії і застосовується у багатьох галузях (соціальні науки, автомобільна промисловість, споживча електроніка, взаємодія людина–комп'ютер і робототехніка) . Завдяки розвитку глибокого навчання, особливо конволюційних нейронних мереж, зроблено значні прориви в цій сфері: CNN демонструють високу здатність автоматично виділяти складні ознаки зображень, що підвищує точність розпізнавання . Огляд існуючих підходів засвідчив різноманіття рішень – від науково-дослідних прототипів до комерційних API – проте чимало з них є закритими або недостатньо гнучкими для кастомізації. Основні проблеми існуючих рішень:

- Недостатня точність у неконтрольованих (природних) умовах через різноманітність поз, змінне освітлення, часткове закриття обличчя тощо .
- Орієнтація систем переважно на постановочні або лабораторні вирази: більшість моделей перевірено на штучно змодельованих зображеннях, тому вони дають гірші результати на спонтанних емоціях .
- Закритість алгоритмів більшості комерційних рішень, що унеможливорює глибоке розуміння їх роботи та обмежує можливості доопрацювання під специфічні потреби.
- Обмежена гнучкість і масштабованість існуючих систем: зазвичай вони побудовані на універсальних наборах даних і готових моделях, через що при адаптації до нових умов потрібно витрачати значні ресурси (що підтверджує необхідність інвестувати у власні дані і навчання моделі) .

З огляду на зазначені обмеження існуючих підходів обґрунтована потреба розробки власної системи розпізнавання емоцій, яка враховуватиме специфічні вимоги. Таким чином, можна зробити висновок: існуючі методи та системи розпізнавання емоцій демонструють високу ефективність у контрольованих

умовах, але мають суттєві обмеження в реальному застосуванні. Розробка власної системи з урахуванням вказаних проблем і технічних вимог дозволить створити гнучке та адаптивне рішення, яке відповідатиме потребам користувачів і актуальним стандартам індустрії.

2 РОЗРОБКА, НАВЧАННЯ ТА ОЦІНКА МОДЕЛІ РОЗПІЗНАВАННЯ ЕМОЦІЙ

2.1 Підготовка даних

Підготовка даних є ключовим етапом будь-якої системи машинного навчання, особливо якщо мова йде про зображення, які мають обмежену роздільну здатність, низьку якість і складну класифікацію, як у випадку з емоціями. Для навчання моделі було обрано відкритий датасет *FER-2013 (Facial Expression Recognition)*, який є загальноприйнятим стандартом у задачах розпізнавання емоцій. Цей набір даних був зібраний для участі у змаганні Kaggle та включає понад 35 тисяч прикладів зображень обличч, кожне з яких представлено у градаціях сірого з роздільною здатністю 48×48 пікселів.



Рис. 2.1 Приклад зображень обличч з датасету FER-2013.

Кожне зображення у FER-2013 пов'язане з міткою, яка вказує одну з семи основних емоцій: злість, відраза, страх, щастя, смуток, подив та нейтральний стан. Ці емоції класифікуються за числовими індексами від 0 до 6. Формат зберігання датасету — CSV-файл, де пікселі представлені у вигляді одного рядка з числовими значеннями, а також вказано розділ даних (навчання, валідація або тест).

Однією з ключових проблем FER-2013 є дисбаланс класів. Наприклад, категорія "відраза" має значно менше прикладів, ніж "щастя" чи "нейтральний

вираз обличчя". Цей дисбаланс може спричинити схильність моделі до переважних класів, що призведе до зниження точності на менш представлених емоціях. Для врахування цього було вжито спеціальні заходи при підготовці вибірок.

Першим кроком було розділення даних на три підмножини: навчальну (training), валідаційну (validation) та тестову (test). Такий поділ дозволяє спочатку навчати модель, потім перевіряти її якість у процесі тренування, а наприкінці — оцінити її узагальнюючу здатність на повністю нових даних.

Усі зображення були нормалізовані до діапазону значень від 0 до 1 для забезпечення стабільності та швидшої збіжності під час навчання. Це досягається шляхом поділу значень пікселів (які варіюються від 0 до 255) на 255. Також усі мітки класів були перетворені у формат one-hot encoding, тобто у вектор з одиницею на місці відповідного класу та нулями на інших позиціях. Така трансформація необхідна для використання функції втрат categorical crossentropy.

Особливу увагу було приділено аугментації даних. Оскільки кількість зображень у деяких класах є недостатньою, а також через обмежений обсяг датасету загалом, використовувалися методи штучного розширення: повороти, зсуви, масштабування, горизонтальні віддзеркалення. Ці методи дозволяють створити нові варіанти зображень, зберігаючи при цьому сутність емоційного виразу.

Важливо, що аугментація відбувалась динамічно під час навчання — щоразу, коли модель бачила зображення, воно могло мати інші характеристики (нахил, розмір, позицію). Це допомогло моделі не запам'ятовувати конкретні приклади, а навчитись розпізнавати узагальнені емоційні патерни.

У результаті підготовки було сформовано якісний навчальний набір, який забезпечив умови для побудови ефективної та стійкої моделі. Цей процес є основою всієї подальшої роботи моделі й напряду впливає на рівень її точності та узагальнення.

2.2 Структура нейронної мережі

Побудована модель розпізнавання емоцій базується на згортковій нейронній мережі (CNN) — одному з найефективніших підходів до обробки зображень. Основна мета архітектури полягає у виявленні характерних ознак обличчя, що дозволяють класифікувати емоційний стан людини.

Модель має послідовну архітектуру, де дані проходять через серію згорткових блоків, які поетапно виділяють візуальні патерни. Усього в мережі 4 таких блоки. Кожен з них містить два згорткові шари з фільтрами малого розміру (3×3), що дозволяють ефективно виявляти локальні ознаки — краї, контури, зміни текстур. Після кожного згорткового шару застосовується пакетна нормалізація, яка стабілізує розподіл активацій, прискорює навчання і підвищує точність. Активаційна функція ReLU (Rectified Linear Unit) використовується для введення нелінійності.

Після двох згорткових шарів у кожному блоці застосовується шар субдискретизації (MaxPooling), який зменшує просторові розміри карти ознак, залишаючи найінформативніші елементи. Це дозволяє значно скоротити обчислювальну складність і зробити модель стійкішою до незначних зміщень на зображенні. Для боротьби з перенавчанням після кожного пулінгу вводиться шар Dropout, який випадковим чином відключає частину нейронів у процесі навчання. Це сприяє кращому узагальненню моделі на нові приклади.

