МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ВСП «ЕКОНОМІКО-ПРАВНИЧИЙ ФАХОВИЙ КОЛЕДЖ ЗАПОРІЗЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ»

Циклова комісія математичних дисциплін та інформаційних технологій

КВАЛІФІКАЦІЙНА РОБОТА

на тему: «РОЗРОБКА НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ІДЕНТИФІКАЦІЇ ЛЮДСЬКИХ ЕМОЦІЙ»

Виконав: здобувач освіти 4 курсу, групи К121-22
Спеціальність 121 Інженерія
програмного забезпечення
(шифр і спеціальність)
Микита ЧОПОРОВ
(ім'я та ПРІЗВИЩЕ)
Керівник Дмитро ДРАЄВСЬКИЙ
(ім'я та ПРІЗВИЩЕ)
керівник Інституту проблем штучного інтелекту, член-кореспондент НАН України, доктор технічних наук, доктор богослов'я,
Рецензент професор, Анатолій ШЕВЧЕНКО
(посада, вчене звання, науковий ступінь, ім'я та ПРІЗВИЩЕ)

Запоріжжя
2026

ЗАПОРІЗЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ВСП «Економіко-правничий фаховий коледж ЗНУ»

Освітньо–професійний ступінь фаховий молодший бакалавр

Спеціальність 121 – Інженерія програмного забезпечення
(шифр і назва)

ЗАТВЕРДЖУЮ

Голова циклової комісії математичних дисциплін та інформаційних технологій

_____ Т.М. Смолянкова
(підпис)

“ 05 ” червня 2026 р.

ЗАВДАННЯ

на кваліфікаційну роботу студенту

ЧОПОРОВУ Микиті Сергійовичу
(ПРІЗВИЩЕ, ім’я та по– батькові)

1. Тема роботи «Розробка нейронної мережі для ідентифікації людських емоцій»

Керівник роботи ДРАЄВСЬКИЙ Дмитро Сергійович
(науковий ступінь, вчене звання, прізвище, ім’я, по батькові)

Затверджені наказом ВСП ЕПФК ЗНУ від «01» грудня 2025 р. № 1728-с

2. Строк подання студентом роботи 05.06.2026

3. Вихідні дані до роботи 1. Постановка задачі.
2. Перелік джерел.

4. Зміст розрахунково– пояснювальної записки (перелік питань, які потрібно розробити)
1.Сучасні нейронні мережі.
2. Розробка проекту нейронних мереж.
3. Програмна реалізація нейронної мережі.

5. Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)
презентація до захисту

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв

7. Дата видачі завдання 01.12.2025

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Розробка плану роботи:		
	постановка задачі	Грудень 2025	виконано
2.	Збір вихідних даних, обробка методичних		
	та теоретичних джерел	Січень 2026	виконано
3.	Розробка першого розділу:		
	огляд інформаційних систем та технологій для їхнього створення	Лютий 2026	виконано
4.	Розробка другого розділу:		
	розробка проєкту інформаційної системи та вибір технологій	Квітень 2026	виконано
5.	Розробка третього розділу:		
	розробка застосунку	Травень 2026	виконано
6.	Оформлення і нормоконтроль кваліфікаційної роботи та перевірка на плагіат	Червень 2026	виконано
7.	Захист кваліфікаційної роботи	16.06.2026	виконано

Здобувач освіти _____ Микита ЧОПОРОВ
(підпис) (ім’я ПРІЗВИЩЕ)

Керівник роботи _____ Дмитро ДРАЄВСЬКИЙ
(підпис) (ім’я ПРІЗВИЩЕ)

Нормоконтроль пройдено

Нормоконтролер _____ Юлія БОРИСОВСЬКА
(підпис) (ім’я ПРІЗВИЩЕ)

РЕФЕРАТ

Кваліфікаційна робота: 47 сторінок, 17 рисунків, 26 джерел.

Об’єкт дослідження – процес автоматичного розпізнавання людських емоцій на основі аналізу зображень обличчя засобами комп’ютерного зору.

Предмет дослідження: методи глибокого навчання, згорткові нейронні мережі та програмні засоби для побудови систем ідентифікації емоцій.

Мета роботи: розробка ефективної нейронної мережі для ідентифікації людських емоцій за зображеннями обличчя з використанням технологій глибокого навчання.

Метод дослідження: аналіз сучасних методів розпізнавання емоцій, проєктування архітектури згорткової нейронної мережі, програмна реалізація системи, навчання моделі та експериментальна оцінка її ефективності.

У кваліфікаційній роботі проведено детальний аналіз згорткових нейронних мереж та розроблено нейронну мережу для ідентифікації людських емоцій по фотографіям

Робота демонструє не лише знання теоретичних аспектів штучного інтелекту та аналізу даних, але й вміння їх успішно реалізувати в практичному застосуванні.

У результаті проведеного дослідження було досягнуто поставленої мети, а саме – розроблено та реалізовано нейронну мережу, яка забезпечує точну ідентифікацію та класифікацію людських емоцій. Ця розробка може бути використана для підвищення ефективності взаємодії людини з комп'ютером, у психологічних дослідженнях, системах безпеки або для покращення рівня обслуговування у відповідних сферах.

НЕЙРОННА МЕРЕЖА, БІБЛІОТЕКА, НАБІР ДАНИХ, НАВЧАННЯ, ТЕСТУВАННЯ, ПРОГНОЗУВАННЯ

SUMMARY

Qualification work: 47 pages, 17 figures, 26 sources.

The object of the study is the process of automatic recognition of human emotions based on the analysis of facial images using computer vision.

Subject of the study: deep learning methods, convolutional neural networks and software tools for building emotion identification systems.

Purpose of the work: development of an effective neural network for identifying human emotions from facial images using deep learning technologies.

Research method: analysis of modern methods of emotion recognition, design of the architecture of the convolutional neural network, software implementation of the system, model training and experimental evaluation of its effectiveness.

In the qualification work, a detailed analysis of convolutional neural networks was conducted and a neural network for identifying human emotions from photographs was developed.

The work demonstrates not only knowledge of the theoretical aspects of artificial intelligence and data analysis, but also the ability to successfully implement them in practical application.

As a result of the study, the set goal was achieved, namely, a neural network was developed and implemented that provides accurate identification and classification of human emotions. This development can be used to improve the efficiency of human-computer interaction, in psychological research, security systems, or to improve the level of service in related areas.

NEURAL NETWORK, LIBRARY, DATA SET, TRAINING, TESTING, PREDICTION

ЗМІСТ

Завдання на кваліфікаційну роботу студенту 2
Реферат 4
Summary 5
Вступ 7
1 Аналіз методів розпізнавання емоцій людини та використання нейронних мереж 9
1.1 Актуальність та сфери застосування технологій розпізнавання облич та емоційних станів 9
1.2 Поняття емоцій та їх роль у житті людини 10
1.3 Методи розпізнавання емоцій людини 11
1.4 Огляд існуючих розробок та методів розпізнавання облич та емоцій 12
1.5 Теоретичні основи штучного інтелекту та аналіз архітектур нейронних мереж для обробки візуальних даних 15
1.6 Висновки до розділу 1 16
2 Проєктування нейронної мережі розпізнавання облич та емоцій 17
2.1 Вибір середовища розробки та огляд прикладних бібліотек 17
2.2 Загальні поняття згорткових мереж 19
2.3 Проєктування архітектури згорткової нейронної мережі для розпізнавання емоцій 22
2.4 Формування та структура набору даних (Dataset) 24
2.5 Попередня обробка, очищення даних та візуальний аналіз розподілу класів 26
2.6 Метрики оцінки якості та аналіз результатів розпізнавання 27
2.7 Висновки до розділу 2 28
3 Розробка нейронної мережі для ідентифікації людських емоцій 30
3.1 Завантаження та попередня обробка набору даних 30
3.2 Етап підготовки та попередньої обробки даних 33
3.3 Архітектура розробленої нейронної мережі 33
3.4 Процес навчання 36
3.5 Аналіз результатів 37
3.6 Результати роботи розробленої нейронної мережі 38
3.7 Аналіз матриці помилок (Confusion Matrix) 39
3.8 Приклади роботи нейронної мережі 41
Висновки 44
Перелік використаних джерел 45

ВСТУП

Розпізнавання облич та емоцій наразі знаходиться на піку актуальності, оскільки ми переходимо від простого збору даних до інтелектуальної взаємодії між машиною та людиною. Сьогодні технології штучного інтелекту та комп'ютерного зору активно впроваджуються в усі сфери життя. Здатність машин розуміти стан людини є критичним кроком до створення ефективної взаємодії "людина-комп'ютер". Розпізнавання людських емоцій має величезний потенціал у таких галузях, як психологічна діагностика, системи безпеки, адаптивне навчання, маркетинг та розробка відеоігор. Незважаючи на значний прогрес, створення стійких до перешкод (наприклад, поганого освітлення чи перекритого обличчя) моделей залишається складною науковою та інженерною проблемою. Саме тому розробка та оптимізація нейронної мережі для точного розпізнавання емоцій є актуальним та своєчасним завданням [1].

Мета дослідження – розробити, навчити та протестувати програмну систему на базі нейронної мережі, здатну автоматично класифікувати базові людські емоції (радість, сум, гнів, здивування, страх, відраза, нейтральний стан) за зображеннями обличчя.

Для досягнення поставленої мети необхідно виконати такі завдання:

- провести аналіз існуючих методів, алгоритмів та архітектур нейронних мереж для розпізнавання образів та емоцій;
- обрати оптимальну архітектуру та інструментальні засоби для реалізації;
- підготувати та обробити набір даних (датасет) для навчання та валідації моделі;
- спроєктувати та програмно реалізувати нейронну мережу;
- провести експериментальне тестування моделі, оцінити її точність та проаналізувати отримані результати.

Об'єкт дослідження – процес автоматизованого розпізнавання та класифікації людських емоцій на основі візуальних даних.

Предмет дослідження – методи, алгоритми та архітектури нейронних мереж, що застосовуються для аналізу міміки та ідентифікації емоційних станів людини.