На виході зі згорткових блоків ознаки зображення перетворюються у вектор за допомогою шару Flatten. Після цього вектор передається у повнозв'язний шар з великою кількістю нейронів (512), що дозволяє моделі сформулювати уявлення про взаємозв'язки між ознаками. І тут також застосовуються нормалізація, ReLU і Dropout для стабільності й регуляризації.

Завершальним етапом є вихідний шар з сімома нейронами — за кількістю класів емоцій. На цьому шарі використовується Softmax-активація, яка перетворює вихід моделі у ймовірнісний розподіл між класами. Найвища ймовірність відповідає передбаченій емоції.

Загалом, модель містить понад 2 мільйони параметрів, що свідчить про її здатність вивчати складні патерни. Така архітектура була вибрана через її збалансованість між глибиною, здатністю до узагальнення та обчислювальною ефективністю. З урахуванням малого розміру вхідних зображень (48×48), модель є оптимальною для задачі розпізнавання емоцій як на фото, так і на відео.

2.3 Налаштування гіперпараметрів

Налаштування гіперпараметрів є критично важливим етапом створення будь-якої глибокої нейронної мережі. Правильно підібрані значення таких параметрів, як розмір пакета (batch size), кількість епох, функція втрат, оптимізатор та швидкість навчання (learning rate), безпосередньо впливають на збіжність моделі, її здатність до узагальнення та ефективність у реальних умовах.

Розмір пакета (batch size)

Було обрано значення `batch_size = 64`, тобто за одну ітерацію модель обробляє 64 зображення. Це компроміс між швидкістю навчання та стабільністю оновлення ваг. Занадто мале значення призводить до нестабільності градієнтів, а надто велике — до перевитрати пам'яті й уповільнення адаптації.

Кількість епох (epochs)

Модель навчалась протягом 50 епох, що дозволило досягти стійкого зниження функції втрат і підвищення точності на валідаційній вибірці. Це число було підбрано емпірично — після 50 епох подальше навчання не давало істотного приросту точності, а іноді навіть призводило до перенавчання.

Функція втрат та метрики

Для задачі мультикласової класифікації було використано функцію `categorical_crossentropy`. Вона підходить для проблем, де кожен приклад належить лише до одного з кількох класів. Як метрику використовували точність (accuracy), яка є найбільш інтуїтивною для інтерпретації результатів класифікації:

Оптимізатор

В якості оптимізатора було обрано Adam (Adaptive Moment Estimation), оскільки він поєднує переваги алгоритмів AdaGrad і RMSProp. Цей оптимізатор

автоматично адаптує швидкість навчання для кожного параметру, що дозволяє досягти швидкої та стабільної збіжності. Початковий коефіцієнт навчання становив 0.001, що є безпечним значенням для моделей середньої глибини. Занадто високі значення могли призвести до нестабільної поведінки моделі.

Callbacks: Збереження ваг та зупинка при перенавчанні

Під час навчання було використано механізми збереження найкращих ваг моделі (ModelCheckpoint) та ранньої зупинки (EarlyStopping), які дозволяють уникнути перенавчання (Рис. 2.2). Ці інструменти зупиняють навчання тоді, коли модель перестає покращуватись, та зберігають тільки найуспішнішу версію мережі.

```
# Рання зупинка, щоб уникнути перенавчання
early_stop = EarlyStopping(monitor='val_accuracy', patience=15,
                           verbose=1, restore_best_weights=True)
# Збереження найкращої моделі (за валідаційною точністю)
checkpoint = ModelCheckpoint('best_fer_model.h5', monitor='val_accuracy',
                             save_best_only=True, verbose=1)
```

Рис. 2.2 – Фрагмент коду, де використано механізми ModelCheckpoint та EarlyStopping.

Аугментація даних під час тренування

Для запобігання перенавчанню та покращення узагальнюючої здатності моделі під час навчання застосовувалась динамічна аугментація зображень (Рис. 2.3). Вона реалізується через генератор зображень, який кожного разу подає на вхід нові трансформовані приклади:

- випадкові повороти (до 10 градусів);
- зсув по ширині/висоті (до 10%);
- масштабування;
- горизонтальні віддзеркалення.

```

from tensorflow.keras.preprocessing.image import ImageDataGenerator
datagen = ImageDataGenerator(
    rotation_range=10,      # випадкове обертання зображення на +/-10 градусів
    width_shift_range=0.1,  # випадковий зсув по ширині
    height_shift_range=0.1, # випадковий зсув по висоті
    zoom_range=0.1,        # випадкове масштабування
    horizontal_flip=True    # горизонтальне віддзеркалення
)

```

Рис. 2.3 Фрагмент коду аугментації даних.

Ці техніки дозволили зімітувати реальні умови зйомки, включаючи зміни положення обличчя, освітлення та масштабу.

2.4 Процес навчання моделі

Після завершення етапів підготовки даних, створення архітектури та налаштування гіперпараметрів, модель була готова до навчання. Процес навчання передбачав поетапну оптимізацію ваг мережі з використанням навчальної вибірки та періодичну перевірку результатів на валідаційній вибірці.

Навчання здійснювалося за допомогою методу пакетного градієнтного спуску, реалізованого в оптимізаторі Adam. В ході кожної епохи нейронна мережа проходила по всіх доступних прикладах навчальної вибірки, обробляючи їх партіями (батчами) по 64 зображення. Після кожної партії ваги моделі оновлювались відповідно до обчислених градієнтів, а після завершення епохи проводилась оцінка на валідаційній вибірці.

Однією з особливостей реалізації було використання генератора зображень з аугментацією, що дозволяв динамічно подавати змінені приклади моделі без дублювання у пам'яті. Це дозволило ефективно збільшити різноманітність даних без суттєвого зростання обсягу.

Навчання тривало 50 епох, однак завдяки використанню механізму ранньої зупинки, воно завершилось раніше, щойно валідаційна точність перестала зростати. Одночасно спрацював механізм збереження найкращих ваг, які використовувались надалі при оцінці моделі.

Важливо зазначити, що навчання моделі вимагало певних обчислювальних ресурсів. Навіть при оптимізації архітектури й використанні невеликих зображень,

повноцінне навчання займало близько 30–60 хвилин на звичайному ноутбуку із встановленим TensorFlow (без GPU-акселерації). У середовищах з графічними прискорювачами (наприклад, Google Colab чи локальна машина з CUDA) час значно скорочується.

По завершенню навчання збережена модель експортувалась у формат .h5, що дозволяє легко інтегрувати її у вебдодаток або інші сервіси. Також можливо експортувати модель у формат TensorFlow SavedModel або ONNX для сумісності з іншими фреймворками.

2.5 Використані бібліотеки та інструменти

Реалізація системи розпізнавання емоцій ґрунтується на використанні потужного та гнучкого інструментарію з відкритим кодом. Для побудови, навчання та оцінки моделі були застосовані бібліотеки, які є стандартом у сфері машинного навчання, комп'ютерного зору та розробки додатків. Нижче подано основні з них із коротким описом їх функціональності у контексті проєкту.

TensorFlow та Keras

Бібліотека TensorFlow (а точніше — її високорівнева обгортка Keras) використовувалась для створення архітектури нейронної мережі, компіляції, навчання, а також для збереження моделі. Keras надає зручний API для побудови послідовних та функціональних моделей глибокого навчання, ідеально підходить для швидкого прототипування.

У проєкті застосовувалися такі компоненти:

- Sequential — для побудови моделі;
- Conv2D, MaxPooling2D, Dropout, Dense — основні шари;
- Adam — оптимізатор;
- to_categorical — для one-hot кодування міток;
- ModelCheckpoint, EarlyStopping — колбеки для контролю навчання.

NumPy

Ключова бібліотека для числових обчислень у Python. Вона використовувалась для обробки масивів пікселів, перетворення даних у формат,

зручний для подачі в модель, а також для маніпуляцій з багатовимірними тензорами.

Pandas

Для зчитування CSV-файлу FER-2013, де кожен запис містить текстове представлення пікселів зображення та мітку емоції, використовувалась pandas. Завдяки цій бібліотеці зручно розділяти дані за призначенням (Training, PublicTest, PrivateTest) та перетворювати текстові пікселі в числовий формат.

Мова програмування

Основною мовою реалізації обрано Python 3.11, яка є лідером у сфері машинного навчання завдяки широкій екосистемі бібліотек, простому синтаксису, активному співтовариству та великій кількості навчальних ресурсів.

2.6 Оцінка точності моделі

Після завершення навчання нейронної мережі постає критичне завдання — оцінити її ефективність. Для цього використовуються як кількісні метрики, так і візуальні інструменти аналізу результатів. У випадку класифікації емоцій точність є головною метрикою, яка дозволяє оцінити, наскільки правильно модель передбачає емоційний стан на нових, не бачених зображеннях.

Загальна точність (Accuracy)

У процесі навчання модель досягла *валідаційної точності близько 66%*, що є доволі високим результатом для задачі емоційної класифікації на обмеженому та складному датасеті FER-2013. Особливо, зважаючи на те, що навіть для людини деякі емоції важко відрізнити на статичному чорно-білому зображенні.

Для оцінки було використано метод `model.evaluate()` (Рис. 2.4):

```
best_model = tf.keras.models.load_model('best_fer_model.h5')
test_loss, test_acc = best_model.evaluate(X_test, y_test_cat, verbose=2)
print(f"Точність на тестових даних: {test_acc*100:.2f}%")
```

Рис. 2.4 – Фрагмент коду використання методу `evaluate()`.

Аналіз помилок

Модель показує найвищу точність для таких емоцій як "Щастя" і "Нейтральність", що частково пояснюється їхньою більшою представленістю в

наборі даних. Найгірші результати — для класів "Відраза" та "Страх", які мають не лише менше прикладів, а й часто візуально схожі на інші емоції. Ці результати демонструють потребу в подальшому покращенні датасету (наприклад, через балансування або використання додаткових джерел даних), а також у потенційній модифікації архітектури чи втрат.

Точність понад 65% у задачі емоційної класифікації на складному датасеті FER-2013 вважається хорошим показником, особливо з урахуванням обмежень по розміру зображень та кількості класів. При цьому, аналіз точності дає змогу побачити, в яких саме класах модель сильна, а в яких — потребує покращення. У наступному розділі буде розглянуто ці помилки глибше.

2.7 Помилки моделі та їх аналіз

Жодна модель машинного навчання не є ідеальною, і навіть при високій загальній точності залишаються помилки класифікації, які необхідно ретельно проаналізувати. Це дозволяє не лише зрозуміти слабкі сторони моделі, а й визначити напрями її покращення. У задачі розпізнавання емоцій такі помилки особливо важливі, адже схожі вирази обличчя можуть мати різні значення залежно від контексту.

Типові помилки

На основі матриці змішування та прикладів виводу моделі можна виділити кілька основних типів помилок:

- Плутання "Смутку" та "Нейтрального" стану(Рис. 2.5)

Це одна з найчастіших помилок. Причина — емоції дійсно мають подібний м'язовий патерн на обличчі, особливо у стані спокою або легкого пригнічення.

- "Страх" і "Подив"

Ці емоції мають візуальну схожість: широко відкриті очі, напружене обличчя. Якщо відсутній контекст, їх складно розрізнити навіть людині.

- "Відраза" класифікується як "Злість" або "Смуток"

Через невелику кількість прикладів у тренувальній вибірці, модель має обмежене уявлення про цю емоцію. Крім того, деякі вирази дійсно містять змішані риси.



Рис. 2.5 – Приклад плутання "Смутку" та "Нейтрального" стану.

Причини помилок

- Низька роздільна здатність зображень: усі обличчя мають розмір 48×48 пікселів, що обмежує деталізацію.
- Наявність шумів у зображеннях (розмиття, тіні, погане освітлення).
- Візуальна схожість емоцій (наприклад, “neutral” $\approx$ “sad”, “happy” $\approx$ “surprise”).
- Дисбаланс класів — модель переорієнтовується на класи з більшою кількістю прикладів.
- Мікроемоції — слабо виражені емоції, які складно зафіксувати у кадрі.

Можливі рішення

Щоб зменшити кількість помилок у майбутньому, можна розглянути такі шляхи покращення:

- Балансування датасету або застосування технік зваженого навчання (class weights).
- Покращення якості зображень — наприклад, використовувати більший розмір (96×96 або 128×128).
- Додавання контексту — аналіз послідовності кадрів (відео) замість окремих зображень.
- Застосування ensemble-моделей — об'єднання кількох моделей для більш стійких результатів.
- Заміна або модифікація функції втрат для боротьби з перекосом у розпізнаванні рідкісних емоцій.

2.8 Порівняння з іншими моделями

Щоб оцінити ефективність розробленої моделі, доцільно порівняти її з іншими популярними підходами у задачах розпізнавання емоцій. Це дозволяє краще зрозуміти, на якому рівні знаходиться розроблена система, а також визначити потенційні напрямки вдосконалення.