Методи дослідження – методи машинного навчання, теорії штучних нейронних мереж, цифрової обробки зображень (комп'ютерного зору) та методи статистичного аналізу для оцінки ефективності розробленої моделі.

Практичне значення отриманих результатів – розроблена нейромережева модель та програмний код можуть бути використані як базовий модуль для інтеграції в складніші системи (наприклад, для моніторингу стану водіїв, систем розумного дому або інструментів аналізу реакції споживачів).

1 АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ ТА ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ

1.1 Актуальність та сфери застосування технологій розпізнавання облич та емоційних станів

На сьогоднішній день технології комп'ютерного зору, зокрема розпізнавання облич та аналіз емоцій, є одним із найбільш динамічних напрямів розвитку штучного інтелекту. Актуальність цієї теми зумовлена глобальним переходом від пасивного збору візуальних даних до побудови систем інтелектуальної взаємодії між машиною та людиною. Впровадження таких рішень дозволяє автоматизувати складні процеси ідентифікації, підвищити рівень безпеки та забезпечити персоналізований підхід у сервісному секторі.

Практична цінність розробки нейронних мереж у цьому напрямі охоплює широкий спектр галузей: безпека та біометричний контроль, комерція, моніторинг безпеки на транспорті, охорона здоров'я та соціальна сфера, освіта та управління персоналом.

Безпека та біометричний контроль – це базовий сегмент застосування, що включає системи контролю доступу на підприємствах, автоматизовану верифікацію в аеропортах та інтеграцію в концепцію «Smart City» для ідентифікації осіб у реальному часі. Використання обличчя як біометричного ключа мінімізує ризики, пов'язані з крадіжкою фізичних носіїв доступу [2].

Технології розпізнавання стають фундаментом для безконтактних платежів та віддаленої верифікації клієнтів у банківському секторі. У ритейлі аналіз емоцій дозволяє оцінювати індекс задоволеності клієнтів безпосередньо в точках продажу, що дає змогу бізнесу оперативно коригувати маркетингові стратегії.

Моніторинг безпеки на транспорті. Одним із критично важливих напрямів є розробка систем моніторингу стану водія. Нейромережі здатні розпізнавати ознаки втоми, засинання або відволікання від дороги, що безпосередньо впливає на зниження рівня аварійності.

Охорона здоров'я та соціальна сфера. Впровадження аналізу емоцій у HealthTech відкриває нові можливості для телемедицини, зокрема для діагностики психоемоційних розладів, депресивних станів або оцінки інтенсивності болю у пацієнтів з обмеженою комунікацією.

Освіта та управління персоналом. В освітніх процесах технології використовуються для автоматичного контролю присутності та систем прокторингу (контролю чесності) під час іспитів. У сфері HR аналіз міміки допомагає об'єктивно оцінювати софтскіли та рівень стресостійкості кандидатів під час відеоінтерв'ю.

Попри широке розповсюдження, сфера залишається відкритою для наукового пошуку. Основними викликами, що потребують вирішення, є підвищення точності розпізнавання в умовах недостатнього освітлення, робота з оклюзіями (маски, окуляри), а також оптимізація моделей для роботи на мобільних пристроях з обмеженими обчислювальними ресурсами. Таким чином, розробка ефективної нейронної мережі для розпізнавання облич та емоцій є не лише технічно складною, але й соціально значущою задачею з високим потенціалом практичного впровадження [3].

1.2 Поняття емоцій та їх роль у житті людини

Емоції є важливою складовою психічної діяльності людини та відіграють ключову роль у процесах сприйняття, мислення, прийняття рішень і міжособистісної комунікації. Вони відображають суб’єктивне ставлення людини до подій навколишнього середовища та внутрішніх станів організму.

Емоційні реакції значною мірою визначають поведінку людини, її соціальну взаємодію та здатність адаптуватися до змін.

У психології емоції розглядаються як складні психофізіологічні процеси, що включають когнітивні, фізіологічні та поведінкові компоненти. Когнітивний компонент пов'язаний з оцінкою ситуації, фізіологічний – зі змінами в організмі, а поведінковий – з зовнішнім проявом емоцій, зокрема мімікою обличчя, жестами та інтонацією голосу.

Однією з найбільш відомих теорій емоцій є теорія базових емоцій, запропонована американським психологом Полом Екманом. Згідно з цією теорією, існує обмежений набір універсальних емоцій, які є спільними для людей незалежно від культури. До них зазвичай відносять радість, сум, гнів, страх, здивування, відразу та нейтральний стан. Саме ці емоції найчастіше використовуються в системах автоматичного розпізнавання [4].

1.3 Методи розпізнавання емоцій людини

Розпізнавання емоцій людини є міждисциплінарною задачею, що поєднує психологію, комп'ютерні науки та штучний інтелект. Залежно від джерела інформації розрізняють декілька основних підходів до визначення емоційного стану, серед яких провідне місце займає аналіз міміки обличчя. Цей метод є одним із найпоширеніших і найбільш інформативних, оскільки мімічні рухи безпосередньо відображають внутрішній стан людини і можуть бути легко зафіксовані за допомогою цифрових зображень або відео. Традиційні методи аналізу обличчя ґрунтувалися на виділенні геометричних ознак, таких як положення очей, брів, рота та їх взаємне розташування. Проте через чутливість до змін освітлення, повороту голови та індивідуальних особливостей, на зміну їм прийшли сучасні системи на базі згорткових нейронних мереж. Вони здатні автоматично виявляти релевантні ознаки без

ручного втручання, що значно підвищує точність та універсальність розпізнавання.

Окрім візуальних даних, вагомим джерелом інформації для аналізу є акустичні характеристики мовлення. Розпізнавання емоцій за голосом базується на дослідженні висоти тону, тембру, енергії сигналу та швидкості мовлення за допомогою спектральних ознак, зокрема мел-кепстральних коефіцієнтів. Даний підхід часто застосовується в телефонних сервісах, голосових асистентах та системах підтримки клієнтів, хоча його ефективність значною мірою залежить від якості аудіозапису та складності обробки фонових шумів.

Поряд із невербальними проявами, якими є міміка та голос, важливу роль відіграє і змістова складова комунікації, що аналізується через текст. Цей напрям передбачає визначення емоційної тональності висловлювань за допомогою методів обробки природної мови та алгоритмів машинного навчання. Такий підхід широко застосовується в аналізі соціальних мереж і чат-ботах. Хоча текстовий аналіз сам по собі не враховує міміку чи інтонацію, його інтеграція з візуальними та акустичними даними дозволяє створювати потужні мультимодальні системи [4].

1.4 Огляд існуючих розробок та методів розпізнавання облич та емоцій

Сучасні системи комп'ютерного зору пройшли шлях від простих геометричних моделей до складних архітектур глибокого навчання, що дозволяють досягати точності понад 90–96% у складних умовах [1, 2].

Процес ідентифікації зазвичай складається з етапів детекції (виявлення), вирівнювання та екстракції ознак [5].

Класичні методи:

- алгоритм Віоли-Джонса базується на ознаках Хаара. Попри високу швидкість, він чутливий до змін освітлення та поворотів голови;
- метод головних компонент – статистичний підхід, що використовується для зменшення розмірності даних при збереженні ключових рис обличчя.

Глибоке навчання:

- RetinaFace та DSFD – моделі, що забезпечують максимальну точність детекції навіть при перекриттях (окуляри, маски), хоча потребують значних обчислювальних ресурсів;
- MediaPipe та SSD – легковажні рішення, оптимізовані для роботи в реальному часі на мобільних пристроях;
- DeepFace (Facebook) та FaceNet (Google) – використовують згорткові мережі для перетворення зображення обличчя у вектор (embedding), що дозволяє порівнювати обличчя як точки у багатовимірному просторі [5, 6, 7, 8, 9].

Методи розпізнавання емоцій (Facial Emotion Recognition, FER). Завдання FER полягає у класифікації виразів обличчя за базовими категоріями (радість, сум, гнів, страх, здивування, огида, нейтральний стан) [10].

Архітектури нейронних мереж:

- CNN (Згорткові мережі) – стандарт індустрії для аналізу статичних зображень. Моделі на кшталт ResNet та VGGNet демонструють високу ефективність у вилученні мімічних ознак;
- ViT (Vision Transformers) – новітній підхід, що аналізує зображення як послідовність патчів, що дозволяє краще вловлювати глобальні залежності між різними частинами обличчя;
- гібридні моделі (CNN-LSTM) – використовуються для аналізу відеопотоку. CNN вилучає ознаки з кадрів, а LSTM (довга короткострокова пам'ять) аналізує їхню зміну в часі для розпізнавання динаміки емоцій.

Підходи до екстракції ознак:

- геометричні – аналіз відстаней між антропометричними точками обличчя (кутики губ, очі, брови);

- текстурні (Appearance-based) – аналіз змін текстури шкіри (зморшки, складки), що виникають при певних емоціях [11, 12, 13, 14, 15].

Сфери застосування та актуальні датасети Для навчання моделей найчастіше використовують набори даних FER2013, AffectNet та CK+. Практичне застосування охоплює [16]:

- Smart Security: Біометрична ідентифікація та пошук осіб;
- HealthTech – моніторинг ментального здоров'я та діагностика емоційних розладів;
- Metaverse та AR – перенесення реальних емоцій користувача на віртуальний аватар у реальному часі;
- автомобільна галузь – системи моніторингу втоми водія для запобігання ДТП [10, 17, 18, 19].

На сьогодні спостерігається тренд на перехід до мультимодальних систем (аналіз обличчя та голос) та оптимізацію важких моделей для роботи на "крайових" пристроях без втрати точності [18, 19].

Пол Екман – найвідоміший у світі психолог, який створив систему кодування рухів обличчя. Його дослідження базових емоцій стали основою для всіх сучасних алгоритмів розпізнавання емоцій.

Ян Лекун – лауреат премії Тюрінга, один із "батьків" згорткових нейронних мереж (CNN). Саме його розробки дозволили ефективно розпізнавати візуальні образи та обличчя.

Тааво Кін та команда Google Brain/DeepMind – розробники архітектур, що лежать в основі FaceNet (революційна модель для верифікації облич).