Базові CNN-архітектури (Simple CNN)

У багатьох початкових реалізаціях використовуються прості згорткові нейронні мережі з 2–3 шарами Conv2D, поєднаними з шарами MaxPooling, Flatten і Dense. Такі моделі можуть досягати точності 50–60% на FER-2013, проте часто схильні до перенавчання або нестабільної роботи при зміні вхідних даних. У порівнянні з ними, створена модель має більшу глибину, регуляризацію (Dropout) і BatchNormalization, що дозволяє їй стабільно досягати 66%+ точності. Вона також має значно кращу здатність до узагальнення.

ResNet, VGGNet, Xception

Сучасні моделі, такі як:

- VGGNet (особливо VGG16) демонструють точність до 68–72%, проте потребують великої кількості параметрів (понад 100 млн) та високих обчислювальних ресурсів.
- ResNet50 досягає ще кращих результатів (~72–75%) завдяки залишковим зв'язкам і глибокій структурі.
- Xception, що базується на глибоких сепарабельних згортках, показує найкращі результати на FER-2013 (до 76–78% при тонкому налаштуванні).

Моя модель — спрощений аналог Xception: вона містить багаторівневі блоки з Conv2D, BatchNormalization і Dropout, що значно підвищує її якість, не потребуючи великої кількості ресурсів. Завдяки цьому вона ідеально підходить для розгортання у вебдодатку.

Моделі на базі MediaPipe та MobileNet

Для реального часу часто використовують легкі моделі, такі як MobileNetV2, інтегровані з MediaPipe. Вони працюють дуже швидко, але зазвичай поступаються за точністю (близько 55–60%), особливо в розпізнаванні "тонких" емоцій.

Порівняно з ними, моя модель:

- точніша, особливо на складних класах;
- гнучкіша, оскільки легко адаптується до нових даних.

Розроблена модель не претендує на абсолютну точність, однак показує відмінний баланс між продуктивністю, швидкістю, точністю та зручністю інтеграції. Це робить її особливо цінною для використання у вебдодатках, де ресурси обмежені, а потреба в реальному часі критична. Вона перевершує базові моделі та мобільні архітектури, і хоча поступається великим мережам, натомість виграє в універсальності й простоті розгортання.

3 РОЗРОБКА ВЕБДОДАТКУ ТА ДЕМОНСТРАЦІЯ ГОТОВОГО ПРОЄКТУ

3.1. Постановка завдання

У межах даної дипломної роботи поставлено завдання розробити інтерактивну систему для розпізнавання емоцій людини за допомогою методів глибокого навчання. Система повинна вміти аналізувати як статичні зображення, так і відео, виявляти обличчя та визначати емоційний стан з використанням попередньо натренованої моделі.

Головна мета полягає у створенні вебдодатку, який би поєднував:

- зручний та інтуїтивно зрозумілий користувацький інтерфейс;
- інтелектуальну серверну обробку, що забезпечує високу точність розпізнавання;
- можливість локального запуску без необхідності підключення до хмарних сервісів.

Основні функціональні вимоги:

- можливість завантаження зображення або відео через вебінтерфейс;
- автоматичне виявлення обличчя на зображенні або у відеокадрі;
- визначення емоції обличчя серед попередньо визначених класів: Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral;
- візуалізація результату у вигляді прямокутника навколо обличчя з підписом відповідної емоції;
- підтримка обробки відео в реальному часі шляхом регулярної відправки кадрів на сервер.

Основні нефункціональні вимоги:

- мінімальна затримка при обробці запитів;
- адаптивний інтерфейс, зручний для використання на різних пристроях;
- стабільна робота з типовими зображеннями та відеофайлами;
- можливість локального розгортання системи без необхідності в Інтернеті.

3.2. Вибір технологічного стеку

Для реалізації системи розпізнавання емоцій на зображеннях і відео було обрано набір технологій, які забезпечують ефективну інтеграцію між вебінтерфейсом, серверною логікою та моделлю глибокого навчання. Вибір стеку здійснювався з урахуванням таких факторів, як легкість розгортання, продуктивність, доступність бібліотек і відповідність вимогам до обробки зображень у реальному часі.

Python і Flask — для серверної частини

Мова програмування Python була обрана як основна мова для серверної логіки через широку підтримку бібліотек глибокого навчання, обробки зображень та швидку розробку прототипів.

Фреймворк Flask використано для створення вебсерверу через такі переваги:

- простота налаштування й запуску;
- можливість обробки HTTP-запитів і взаємодії з фронтендом;
- легке масштабування та модульність.

OpenCV та DNN (SSD) — для виявлення облич

Для знаходження облич на зображеннях і кадрах відео використовується OpenCV у поєднанні з DNN-моделлю `res10_300x300_ssd_iter_140000.caffemodel`. Такий підхід дозволяє:

- ефективно виявляти обличчя без використання окремих сторонніх сервісів;
- працювати локально без підключення до мережі;
- легко масштабуватись на відеопотоках.
- HTML/CSS/JavaScript — для фронтенду
- Фронтенд реалізовано із використанням стандартних вебтехнологій:
- HTML — для розмітки інтерфейсу;
- CSS — для стилізації та забезпечення адаптивності;
- JavaScript — для обробки подій, завантаження файлів, взаємодії з сервером через fetch.

3.3. Архітектура вебдодатку

Система розпізнавання емоцій реалізована як класичний клієнт-серверний вебдодаток. Клієнтська частина (фронтенд) відповідає за взаємодію з користувачем, а серверна (бекенд) — за обробку вхідних зображень/кадрів, розпізнавання облич і визначення емоцій.

Загальна схема взаємодії

Нижче наведено загальну логіку роботи вебдодатку:

1. Користувач завантажує зображення або відео через інтерфейс сайту.
2. Залежно від типу файлу (зображення або відео), файл передається на сервер одним із двох маршрутів:
 - /detect_image — для фото;
 - /detect_frame — для кадру з відео.
3. На сервері:
 - Виявляються обличчя за допомогою нейромережевої моделі OpenCV (SSD).
 - Область обличчя нормалізується, перетворюється на відтінки сірого та масштабовується до 48x48.
 - Модель emotion_model.h5 визначає ймовірну емоцію.
4. Сервер повертає клієнту список виявлених облич із координатами та емоцією у форматі JSON.
5. Фронтенд відображає ці результати у вигляді прямокутників із підписами на canvas.

Компоненти архітектури

Фронтенд:

- index.html — HTML-розмітка сторінки;
 - style.css — стилізація елементів інтерфейсу;
 - script.js — логіка завантаження зображень/відео, взаємодія з сервером.
- Бекенд (app.py):
- Flask-сервер, який обробляє маршрути /detect_image, /detect_frame;
 - Обробка зображень через OpenCV;
 - Завантаження моделі emotion_model.h5 для прогнозування емоцій.