Джефрі Гінтон – класик глибокого навчання, чиї роботи по архітектурах нейромереж дозволили автоматизувати екстракцію ознак з людського обличчя.

Рана ель-Каліубі – розробниця та дослідниця у галузі "емоційного штучного інтелекту", співзасновниця компанії Affectiva, яка спеціалізується на розпізнаванні емоцій для маркетингу та медицини.

Україна також має потужну школу комп'ютерного зору, представлену як науковцями, так і практиками.

1.5 Теоретичні основи штучного інтелекту та аналіз архітектур нейронних мереж для обробки візуальних даних

В основі сучасних систем розпізнавання облич та емоцій лежить штучний інтелект (ШІ) – галузь комп’ютерних наук, що займається створенням систем, здатних виконувати завдання, які зазвичай потребують людського інтелекту. Найбільш ефективним підходом у межах ШІ для вирішення задач комп'ютерного зору є машинне навчання, а саме його підмножина – глибоке навчання.

Глибоке навчання базується на використанні багатошарових штучних нейронних мереж, які здатні самостійно вилучати ієрархічні ознаки з сирих даних (пікселів зображення). Для задач аналізу облич сьогодні домінують дві основні архітектурні парадигми:

1. Згорткові нейронні мережі (Convolutional Neural Networks, CNN)/

CNN є класичним та найбільш розповсюдженим інструментом для обробки зображень. Їхня архітектура імітує роботу біологічних нейронів зорової кори.

Принцип роботи: Через операцію згортки (convolution) мережа застосовує набір фільтрів, які спочатку фіксують прості ознаки (краї, лінії, кути), а на глибших шарах – складні структури (очі, ніс, контур обличчя).

Переваги: Стійкість до зсувів та поворотів об’єкта на зображенні, відносна обчислювальна ефективність та наявність великої кількості попередньо навчених моделей (наприклад, ResNet, MobileNet, VGG).

2. Трансформери візуальних образів (Vision Transformers, ViT).

Це новітній підхід, запозичений з обробки природної мови (NLP), який демонструє вражаючі результати в останні роки.

Принцип роботи: Зображення розбивається на окремі фрагменти (патчі), які подаються на вхід мережі як послідовність токенів. Механізм Self-Attention (самоуваги) дозволяє моделі одночасно аналізувати взаємозв'язки між усіма частинами обличчя, незалежно від відстані між ними.

Переваги: ViT краще за CNN вловлюють глобальний контекст зображення, що критично важливо для розпізнавання складних емоцій, де стан людини визначається не окремою рисою, а цілісним виразом обличчя.

Вибір між CNN та ViT залежить від технічних обмежень проєкту. CNN залишаються оптимальним вибором для систем реального часу та мобільних додатків завдяки швидкості роботи. У той же час ViT забезпечують вищу точність на великих обсягах даних, проте потребують значних обчислювальних потужностей (GPU).

1.6 Висновки до розділу 1

У першому розділі було розглянуто поняття емоцій та їх значення у житті людини, а також проаналізовано основні методи розпізнавання емоцій. Особливу увагу приділено сучасним підходам, заснованим на використанні нейронних мереж.

Проведений аналіз показав, що найбільш ефективним та перспективним методом є розпізнавання емоцій за зображенням обличчя з використанням згорткових нейронних мереж. Це обґрунтовує вибір даного підходу для подальшої практичної реалізації у межах дипломної роботи.

2 ПРОЄКТУВАННЯ НЕЙРОННОЇ МЕРЕЖІ РОЗПІЗНАВАННЯ ОБЛИЧ ТА ЕМОЦІЙ

Для практичної реалізації розробленої архітектури нейронної мережі було обрано мову програмування Python, яка є стандартом у галузі Data Science. Обробка відеопотоку та попереднє перетворення зображень (зміна розміру, конвертація кольорів) здійснювалися за допомогою бібліотеки OpenCV [5]. Побудова та навчання глибокої нейронної мережі для класифікації емоцій базувалися на фреймворку TensorFlow та його високорівневому API Keras [6, 16], що дозволяє ефективно використовувати потужності графічних процесорів (GPU) для прискорення обчислень.

2.1 Вибір середовища розробки та огляд прикладних бібліотек

Першим етапом розробки нейромережевої моделі є формування стійкої програмної екосистеми. Для реалізації поставлених завдань було обрано мову програмування Python, яка завдяки лаконічному синтаксису та потужній підтримці спільноти є де-факто стандартом у галузі штучного інтелекту.

Для побудови ефективного конвеєра (pipeline) обробки візуальних даних необхідне використання спеціалізованих бібліотек та фреймворків:

OpenCV (Open Source Computer Vision Library) – ця бібліотека є фундаментальною для завдань комп'ютерного зору. У межах даного проєкту OpenCV використовується для:

- захоплення та зчитування відеопотоку в реальному часі;
- попередньої обробки кадрів (зміна розміру, перетворення у відтінки сірого для зменшення обчислювального навантаження);
- візуалізації результатів (малювання обмежувальних рамок навколо облич та виведення тексту з назвою розпізнаної емоції) [21];

TensorFlow – це платформа з відкритим кодом для машинного навчання від Google, а Keras – високорівневий API, що працює поверх неї. Їх обрано для:

- опису архітектури нейронної мережі (додавання згорткових, пулінгових та повнозв'язних шарів);
- керування процесом навчання (налаштування оптимізаторів, функцій втрат та метрик точності);
- збереження та подальшого завантаження навченої моделі для інференсу (використання) [22].

NumPy – математична бібліотека, необхідна для роботи з багатовимірними масивами. Оскільки будь-яке цифрове зображення сприймається комп'ютером як матриця чисел (пікселів), NumPy забезпечує швидкі обчислення та маніпуляції з цими даними на низькому рівні [23, 24].

Pandas та Matplotlib – ці інструменти використовуються на етапі аналізу даних та підготовки датасету. Pandas дозволяє структурувати інформацію про зображення (мітки класів, шляхи до файлів). Matplotlib застосовується для побудови графіків навчання (динаміка точності та втрат по епохах), що необхідно для візуального аналізу якості моделі та виявлення ефекту перенавчання (overfitting) [25].

MediaPipe (опціонально) – для задачі детекції облич (виявлення їх на кадрі перед розпізнанням емоцій) доцільно використовувати рішення від Google – MediaPipe Face Detection. Воно базується на легковажних моделях, що забезпечують високу швидкість роботи навіть на пристроях з низькою потужністю [26].

Обраний стек технологій дозволяє створити гнучку систему, що поєднує високу швидкість обробки зображень (завдяки OpenCV) та точність глибокого навчання (завдяки TensorFlow). Використання мови Python забезпечує можливість швидкого прототипування та подальшої масштабованості системи.

2.2 Загальні поняття згорткових мереж

Згорткові нейронні мережі – це клас нейронних мереж із прямим зв’язком, які підходять для операцій з двовимірними даними, такими як зображення. CNN подібні до звичайних нейронних мереж, де вони складаються з нейронів, які мають вагу та зміщення. Нейрони в CNN отримують вхідні дані, виконують точковий добуток, за яким слідує нелінійність, а потім застосовують функцію втрат до шару класифікації. Основна відмінність між CNN і звичайними мережами з прямим зв’язком полягає в тому, що CNN краще справляються з двовимірними вхідними даними, і тому вони в основному використовуються в класифікації зображень. CNN зазвичай починають зі згорткового шару, де він приймає вхідні зображення та розкладає їх на різні карти об’єктів, такі як краї, лінії, криві тощо.

До витягнутих карт об’єктів по всій мережі застосовуються кілька процесів. Витягнуті карти об’єктів з останнього шару (як правило, повністю підключеного шару) класифікуються на вихідні класи за допомогою класифікатора, такого як класифікатор SoftMax [14]. Типова згортка нейронна мережа складається з наступних елементів:

- кількість згорткових і повнозв'язаних шарів, де виконується більшість операцій;
- об’єднання шарів, які використовуються для уникнення переобладнання; шар класифікації, щоб класифікувати кінцеві результати за класами;
- інші шари за потребою.

Шар в CNN складається з тривимірних об'ємів нейронів, як показано на рисунку 2.1 (ширина, висота і глибина, а слово глибина відноситься до того, що називається «картами функцій або картами активації», а не кількістю шарів у CNN).

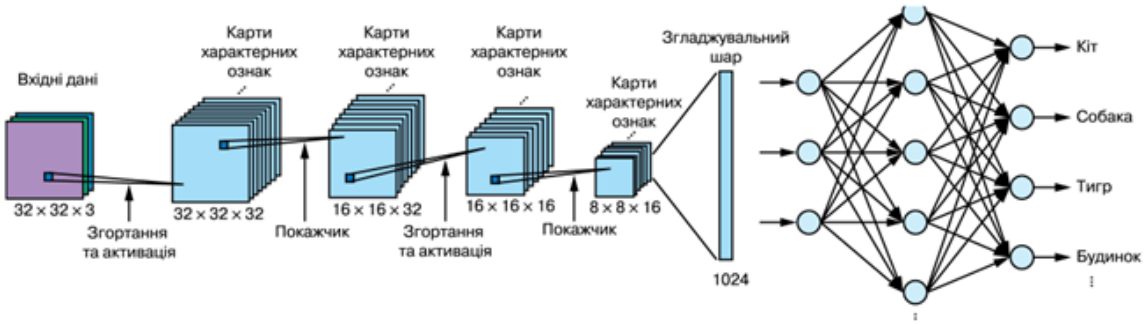


Рисунок 2.1 – Модель згорткової нейронної мережі

Згортковий шар розглядається як основний будівельний блок CNN, і він містить більшість операцій у моделі CNN. Згортковий шар, по суті, виконує математичну операцію під назвою згортка, яка включає тривимірне множення накопичення (MACC) операції. На рисунку 2.2 показано ядро/фільтр (вибір значень фільтра залежить від призначених функцій, а вхідні зображення повинні ділитися на 2 багато разів) ваг, які помножуються на набір вхідних даних (приймальна область), а зважені вхідні дані підсумовуються разом.

Зміщення, значення якого зазвичай одиниця, додається до підсумованих зважених вхідних даних, щоб забезпечити спрацьовування нейронів. Функція активації, така як випрямлена лінійна одиниця (ReLU), застосовується до накопиченої суми, щоб внести нелінійність і обмежити вихід до розумного діапазону. Результати функції активації передаються відповідним нейронам наступного шару.

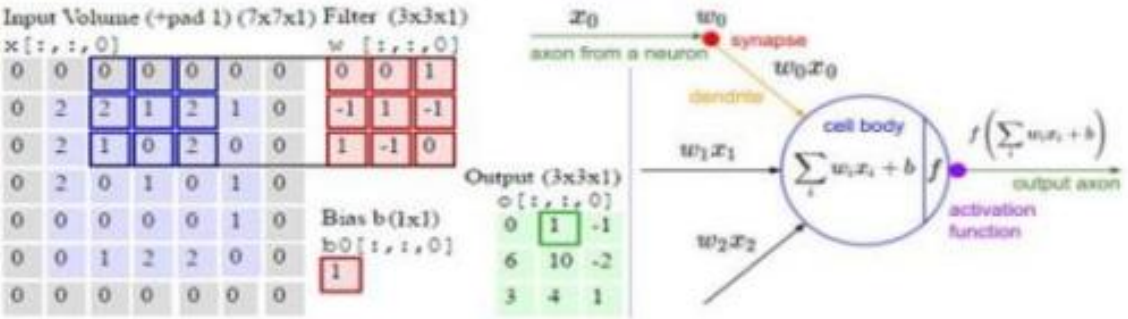


Рисунок 2.2 – Згортковий шар

Три гіперпараметри контролюють розмір вихідних даних: глибина, Stride, Zero-adding, де крок – це швидкість ковзання, з якою фільтр ковзає(найпоширенішим значенням слайда є 1, коли фільтр переміщується на один піксель вправо за раз), а нульове заповнення – це процес, який застосовується до межі вхідних даних, щоб допомогти контролювати просторовий розмір виводу та зберегти інформацію на межі. У CNN загальна кількість параметрів(ваги та зміщення) менша, ніж у звичайних мережах із прямим зв’язком, тоді як не всі нейрони підключені один до одного.

Загальна кількість параметрів зменшується, оскільки застосовуються рецептори локального поля (локальна зв’язність), де нейрони підключаються лише до відповідного локального поля без необхідності підключатися до всіх входів (пікселів на зображенні чи нейронів на карті ознак).

Рецептор поля є загальним для всіх нейронів наступного шару. Наприклад, якщо є N прихованих шарів і рецепторне поле 5×5×3, то загальна кількість параметрів дорівнює - (5×5×3×N) + (N зміщень). На рисунку 2.3 показано вхідне зображення, згорнуте за допомогою фільтра (зеленим), що створює відповідні карти активації (синім кольором). Інша карта активації (зелена) також була створена іншим фільтром з таким же розмір, але різні значення фільтра.

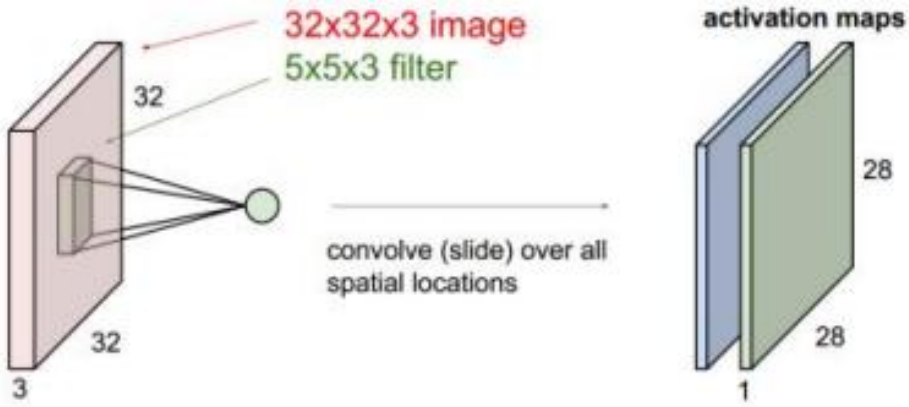


Рисунок 2.3 – CNN 5 x 5

Функція активації застосовується до кожного вхідного пікселя, щоб забезпечити нелінійність мережі, а також позбутися від непотрібної інформації. Серед різних функцій активації, Sigmoid, Tanh і ReLU є найбільш часто використовуваними функціями активації. Функція активації застосовується до кожного вхідного пікселя для забезпечення нелінійності в мережах, а також для позбавлення від непотрібної інформації. Серед різних функцій активації, Sigmoid, Tanh і ReLU є найбільш часто використовуваними функціями активації.

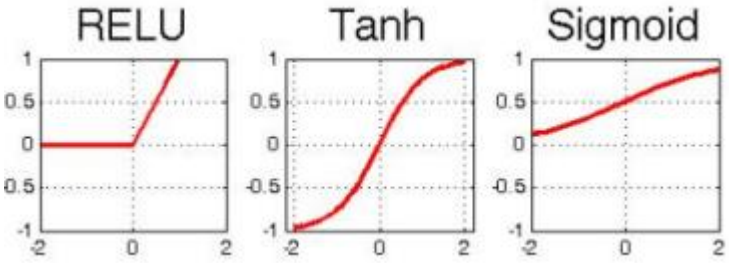


Рисунок 2.4 – Функції активації

Нормалізація або нормалізація локальної реакції (LRN) реалізує латеральне гальмування [16], демпфуючи відповіді, які є рівномірно великими в будь-якому локальному околиці. Перед відправкою зважених входів (виходів) згортки до нелінійності, рівень нормалізації нормалізує виходи залежно від сусідніх нейронів, щоб допомогти привести вхідні дані до ReLU до загального масштабу. Рівень LRN був введений в архітектуру AlexNet [8], але менш поширений в останніх CNN [22].

2.3 Проєктування архітектури згорткової нейронної мережі для розпізнавання емоцій

Розробка архітектури нейронної мережі для задачі розпізнавання емоцій потребує балансу між глибиною мережі (для високої точності) та її

обчислювальною складністю (для роботи в реальному часі). У межах даної роботи було спроєктовано багаторівневу архітектуру CNN, яка складається з наступних функціональних блоків: вхідний шар, блок екстракції ознак, регуляризація, повно зв'язний блок і вихідний шар (рис. 2.5).

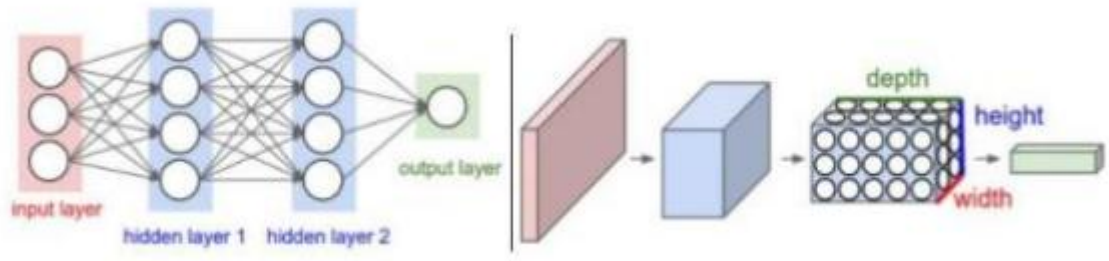


Рисунок 2.5 – Модель згорткової нейронної мережі

Вхідний шар (Input Layer): зображення, що надходять з камери або дата сету, попередньо ресайзяться до фіксованого розміру (наприклад, 48x48 або 64x64 пікселі) та конвертуються у відтінки сірого (1 канал). Це дозволяє мережі зосередитися на мімічних зморшках та контурах обличчя, ігноруючи несуттєву інформацію про колір шкіри чи освітлення.

Блок екстракції ознак (Feature Extraction Blocks). Цей блок складається з декількох послідовних згорткових шарів (Convolutional Layers):

- Згортка (Conv2D) – використовуються фільтри розміром 3x3 для виявлення локальних паттернів (кутики губ, вигин брів).
- Функція активації ReLU – забезпечує нелінійність моделі, дозволяючи їй вивчати складні залежності.
- Пакетна нормалізація (Batch Normalization) – прискорює навчання та робить мережу стійкішою до розкиду значень пікселів.
- Пулінг (MaxPooling) – зменшує розмірність ознак у два рази, залишаючи лише найважливіші сигнали. Це дозволяє системі бути інваріантною до невеликих зсувів обличчя в кадрі.

Регуляризація (Dropout): для запобігання ефекту перенавчання (overfitting), коли мережа просто "зазубрює" тренувальні фото, в архітектуру впроваджено шари Dropout. Вони випадковим чином вимикають частину нейронів (наприклад, 25-50%) під час навчання, змушуючи мережу шукати більш універсальні ознаки емоцій.

Повнозв'язний блок (Fully Connected / Dense Layers): після серії згорток дані "розгортаються" у довгий вектор (Flatten) і проходять через щільні шари нейронів. Саме тут відбувається фінальний логічний висновок на основі всіх виявлених ознак (наприклад: "підняті кутики губ та примружені очі = радість").

Вихідний шар (Output Layer): Кількість нейронів на виході відповідає кількості емоцій, що розпізнаються (в даному випадку їх 7: гнів, огида, страх, щастя, сум, здивування, нейтрально). Використовується функція активації Softmax, яка перетворює вихідні сигнали у ймовірності. Результатом роботи мережі є клас із найвищим відсотком впевненості.

2.4 Формування та структура набору даних (Dataset)

Для навчання та тестування розробленої архітектури нейронної мережі було використано один із найбільш релевантних у галузі наборов даних – FER2013 (Facial Expression Recognition 2013). Цей датасет був створений для платформи Kaggle і складається з тисяч зображень облич, що демонструють різні емоційні стани.