Модель:

- emotion_model.h5 — натренована глибока нейронна мережа на базі CNN (Convolutional Neural Network), яка класифікує емоції.
- Алгоритм виявлення облич:
- deploy.prototxt + res10_300x300_ssd_iter_140000.caffemodel — модель на основі SSD для детекції облич, реалізована через cv2.dnn.

Життєвий цикл запиту

Для фото:

- Користувач завантажує зображення.
- Зображення надсилається через AJAX-запит на /detect_image.
- Сервер повертає JSON з результатами.
- Результати виводяться поверх imageCanvas.
- Для відео:
- Користувач завантажує відео.
- Скрипт кожні 500 мс знімає кадр і надсилає його на /detect_frame.
- Отримані результати наносяться на overlayCanvas, який накладено поверх відеоплеєра.

3.4. Опис реалізації фронтенду

Фронтенд вебдодатку реалізовано за допомогою стандартних вебтехнологій — HTML, CSS і JavaScript. Основною метою фронтенду є забезпечення зручного способу взаємодії користувача з системою: завантаження зображення або відео, передача даних на сервер, відображення результатів розпізнавання у зрозумілому вигляді.

HTML-структура інтерфейсу

HTML-файл відповідає за основну розмітку сторінки. У ньому передбачено такі елементи:

- заголовок сторінки;
- дві кнопки для завантаження фото та відео;

- універсальний контейнер `#mediaContainer`, який може містити або зображення (`<canvas id="imageCanvas">`), або відео (`<video id="videoPlayer">`);
- додатковий `canvas #overlayCanvas` для нанесення результатів поверх відео;
- кнопка "Розпізнати", яка ініціює обробку.

CSS-оформлення інтерфейсу

Стилізація забезпечує приємний та чистий вигляд сторінки:

- фон — світло-сірий, з центральним розміщенням медіаобласті;
- кнопки оформлені у вигляді елементів із заокругленими краями;
- `#mediaContainer` — основна зона для перегляду/аналізу медіа;
- `#overlayCanvas` розташовано абсолютно поверх відео для відображення емоційних рамок.

JavaScript-логіка роботи

Файл `script.js` відповідає за реакцію на дії користувача. Головні етапи:

1. Обробка завантаження зображення:

- При виборі зображення воно зчитується через `FileReader` (Рис. 3.1);
- Завантажене зображення відображається у `canvas`;
- Зображення надсилається на сервер через `fetch` POST-запит на `/detect_image`.

```
// Завантаження фото
imageInput.addEventListener('change', () => {
  const file = imageInput.files[0];
  if (!file) return;

  imageCanvas.style.display = 'block';
  videoPlayer.style.display = 'none';
  overlayCanvas.style.display = 'none';
  currentMode = 'image';

  const reader = new FileReader();
  reader.onload = function (e) {
    const img = new Image();
    img.onload = function () {
      imageCanvas.width = img.width;
      imageCanvas.height = img.height;
      imageContext.clearRect(0, 0, imageCanvas.width, imageCanvas.height);
      imageContext.drawImage(img, 0, 0);
    };
    img.src = e.target.result;
  };
  reader.readAsDataURL(file);
});
```

Рис. 3.1 – Фрагмент коду завантаження зображення.

2. Обробка завантаження відео:

- Відео завантажується у відеоплеєр;
- Поверх нього активується overlayCanvas;
- Скрипт кожні 500 мс знімає кадр і відправляє його на /detect_frame.

3. Візуалізація результатів:

Після отримання списку облич (координати + емоція), на відповідному canvas малюються червоні прямокутники та підписи.

Особливості canvas-візуалізації

- Для зображення використовується один canvas, на якому одночасно малюється і фото, і прямокутники.
- Для відео використовується два шари:
 - videoPlayer — основний плеєр;
 - overlayCanvas — прозорий шар для накладання емоцій;
- Масштаб координат розраховується динамічно, щоб емоції правильно накладались навіть при зміні розміру плеєра.

3.5. Опис реалізації бекенду

Серверна частина вебдодатку реалізована за допомогою фреймворку Flask. Вона відповідає за обробку HTTP-запитів, завантаження файлів, виявлення облич на зображеннях/кадрах та визначення емоцій за допомогою попередньо натренованої моделі.

Запуск Flask-серверу

Запуск виконується за допомогою стандартної конструкції (Рис. 3.2):

```
if __name__ == '__main__':
    app.run(debug=True)
```

Рис. 3.2 – Фрагмент коду запуску серверу.

Це дозволяє локально запустити сервер на `http://127.0.0.1:5000/`.

Завантаження моделі

При запуску застосунку відбувається завантаження натренованої моделі розпізнавання емоцій:

- `model = tf.keras.models.load_model('emotion_model.h5')`

Виявлення облич

Для пошуку облич використовується попередньо натренована SSD-модель на базі OpenCV. Цей підхід забезпечує швидке та достатньо точне виявлення облич на будь-якому фоні.

Передобробка зображення(Рис 3.3)

Для кожного виявленого обличчя:

- зображення обрізається до області обличчя;
- конвертується у відтінки сірого;
- застосовується CLAHE (контрастна гістограмна еквалізація);
- змінюється розмір до 48x48;
- нормалізується значення пікселів (0–1);
- змінюється форма тензора під модель.

```
gray = cv2.cvtColor(face_roi, cv2.COLOR_BGR2GRAY)
clahe = cv2.createCLAHE(clipLimit=2.0, tileGridSize=(8, 8))
gray = clahe.apply(gray)
resized = cv2.resize(gray, (48, 48))
normalized = resized.astype('float32') / 255.0
reshaped = np.expand_dims(np.expand_dims(normalized, -1), 0)

prediction = model.predict(reshaped)
emotion_idx = int(np.argmax(prediction))
emotion = emotion_labels[emotion_idx]
```

Рис. 3.3 – Фрагмент коду передобробки зображень.

Прогноз емоції

Оброблене зображення передається у модель (Рис 3.4):

```
prediction = model.predict(reshaped)
emotion_idx = int(np.argmax(prediction))
emotion = emotion_labels[emotion_idx]
```

Рис. 3.4 Фрагмент коду передачі зображення для подальшої обробки.

Список підтримуваних емоцій:

['Angry', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', 'Neutral']

API-маршрути

/ — головна сторінка:

Повертає HTML-шаблон:

```
@app.route('/')
def index():
    return render_template('index.html')
```

Рис. 3.5 Фрагмент коду який повертає HTML-шаблон.

/detect_image — розпізнавання на зображенні:

Приймає файл image, викликає функцію detect_faces_and_emotions, повертає результат у JSON.