Використаний набір даних містить понад 35,000 зображень, що розподілені за наступними параметрами:

- розмір зображень – 48x48 пікселів (фіксований розмір для всіх екземплярів);
- колірна модель – Grayscale (відтінки сірого), що дозволяє моделі ігнорувати варіації кольору шкіри та фокусуватися на геометрії обличчя;

- класифікація – зображення розділені на 7 основних категорій (емоцій): 0=Angry (Гнів), 1=Disgust (Огида), 2=Fear (Страх), 3=Happy (Радість), 4=Sad (Сум), 5=Surprise (Здивування), 6=Neutral (Нейтрально).

Для об'єктивної оцінки якості моделі та уникнення перенавчання (overfitting), загальний обсяг даних було розділено на три підмножини:

1. Training Set (Навчальна вибірка) – 80%: використовується безпосередньо для корегування ваг нейронної мережі під час ітерацій навчання.

2. Validation Set (Валідаційна вибірка) – 10%: слугує для перевірки точності моделі на кожній епосі навчання та підбору оптимальних гіперпараметрів.

3. Test Set (Тестова вибірка) – 10%: фінальна перевірка навченої моделі на абсолютно нових даних, які мережа раніше не «бачила».

Оскільки деякі класи в датасеті можуть бути незбалансованими (наприклад, зображень «Радості» більше, ніж «Огиди»), було застосовано техніку аугментації даних. Вона полягає у штучному створенні нових варіацій зображень шляхом випадкових поворотів на 10-15 градусів, горизонтального віддзеркалення, масштабування та зсуву, що дозволяє системі стати «робастною» – стійкою до різних ракурсів зйомки та нахилів голови користувача в реальних умовах.

Використання структурованого та анотованого датасету FER2013 забезпечує необхідну статистичну базу для вивчення мімічних ознак, а застосування аугментації дозволяє підвищити генералізаційну здатність моделі.

2.5 Попередня обробка, очищення даних та візуальний аналіз розподілу класів

Якість навчання нейронної мережі безпосередньо залежить від чистоти вхідних даних. Набір даних FER2013, містить певну кількість «шумів»: некоректно розмічені зображення, кадри з низькою контрастністю або зображення, де обличчя перекрите сторонніми предметами.

Для підвищення точності моделі було проведено етап препроцесингу, що включав видалення дублікатів, виключення ідентичних кадрів, які могли потрапити до вибірки під час автоматизованого збору даних. Також була проведена фільтрація за рівнем впевненості та видалення зображень, на яких алгоритм детекції облич не зміг чітко локалізувати ключові антропометричні точки. Вирівнювання яскравості та контрастності зображень дозволило моделі однаково ефективно розпізнавати емоції як на занадто темних, так і на переосвітлених фото.

Важливим етапом аналізу є побудова гістограми розподілу кількості зображень за категоріями емоцій. У типовому датасеті FER2013 спостерігається значний дисбаланс (рис. 2.6):

- домінуючі класи – «Happy» (Радість) та «Neutral» (Нейтрально) зазвичай мають найбільшу кількість прикладів;
- дефіцитні класи – «Disgust» (Огида) часто містить у 5-10 разів менше зображень, ніж інші категорії.

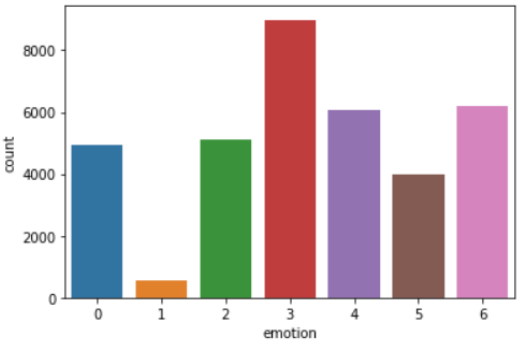


Рисунок 2.6 – Графік розподілу зображень по категоріям

Оскільки нейромережа схильна краще розпізнавати ті класи, яких у вибірці більше, було застосовано метод зважування класів. Під час розрахунку функції втрат помилка на рідкісних класах (наприклад, «Огида») отримує більшу вагу, що змушує модель приділяти їм більше «уваги» під час корекції ваг нейронів.

Останнім кроком підготовки є масштабування значень пікселів з діапазону [0, 255] до діапазону [0, 1]. Це прискорює збіжність градієнтного спуску та запобігає проблемі «вибуху градієнтів» на початкових етапах навчання.

Проведене очищення та збалансування даних дозволило мінімізувати вплив шумів на процес навчання та забезпечити рівномірну точність розпізнавання для всіх семи категорій емоцій.

2.6 Метрики оцінки якості та аналіз результатів розпізнавання

Для об’єктивної оцінки ефективності розробленої нейронної мережі недостатньо використовувати лише загальну точність, оскільки в умовах дисбалансу класів цей показник може бути оманливим. Для детального аналізу було застосовано комплексний підхід із використанням матриці помилок.

У роботі розраховуються наступні показники для кожного класу емоцій:

- Precision (Точність) – відображає частку правильно розпізнаних емоцій певного типу серед усіх випадків, які модель класифікувала як цей тип.
- Recall (Повнота) – показує, яку частку реальних емоцій певного типу з тестової вибірки вдалося знайти моделі.
- F1-Score – гармонійне середнє між Precision та Recall, що є найкращим показником якості для незбалансованих даних.

Найбільш інформативним інструментом аналізу є Confusion Matrix. Це таблиця, де по горизонталі відкладено реальні мітки емоцій, а по вертикалі – прогнози нейромережі.

Діагональні елементи – відображають кількість правильних прогнозів. Чим вищі значення на діагоналі, тим краще працює модель.

Позадіагональні елементи вказують на типи помилок. Наприклад, аналіз матриці часто виявляє типові «перехресні» помилки:

- Fear (Страх) та Surprise (Здивування) – ці емоції часто плутаються через схожість мімічних ознак (широко розплющені очі);
- Sad (Сум) та Neutral (Нейтрально) – незначні зміни в кутиках губ можуть призводити до помилкової класифікації суму як спокійного стану;
- Happy (Радість) – зазвичай має найвищий показник розпізнавання (понад 90%), оскільки усмішка є найбільш вираженою та унікальною ознакою.

Для моніторингу процесу навчання використовуються графіки залежності Loss (втрат) та Accuracy (точності) від номера епохи.

Збіжність – поступове зниження кривої втрат свідчить про успішне навчання.

Діагностика – якщо точність на навчальній вибірці зростає, а на валідаційній починає падати – це сигнал про overfitting (перенавчання), що потребує корекції коефіцієнта Dropout або обсягу даних. Використання матриці помилок дозволило ідентифікувати найбільш складні для розпізнавання емоції та внести корективи в архітектуру мережі. Отримані результати підтверджують високу здатність моделі до генералізації на нових обличчях, не представлених у навчальній вибірці.

2.7 Висновки до розділу 2

У другому розділі було спроєктовано систему розпізнавання емоцій на основі архітектури згорткових нейронних мереж (CNN) із використанням мови Python та бібліотек OpenCV і TensorFlow. Проведено детальну аналіз для підготовки та очищення датасету FER2013, щоб мінімізувати вплив цифрових шумів та дисбалансу класів на процес навчання. Описано структуру моделі,

включаючи шари екстракції ознак, регуляризації та фінальної класифікації емоційних станів. За допомогою матриці помилок та метрик точності виконано аналіз результатів, який підтвердив здатність мережі ефективно ідентифікувати базові мімічні вирази. Таким чином, програмна система є готовим інструментом для інтеграції в прикладні рішення, де необхідний автоматизований аналіз психоемоційного стану користувача в реальному часі.

3 РОЗРОБКА НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ІДЕНТИФІКАЦІЇ ЛЮДСЬКИХ ЕМОЦІЙ

3.1 Завантаження та попередня обробка набору даних

Сьогодні технології комп'ютерного зору пройшли шлях від експериментальних розробок до фундаментальної складової цифрової екосистеми. Серед них розпізнавання облич займає особливе місце, оскільки воно стосується безпосередньої ідентифікації людини, що відкриває широкі можливості для автоматизації та безпеки.

Традиційні методи обробки зображень були чутливими до змін освітлення, ракурсів та виразу обличчя. Поява згорткових нейронних мереж (CNN) здійснила революцію, дозволивши системам самостійно навчатися виділяти ієрархічні ознаки. Це підвищило точність ідентифікації до високого рівня.

Розпізнавання облич стало пріоритетним методом автентифікації у смартфонах, банківських застосунках та системах контролю доступу (СКУД), замінюючи традиційні паролі та відбитки пальців.

Зі зростанням кількості камер відеоспостереження та обсягів мультимедійного контенту в мережі виникла потреба в інтелектуальному аналізі даних у реальному часі. CNN дозволяють ефективно обробляти величезні потоки інформації, забезпечуючи швидкий пошук та верифікацію об'єктів.

Таким чином, дослідження та вдосконалення архітектур CNN для розпізнавання облич є не просто технічним завданням, а необхідністю для розбудови безпечного, швидкого та зручного цифрового суспільства.

На перших етапах розробки були проаналізовані та встановлені всі необхідні бібліотеки, такі як: numpy, pandas, seaborn, tensorflow та інші. Наступним кроком був огляд та аналіз відповідних датасетів. Був обраний

датасет з підготовленими фотографіями людей, поділений на 7 категорій (рис. 3.1).

```
emotion_label_to_text = {0:'anger', 1:'disgust', 2:'fear', 3:'happiness', 4: 'sadness', 5: 'sur
prise', 6: 'neutral'}
```

Рисунок 3.1 – Розподіл зображень по категоріям

Де 0 - злість, 1 - огида, 2 - страх, 3 - щастя, 4 - смуток, 5 - подив, 6 - нейтральне.

Таким чином, більшість об’єктів належать до трьох класів: 3 (Щастя), 4 (Сум) та 6 (Нейтральна емоція), і саме ці три класи є об’єктом нашого дослідження.

На рисунках 3.2 і 3.3 наведені приклади зображень людей у підготовленому датасеті. Можна звернути увагу, що підібрані обличчя людей різних національностей, різного віку, різної довжини волосся, а деякі обличчя прикриті частково одягом.



Рисунок 3.2 – Зображення людей

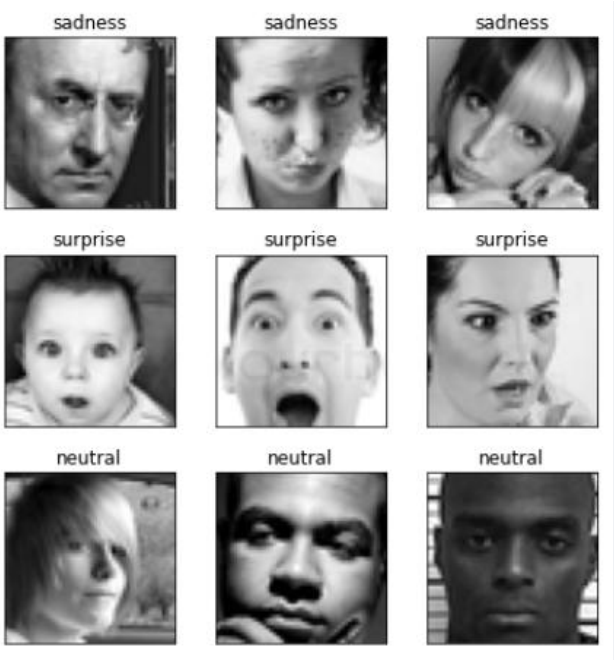


Рисунок 3.3 – Зображення облич з емоціями

Після аналізу обраного датасету були зроблені наступні кроки.

1. Підготовка даних

У завданнях з розпізнавання образів успіх моделі на 70% залежить від якості підготовки даних.

2. Масштабування

Піксельні значення зображень (зазвичай від 0 до 255) перетворюються на діапазон [0, 1]. Це прискорює збіжність нейронної мережі під час навчання.

3. Ресайзинг

Усі вхідні зображення приводяться до єдиного стандарту розміру (наприклад, 48x48 пікселів), щоб мережа могла працювати з фіксованим входом.

4. Розподіл

Дані діляться на три частини: Train (для навчання), Validation (для підбору гіперпараметрів) та Test (для остаточної оцінки якості).

3.2 Етап підготовки та попередньої обробки даних

На цьому етапі ми готуємо дані для нейронної мережі. Комп’ютер не бачить зображення як ми; для нього це велика таблиця чисел (матриця пікселів).

Зміна розміру: усі вхідні зображення приводяться до однакового розміру (наприклад, 48x48), щоб архітектура мережі була стабільною.

Нормалізація: значення пікселів (0-255) діляться на 255, щоб привести їх до діапазону [0, 1]. Це потрібно для того, щоб градієнти під час навчання не ставали занадто великими, що дозволяє мережі навчатися швидше та стабільніше.

Аугментація (опціонально): якщо даних мало, ми штучно їх розширюємо (повороти, віддзеркалення), щоб модель навчилася бути стійкою до різних ракурсів обличчя.

3.3 Архітектура розробленої нейронної мережі

Для роботи з зображеннями найкраще підходять згорткові нейронні мережі (CNN), оскільки вони вміють виділяти ієрархічні ознаки: спочатку краї та лінії, потім геометричні фігури (очі, ніс, губи), а на фінальних шарах – цілісні емоції.

Conv2D (Згорткові шари) виконують фільтрацію зображення для пошуку важливих деталей.

MaxPooling зменшує розмірність карти ознак, допомагаючи мережі фокусуватися на найважливіших особливостях і запобігаючи перенавчанню (overfitting).

Flatten перетворює двовимірну матрицю ознак у лінійний вектор для подальшої класифікації.

Dense (Повнозв’язні шари) приймають виділені ознаки та роблять фінальне рішення. Останній шар має кількість нейронів, що дорівнює кількості емоційних класів, з функцією активації Softmax.

Згорткові шари (Conv2D) пересуваються по зображенню, шукаючи характерні ознаки. Перші шари знаходять прості лінії та кути, наступні – складніші форми (очі, ніс, губи).

Шари підсилення (MaxPooling) зменшують розмір матриці, залишаючи лише найважливіші ознаки. Це допомагає мережі зосередитися на головному і зменшує обчислювальну складність.

Функції активації (ReLU) додають нелінійність. Без них мережа була б просто серією лінійних рівнянь і не змогла б вивчити складні закономірності людських емоцій.

Повнозв’язні шари (Dense) у самому кінці вони збирають усі виділені ознаки разом і приймають рішення: "Це злість" або "Це радість".

На рисунку 3.4 представлена архітектура згорткової нейронної мережі, що є типовою схемою моделі для розпізнавання емоцій на обличчі. Процес обробки даних починається з вхідного шару, який приймає зображення розміром 48 на 48 пікселів. Наявність лише одного колірного каналу вказує на те, що на вхід подаються чорно-білі зображення, тоді як змінний розмір батчу на схемі дозволяє обробляти різну кількість зображень одночасно під час навчання. Далі дані проходять через три послідовні блоки виділення ознак, основою яких є згорткові шари (Conv2D). Вони відповідають за пошук візуальних патернів, починаючи від простих контурів і закінчуючи складними рисами обличчя, причому кількість фільтрів у цих шарах поступово зростає від 64 до 256, дозволяючи мережі вивчати дедалі глибші залежності.

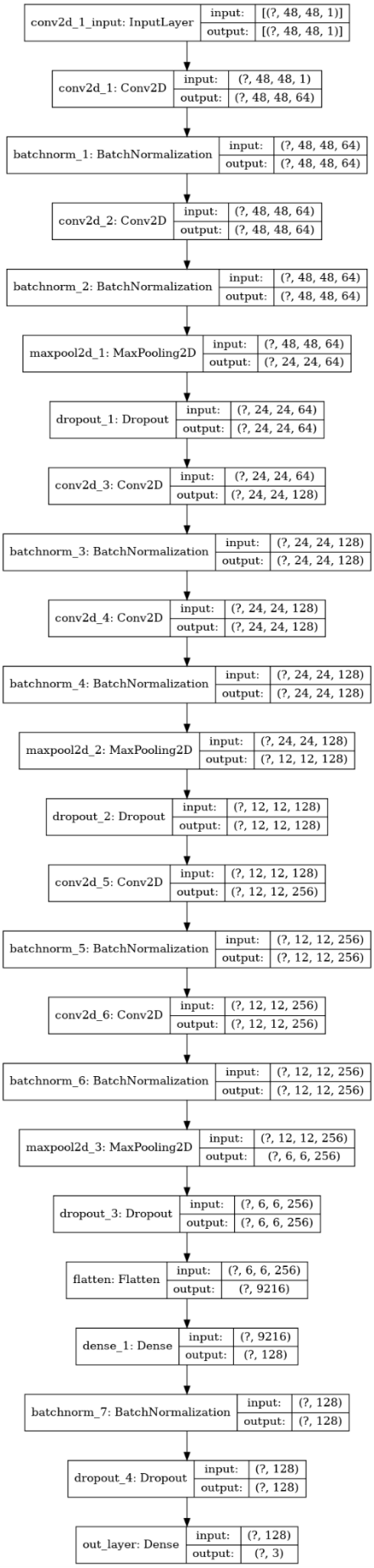


Рисунок 3.4 – Архітектура нейронної мережі

Для стабілізації та прискорення процесу навчання після згорток застосовується пакетна нормалізація, а шари субдискретизації поступово зменшують просторовий розмір карт ознак з 48x48 до 6x6, залишаючи лише найважливішу інформацію та знижуючи обчислювальне навантаження.

Щоб запобігти ефекту перенавчання, використовуються шари проріджування, які випадковим чином тимчасово вимикають частину нейронів. Після завершення етапу виділення ознак багатовимірні дані перетворюються у звичайний одновимірний масив із 9216 значень за допомогою шару вирівнювання. Цей вектор передається до повнозв'язного шару зі 128 нейронами для фінального аналізу виділених характеристик. Завершується архітектура вихідним шаром із трьома нейронами, що свідчить про налаштування цієї конкретної моделі на класифікацію трьох окремих емоційних станів, формуючи загалом класичну, добре збалансовану архітектуру для аналізу візуальних даних.

3.4 Процес навчання

Функція втрат (Loss Function) зазвичай використовується Categorical Crossentropy, яка оцінює різницю між передбаченням мережі та реальною емоцією.

Оптимізатор (Optimizer) Adam. – це математичний механізм, який коригує ваги нейронів після кожної ітерації, щоб мінімізувати помилку.

Епохи – кількість повних проходів через увесь набір даних. Важливо слідкувати, щоб помилка на валідаційній вибірці не почала зростати (ознака перенавчання).

Наступний етап навчання – це процес, де модель помиляється, аналізує помилку і виправляє себе.

Функція втрат (Loss Function) математично вимірює різницю між тим, що модель передбачила, і тим, що було насправді. Наприклад, якщо мережа

впевнена, що емоція - "сум", а насправді - "радість", функція втрат видасть велике значення помилки.

Оптимізатор (наприклад, Adam) – алгоритм, який під час кожного кроку (ітерації) трохи коригує ваги нейронів у мережі, щоб зменшити помилку.

Зворотне поширення – процес, при якому сигнал помилки передається від кінця мережі до початку, оновлюючи ваги.

3.5 Аналіз результатів

Для аналізу результатів важливо оцінювати не тільки загальну точність (Accuracy), а й матрицю помилок (Confusion Matrix):

Вона показує, які емоції модель плутає найчастіше (наприклад, «сум» часто плутають із «нейтральним» виразом).

Це дає підказку, що потрібно змінити: можливо, збільшити кількість прикладів для певної емоції або додати шари регуляризації (Dropout).

Після навчання потрібно перевірити, чи модель справді вивчила емоції, чи просто "зазубрила" навчальні приклади.

Графіки Loss та Accuracy: Якщо помилка на валідації (Validation Loss) починає зростати, поки на навчанні (Train Loss) вона падає – це ознака перенавчання (overfitting). Модель стала "надто розумною" для тренувальних даних і втратила здатність до узагальнення.

Матриця помилок (Confusion Matrix): Це найважливіший інструмент аналізу. Вона показує, які емоції модель плутає між собою (наприклад, чи плутає вона "страх" з "подивом").