/detect_frame — розпізнавання на кадрі відео:

Обробка майже ідентична до detect_image, але очікує файл frame.

Формат відповіді

Сервер повертає список словників такого вигляду:

```
[{"x": 120, "y": 85, "w": 75, "h": 75, "emotion": "Happy"}]
```

Ці координати використовуються фронтом для відображення результатів.

3.7. Демонстрація роботи проєкту

У цьому підрозділі наведено практичну демонстрацію функціональності вебдодатку, розробленого в межах дипломної роботи. Додаток дозволяє завантажувати зображення або відео, виявляти на них обличчя та визначати емоційний стан з використанням моделі глибокого навчання. Результати візуалізуються у реальному часі без необхідності сторонніх сервісів.

Запуск локального серверу

Після запуску app.py відкривається локальний сервер за адресою <http://127.0.0.1:5000/>. Інтерфейс виглядає наступним чином:

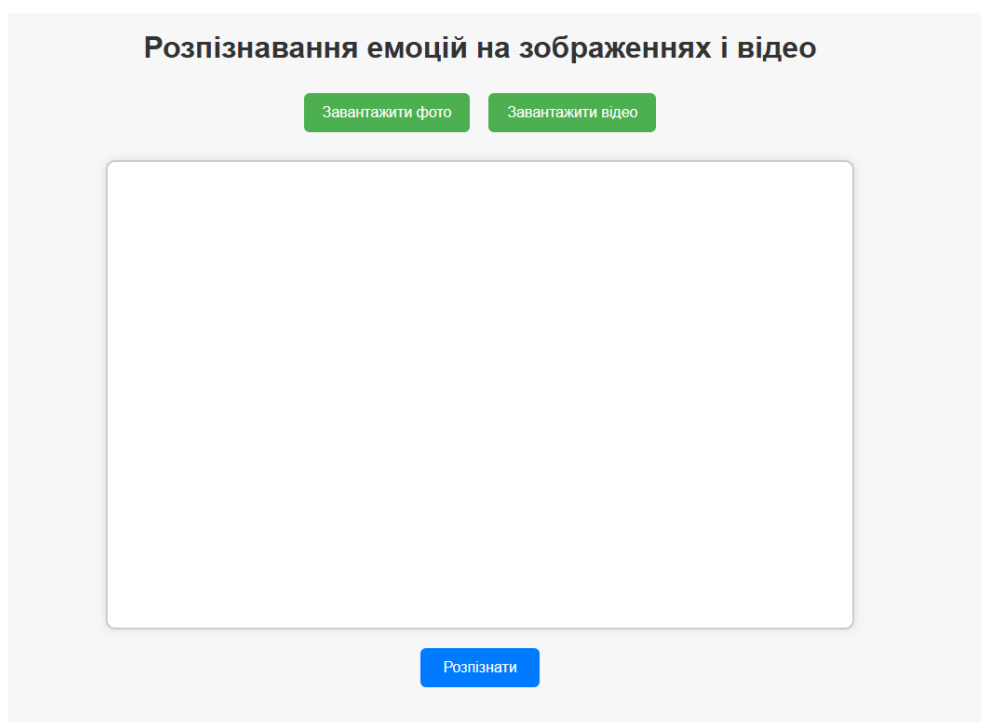


Рис. 3.6 Інтерфейс вебдодатку після запуску.

Аналіз зображення

- Користувач натискає кнопку «Завантажити фото» та обирає зображення з пристрою.
- Зображення відображається у canvas на сторінці.
- Після натискання кнопки «Розпізнати», зображення надсилається на сервер, де для кожного з них визначається емоція.
- У результаті на зображенні з'являються червоні прямокутники та підписи емоцій.

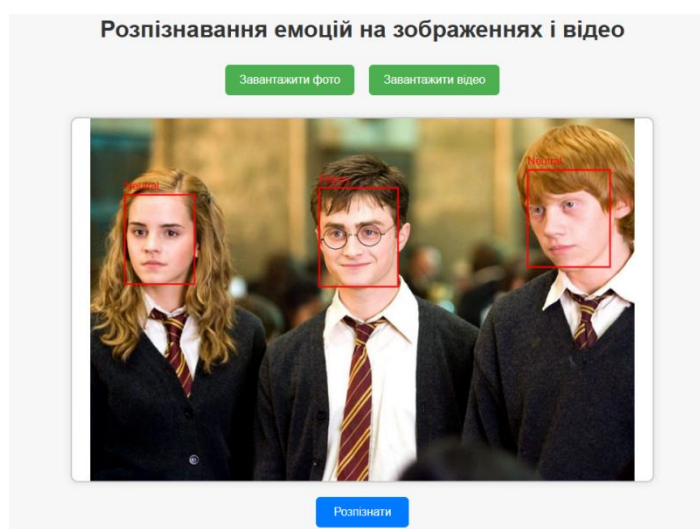


Рис. 3.7 Приклад роботи системи з фото

Аналіз відео

- Користувач завантажує відеофайл за допомогою кнопки «Завантажити відео».
- Відео відтворюється у вбудованому плеєрі.
- При натисканні «Розпізнати», кожні 500 мс з відео береться один кадр.
- Кадри надсилаються на сервер для обробки.
- Результати (обличчя + емоції) накладаються на overlayCanvas поверх відео.

Інтерактивність і зручність

- Весь процес відбувається без перезавантаження сторінки, завдяки JavaScript та технології fetch.
- Інтерфейс мінімалістичний, не перевантажений — користувач бачить лише основні дії.
- Виведення результатів відбувається миттєво, що робить додаток зручним для демонстрації роботи моделі.

Аналіз результатів

Під час тестування додатку на зображеннях і відео було помічено:

- високу точність у випадках з чіткими фронтальними обличчями;
- стабільну роботу при послідовній обробці кадрів з відео.

3.9. Можливості для розширення

Незважаючи на завершеність функціональної частини, розроблений вебдодаток має потенціал для подальшого вдосконалення та розширення. Нижче розглянуто напрямки, які можуть бути реалізовані в межах майбутніх проєктів або у виробничому застосуванні.

Підключення вебкамери для потокового розпізнавання

Одним із найпопулярніших сценаріїв є реальне використання вебкамери замість відеофайлів. Це дозволило б реалізувати повноцінну аналіз емоцій у реальному часі або використовувати додаток для живих консультацій, занять або моніторингу настрою.

Технічно це можливо шляхом використання API `getUserMedia` на фронтенді та захоплення кадрів через WebRTC або HTML5 відеопотік.