3.6 Результати роботи розробленої нейронної мережі

Аналіз графіків навчання – це один із найважливіших етапів роботи Data Scientist, адже саме вони розповідають правду про те, як почувається ваша модель (рис. 3.5).

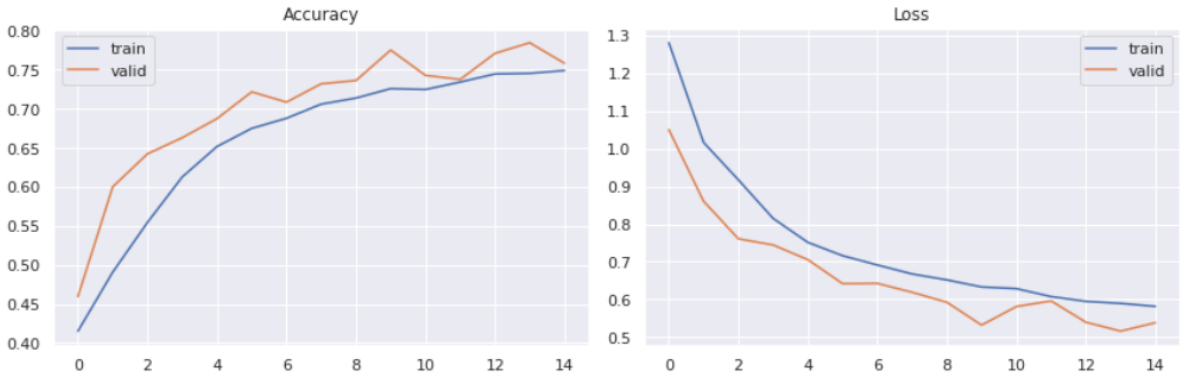


Рисунок 3.5 – Графіки навчання моделі

Точність (Accuracy): Модель досягає піку приблизно на рівні 0.75 - 0.78 (75-78%). Обидві лінії (train та valid) рухаються паралельно, що є ознакою гарного навчання.

Втрати (Loss): Обидві лінії стабільно знижуються. Це означає, що модель продовжує вчитися.

У більшості випадків валідаційна точність нижча за тренувальну. Якщо ж вона вище, це зазвичай означає:

1. Під час навчання використовуються шари Dropout (випадкове вимкнення нейронів), щоб модель не зазубрювала дані. Це робить навчання "важчим". Під час валідації Dropout вимкнено, і модель використовує "повну потужність", тому працює трохи краще.

2. Якщо використовувати аугментацію даних, тренувальні дані стають складнішими, тому модель показує на них нижчу точність, ніж на "чистих" валідаційних даних.

Справжнє перенавчання виглядає так: тренувальна помилка (Loss) падає до нуля, а валідаційна помилка починає різко зростати. У нашому випадку валідаційна помилка (Loss) продовжує плавно знижуватися навіть на 14-й епосі. Валідаційна точність не падає.

Отже, модель не перенавчена, вона, швидше за все, досягла свого плато – вона більше не може вивчити нові закономірності з тих даних, які має.

Використання Violin Plot дозволило оцінити розподіл метрик за весь період навчання. Це підтверджує, що модель більшу частину часу перебуває в зоні оптимальної роботи (висока точність, низька помилка) і не демонструє критичних розбіжностей між тренувальним та валідаційним наборами, що свідчить про хорошу здатність моделі до узагальнення (рис. 3.6).

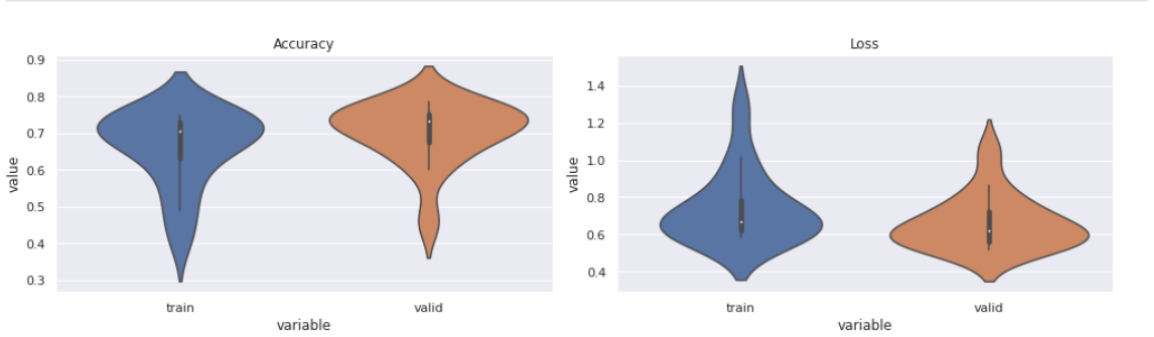


Рисунок 3.6 – Використання Violin Plot

3.7 Аналіз матриці помилок (Confusion Matrix)

Матриця помилок дозволяє оцінити якість класифікації для кожного з трьох класів (0, 1, 2). Головна діагональ (від верхнього лівого до нижнього правого кута) відображає кількість правильно класифікованих зразків, тоді як інші клітинки показують помилкові класифікації.

Модель демонструє відмінні результати для Класу 0. Із загальної кількості зразків (899) правильних передбачень – 822. Це вказує на те, що ознаки цього класу є найбільш чіткими та легко розпізнаваними для моделі.

Клас 1 є найбільш проблемним для моделі. Спостерігається значна плутанина: модель часто помилково класифікує зразки Класу 1 як Клас 2 (290 помилок). Це свідчить про те, що або межі між цими класами занадто розмиті, або модель не має достатньо інформативних ознак для їх розділення.

Модель впевнено розпізнає Клас 2 (523 правильні відповіді), проте також має невеликі труднощі з диференціацією Класу 1.

На рисунку 3.7 зображена матриця помилок на основі вхідних даних.

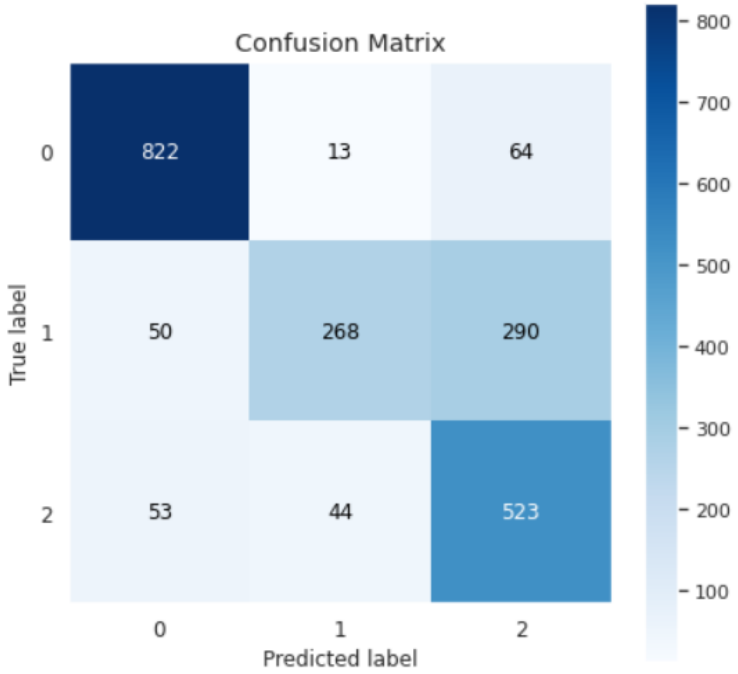


Рисунок 3.7 – Матриця помилок

Загальна точність (Accuracy): модель правильно класифікувала 2127, 75.8% зразків.

Дисбаланс даних – кількість прикладів для Класу 0 (899) значно перевищує кількість прикладів для Класу 1 (608) та Класу 2 (620). Це може бути причиною того, що модель "краще вивчила" Клас 0.

3.8 Приклади роботи нейронної мережі

Після аналізу отриманих результатів було прийняте рішення протестувати розроблену нейронну мережу на розпізнавання тільки трьох категорій емоцій – щастя, сум та нейтральне.

В результаті отримали визначення емоцій як зображено на рисунку 3.8.

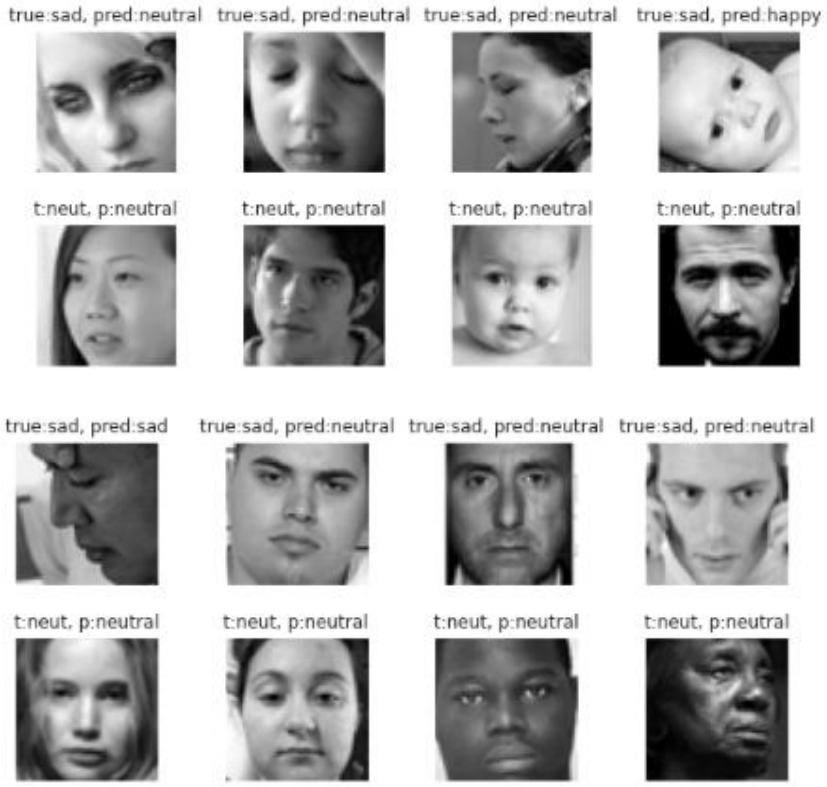


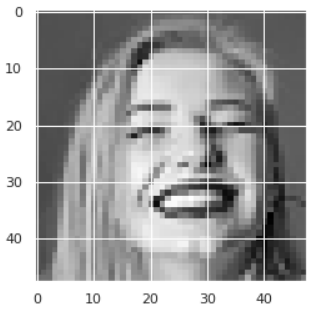
Рисунок 3.8 – Результати роботи нейромережі

Коли візуальна різниця між «Нейтральним» і «Сумним» ледь помітна наприклад, через злегка опущені кутики рота чи «важкі» повіки – моделі важко, бо ознаки недостатньо чіткі для розділення. У такій ситуації зазвичай допомагають кілька речей: нормалізація за суб'єктом: Якщо в наборі даних є кілька фото однієї людини, модель може навчитися розуміти, як виглядає саме її нейтральне обличчя порівняно з сумним. Іноді самого обличчя замало; мова тіла або навіть нахил голови можуть бути вирішальними факторами. Замість того, щоб змушувати модель обирати щось одне, деякі

дослідники використовують ймовірнісні цілі (наприклад, 60% - нейтральне, 40% - сумне), щоб врахувати цю людську неоднозначність.