Покращення точності моделі

Поточна модель працює добре, однак має потенціал для донавчання на більш якісних/збалансованих наборах, використання глибших моделей, наприклад ResNet, EfficientNet, включення attention-механізмів або ConvLSTM для послідовного аналізу відео.

Підтримка додаткових емоцій або мультикласифікація

Датасет FER2013 обмежується 7 емоціями, але можливо додати класи, як-от *confused*, *tired*, *bored*, реалізувати мультиемоційну оцінку (наприклад: 60% *happy*, 40% *surprise*) та поєднати це з нейропсихологічними шкалами.

Створення системи журналювання або бази даних

Інтеграція з базою даних (наприклад SQLite, MongoDB або PostgreSQL) дозволила б:

- зберігати історію розпізнань;
- формувати аналітичні звіти;
- створити профіль користувача або аналітику емоцій за період.

Розгортання на хмарній платформі

Для демонстрації системи онлайн або публічного доступу можна:

- розгорнути додаток на Heroku, Render, Vercel (фронт), AWS чи Google Cloud;
- обгорнути додаток у Docker-контейнер;
- забезпечити захист запитів, CORS, сертифікати HTTPS.

Інтеграція з іншими сервісами

Систему можна розширити до складнішої платформи з:

- розпізнаванням мовлення + емоцій (аудіо+відео);
- використанням у чатботах, освітніх платформах, e-health;

ВИСНОВКИ

У дипломній роботі було повністю реалізовано систему для автоматичного розпізнавання емоцій людини на зображеннях і відео з використанням методів глибокого навчання. Система поєднує в собі технології комп'ютерного зору, штучного інтелекту та веброзробки, демонструючи приклад прикладного використання сучасних досягнень у сфері машинного навчання.

У ході виконання роботи було вирішено наступні ключові завдання:

- Проведено аналіз сучасних підходів до розпізнавання емоцій, а також розглянуто основні алгоритми, які застосовуються для класифікації виразів обличчя.
- Обрано архітектуру рішення, яке базується на клієнт-серверній моделі: фронтенд реалізовано на HTML/CSS/JavaScript, а бекенд — на Python з використанням фреймворку Flask.
- Розроблено модель глибокого навчання на основі згорткової нейронної мережі (CNN), яка була навчена на датасеті FER2013. Модель здатна розпізнавати 7 базових емоцій: Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral.
- Інтегровано модель у вебсервіс, який дозволяє користувачу завантажити зображення або відео для аналізу емоцій.
- Реалізовано функціонал обробки відеопотоку, де система знімає кадри кожні 500 мс, виконує розпізнавання емоцій у режимі, наближеному до реального часу, та візуалізує результати без перезавантаження сторінки.
- Забезпечено автоматичне виявлення облич за допомогою DNN-моделі на базі SSD (Single Shot Detector), що забезпечує стабільну роботу навіть при наявності кількох облич у кадрі.
- Подолано низку технічних проблем, серед яких: помилки з декодуванням зображень, невідповідність координат при відображенні результатів, стабільність canvas-рендерінгу у відео тощо.
- Розроблено інтуїтивно зрозумілий та візуально охайний інтерфейс, який не потребує додаткових інструкцій для користувача.

Особливу увагу було приділено інтерактивності інтерфейсу та стабільності обробки. Всі процеси — від завантаження файлу до виведення результатів — виконуються автоматично та асинхронно. Це робить систему зручною для реального використання у сфері освіти, психології, телемедицини або взаємодії з клієнтами.

Таким чином, поставлену мету — створення системи розпізнавання емоцій із вебінтерфейсом на основі глибокого навчання — успішно досягнуто. Система є готовою до використання та може бути адаптована під потреби конкретного користувача або організації.

ПЕРЕЛІК ПОСИЛАНЬ

1. How Does Facial Emotion Recognition Express Your Feelings?. *ARATEK | The World Leader in Biometrics Technology*. URL: <https://www.aratek.co/news/how-does-facial-emotion-recognition-express-your-feelings#:~:text=Facial%20Emotion%20Recognition%20,insights%20for%20many%20different%20uses>
2. Fernandes L. New Trends in Emotion Recognition Using Image Analysis by Neural Networks, A Systematic Review - PMC. *PMC Home*. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10458371/#:~:text=Facial%20emotion%20recognition%20,In%20this%20context>
3. Welcome to Flask – Flask Documentation (3.1.x). *Flask Documentation*. URL: <https://flask.palletsprojects.com/en/stable/>
4. Facial Emotion Recognition Using Transfer Learning in the Deep CNN / M. A. H. Akhand et al. *Electronics*. 2021. Vol. 10, no. 9. P. 1036. URL: <https://doi.org/10.3390/electronics10091036>
5. Kusuma G. P., Jonathan J., Lim A. P. Emotion Recognition on FER-2013 Face Images Using Fine-Tuned VGG-16. *Advances in Science, Technology and Engineering Systems Journal*. 2020. Vol. 5, no.6. P. 315–322. URL: <https://doi.org/10.25046/aj050638>
6. A Combined CNN Architecture for Speech Emotion Recognition / R. Begazo et al. *Sensors*. 2024. Vol. 24, no. 17. P. 5797. URL: <https://doi.org/10.3390/s24175797>
7. Rafalski K. CNN Optimization: Strategies for Enhancing Deep Learning Performance. *Software Development Company | Netguru*. URL: <https://www.netguru.com/blog/cnn-optimization#:~:text=Data%20augmentation%20also%20serves%20as,like%20rotations,%20flips,%20and%20crops>
8. A review of convolutional neural networks in computer vision / X. Zhao et al. *Artificial Intelligence Review*. 2024. Vol. 57, no. 4. URL: <https://doi.org/10.1007/s10462-024-10721-6>
9. Cozzi P. Human emotion recognition: the latest advances in emotion AI. *Tech4Future*. URL: <https://tech4future.info/en/human-emotion-recognition-ai/>
10. Bengio Y., Courville A., Goodfellow I. *Deep Learning*. MIT Press, 2016. 775 p.
11. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv.org*. URL: <https://arxiv.org/abs/1704.04861>

12. Wang Y., Chen L., Gao H. A Survey on Facial Expression Recognition of Static and Dynamic Images. *arXiv*, 2024. URL: <https://arxiv.org/abs/2408.15777v1>
13. Zhao Y.-F., Li B., Sun J. Advances in Facial Expression Recognition: A Survey of Methods. *Information*, 2024. Vol. 15, No. 3. Article 135. URL: <https://www.mdpi.com/2078-2489/15/3/135>
14. TensorFlow. Keras guide – Get started with Keras. TensorFlow Official Guide, 2024. URL: <https://www.tensorflow.org/guide/keras>
15. Maurício J., Domingues I., Bernardino J. Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. *Applied Sciences*, 2023. URL: <https://doi.org/10.3390/app13095521>
16. Optimization of Convolutional Neural Network Hyperparameter for Feature Extraction using Metaheuristic Methods. *arXiv*, 2024. URL: <https://doi.org/10.48550/arXiv.2412.17956>
17. Wang Y., Song W., Tao W. et al. A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances. *arXiv*, 2022. URL: <https://doi.org/10.48550/arXiv.2203.06935>
18. Lémin X., Muszyński M. (eds.). *Applied Affective Computing*. ACM Books, 2022, pp. 1–50, 110–150.
19. JavaScript – MDN Web Docs. Mozilla Developer Network, 2024. URL: <https://developer.mozilla.org/en-US/docs/Web/JavaScript>
20. Guo R., Guo H., Wang L. et al. Development and application of emotion recognition technology – a systematic literature review. *BMC Psychology*, 2024. URL: <https://doi.org/10.1186/s40359-024-01581-4>

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

КАФЕДРА ШТУЧНОГО ІНТЕЛЕКТУ

КВАЛІФІКАЦІЙНА РОБОТА НА ТЕМУ : «СИСТЕМА РОЗПІЗНАВАННЯ ЕМОЦІЙ НА ФОТО ТА ВІДЕО ЗА ДОПОМОГОЮ МЕТОДІВ ГЛИБОКОГО НАВЧАННЯ»

Виконав: здобувач вищої освіти гр. ШІД-42
Шклярєнко Владислав Володимирович

Керівник: кандидат технічних наук
Шантір А. С.

1

Мета: розробка ефективної системи розпізнавання емоцій на фото та відео на основі методів глибокого навчання для практичного застосування у вебдодатку.

Об'єкт: процеси аналізу та розпізнавання емоцій за допомогою алгоритмів глибокого навчання у зображеннях та відеоданих.

Предмет: процеси інтеграції та використання нейронної мережі для розпізнавання емоцій у веб-застосунку, включаючи обробку зображень і відео, передобробку даних, візуалізацію результатів та організацію взаємодії між фронтендом і бекендом.

Завдання:

- Аналіз існуючих підходів та методів розпізнавання емоцій на зображеннях і відео.
- Розробка та тренування нейронної мережі для класифікації емоцій за допомогою TensorFlow/Keras.
- Розробка веб-додатку, в який буде інтегровано модель.
- Тестування на реальних даних та оцінка ефективності.
- Рекомендації щодо покращення системи та можливих напрямків подальших досліджень

2

ВИКОРИСТАНІ ТЕХНОЛОГІЇ

FrontEnd: HTML, CSS, JavaScript.

BackEnd: Python.

Бібліотеки:

- TensorFlow/Keras - для створення та навчання моделі.
- Flask - для розробки бекенду веб-додатка.



K Keras

Датасет: FER2013



3

Розробка моделі

Модель розроблена на основі архітектури CNN(Згортова нейронна мережа) та натренована на датасеті FER-2013. Модель має послідовну архітектуру, де дані проходять через серію згорткових блоків, які поетапно виділяють візуальні патерни.

Плюси:

- Захист від перенавчання;
- Динамічна аугментація даних;
- Оптимізованість під обчислювальні ресурси;
- Ефективне виявлення локальних патернів(краї, контури, текстури).

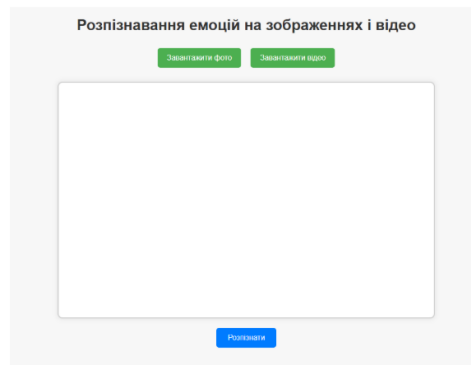
Мінуси:

- Плутання схожих емоцій;
- Недостатня точність для критичних застосунків;
- Дисбаланс класів датасету.

4

Демонстрація проєкту

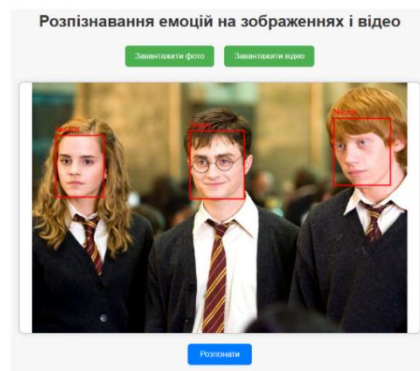
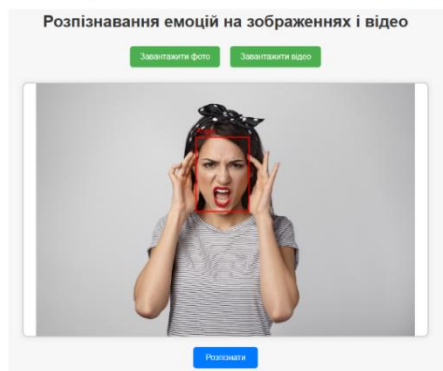
Інтерфейс веб-додатку



5

Демонстрація проєкту

Розпізнавання емоцій на фото



6

Демонстрація проєкту

Розпізнавання емоцій на відео



7

ЕФЕКТИВНІСТЬ

У процесі навчання модель досягла близько 66% точності. Для збільшення ефективності можна використати готові, більш потужні моделі або удосконалити свою модель та натренувати її на реалістичнішому датасеті(наприклад AffectNet), але це потребує більших ресурсів(потужний ПК, більше часу).

Актуальність та сфери застосування

У сучасному світі важливим стає розуміння емоцій користувача в системах взаємодії людини і комп'ютера. Система розпізнавання емоцій підвищить ефективність та адаптивність цифрових сервісів.

Сфери застосування:

- Безпека - виявлення агресивних чи тривожних емоцій у публічних місцях.
- Ігрова індустрія - персоналізація досвіду гравця.
- Маркетинг - аналіз реакції клієнтів на контент.

9

ВИСНОВКИ

Під час виконання роботи було розроблено систему розпізнавання емоцій на фото на відео за допомогою методів глибокого навчання. Система показала задовільну точність на тестових даних(~66%) та зручність інтеграції у веб середовище. Проведено аналіз сфер застосування та актуальності розпізнавання емоцій у IT-системах. Визначено можливі шляхи подальшого вдосконалення, зокрема: використання більш наповнених датасетів та потужніших моделей.