Далі я протестував розроблену нейронну мережу на зображеннях зірок і отримав наступні результати – рисунки 3.9-3.12.

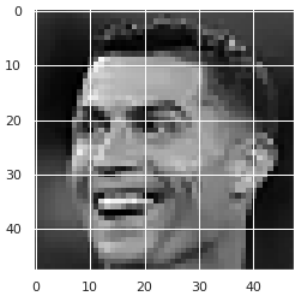
In [41]:
```
happy_girl_48x48 = cv2.imread("/kaggle/input/testemotion/happy_girl_48x48.png")
#Convert an image from BGR to grayscale mode
happy_girl_48x48 = cv2.cvtColor(happy_girl_48x48, cv2.COLOR_BGR2GRAY)
happy_girl_48x48 = happy_girl_48x48.astype(np.float32)
pyplot.imshow(happy_girl_48x48, cmap='gray')
pyplot.show()
mapper[model.predict_classes(happy_girl_48x48.reshape(1,48,48,1))[0]]
```



Out[41]:
```
'happy'
```

Рисунок 3.9 – Тестування розробленої нейромережі

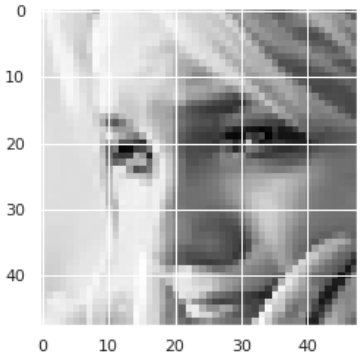
In [38]:
```
ronaldo_happy_48x48 = cv2.imread("/kaggle/input/testemotion/ronaldo_happy_48x48.png")
#Convert an image from BGR to grayscale mode
ronaldo_happy_48x48 = cv2.cvtColor(ronaldo_happy_48x48, cv2.COLOR_BGR2GRAY)
ronaldo_happy_48x48 = ronaldo_happy_48x48.astype(np.float32)
pyplot.imshow(ronaldo_happy_48x48, cmap='gray')
pyplot.show()
mapper[model.predict_classes(ronaldo_happy_48x48.reshape(1,48,48,1))[0]]
```



Out[38]:
```
'happy'
```

Рисунок 3.10 – Тестування розробленої нейромережі

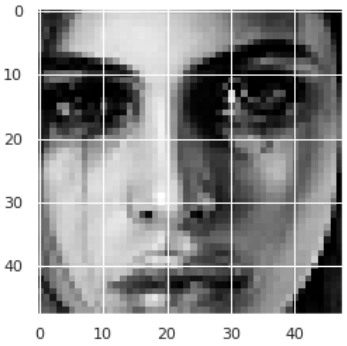
In [35]:
```
sad_britney = cv2.imread("/kaggle/input/testemotion/britney_sad_48x48.png")
#Convert an image from BGR to grayscale mode
sad_britney_bw = cv2.cvtColor(sad_britney, cv2.COLOR_BGR2GRAY)
sad_britney_bw = sad_britney_bw.astype(np.float32)
pyplot.imshow(sad_britney_bw, cmap='gray')
pyplot.show()
mapper[model.predict_classes(sad_britney_bw.reshape(1,48,48,1))[0]]
```



Out[35]:
```
'sad'
```

Рисунок 3.11 – Тестування розробленої нейромережі

In [39]:
```
sad_girl_48x48 = cv2.imread("/kaggle/input/testemotion/sad_girl_48x48.png")
#Convert an image from BGR to grayscale mode
sad_girl_48x48 = cv2.cvtColor(sad_girl_48x48, cv2.COLOR_BGR2GRAY)
sad_girl_48x48 = sad_girl_48x48.astype(np.float32)
pyplot.imshow(sad_girl_48x48, cmap='gray')
pyplot.show()
mapper[model.predict_classes(sad_girl_48x48.reshape(1,48,48,1))[0]]
```



Out[39]:
```
'neutral'
```

Рисунок 3.12 – Тестування розробленої нейромережі

ВИСНОВКИ

В результаті кваліфікаціцйної роботи було розроблено нейронну мережу для розпізнавання емоцій людей на фотографіях.

Розроблена модель демонструє базову здатність до класифікації емоцій з точністю близько 75-78%, проте поточна архітектура потребує балансування даних та корекції ваг класів для підвищення чутливості до виражених емоційних станів, відмінних від нейтрального

Точність розпізнавання емоцій нейронною мережею залежить від трьох основних факторів: якість та обсяг даних – моделі потрібна велика кількість різноманітних прикладів. Якщо даних мало, мережа просто не матиме достатньо бази для порівняння. Об’єктивність розмітки – оскільки емоції суб'єктивні, успіх залежить від того, наскільки чітко люди визначили класи. Якщо навіть людина вагається, чи обличчя нейтральне, чи сумне, нейромережа отримає суперечливі сигнали під час навчання. Індивідуальні особливості – анатомія обличчя кожної людини унікальна. Базова міміка одного суб’єкта може збігатися за ознаками з емоцією іншого, тому найсучасніші системи намагаються враховувати контекст або порівнювати обличчя з іншими фото тієї ж самої людини.

У підсумку можна зазначити, що матеріали, у роботі викладені у повному обсязі, а також можна стверджувати, що всі поставлені завдання до кваліфікаційної роботи реалізовано, отже, мету дослідження досягнуто.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Роль невербальних засобів комунікації: жести, міміка, тілесна позиція, інтонація. *Optima*. URL: https://studfile.net/preview/10355129/page:3/ (дата звернення 21.04.2026).

2. Face Recognition Technology (FERET). *Data Science Today*. URL: https://datasciencetoday.net/index.php/en-us/datasets/[33-face-recognitiontechnology-feret (дата звернення: 25.04.2026).

3. Frontiers. URL: https://www.frontiersin.org (дата звернення: 26.05.2026).

4. Емоції, їх значення та вплив на життя людей. *Rozmova*. URL: https://www.rozmova.me/blog/emociyi-yih-znachennya-ta-vpliv-na-zhittya-lyudey (дата звернення: 27.04.2026).

5. Роль невербальних засобів комунікації: жести, міміка, тілесна позиція, інтонація URL: https://optima.study/blog/rol-neverbal-nykh-zasobiv-komunikatsiyi-zhesty-mimika-tilesna-pozytsiya-intonatsiya (дата звернення 27.05.2026).

6. Науковий портал Львівської політехніки. URL: https://science.lpnu.ua (дата звернення: 26.05.2026).

7. Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки. URL: https://tech.vernadskyjournals.in.ua (дата звернення: 26.05.2026).

8. Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки. URL: https://tech.vernadskyjournals.in.ua (дата звернення: 26.05.2026).

9. Discover scientific knowledge and stay connected to the world of science. ResearchGate. URL: https://www.researchgate.net (дата звернення: 26.05.2026).

10. Найкраща модель розпізнавання облич у 2025 році? А також, як створити її з нуля для використання в промисловості? Reddit. URL: https://www.reddit.com (дата звернення: 26.05.2026).

11. ACM Digital Library. URL: https://dl.acm.org (дата звернення: 26.05.2026).

12. Nature. URL: https://www.nature.com (дата звернення: 26.05.2026).

13. Codefinity. URL: https://codefinity.com (дата звернення: 22.05.2026).

14. Електронний архів ХНУРЕ. URL: https://openarchive.nure.ua (дата звернення: 12.05.2026).

15. Електронний архів КПІ ім. Ігоря Сікорського (ELA KPI). URL: https://ela.kpi.ua (дата звернення: 16.05.2026).

16. SpringerLink. URL: https://link.springer.com (дата звернення: 06.05.2026).

17. Centre for Research and Technology Hellas (CERTH). URL: https://www.iti.gr (дата звернення: 02.05.2026).

18. Kristall Systems. URL: https://kristall-systems.net.ua (дата звернення: 25.05.2026).

19. MDPI. URL: https://www.mdpi.com (дата звернення: 24.05.2026).

20. arXiv. URL: https://arxiv.org (дата звернення: 23.05.2026).

21. Вступ до OpenCV. Комп'ютерний зір. IT Master. URL: https://itmaster.biz.ua/programming/vision/opencv.html (дата звернення: 26.05.2026).

22. Tensorflow що це і які основні можливості застосування. FoxmindEd. URL: https://foxminded.ua/tensorflow-shcho-tse/ (дата звернення: 22.05.2026).

23. Python Підручник NumPy для початківців: навчайтеся на прикладах. Guru99. URL: https://www.guru99.com/uk/numpy-tutorial.html (дата звернення: 26.05.2026).

24. Інтеграція Matplotlib та Pandas. Metanit. URL: https://metanit.com/python/pandas/9.2.php (дата звернення: 20.05.2026).

25. MEDIAPIPE – AI инструмент для Computer vision. Habr. URL: https://habr.com/ru/articles/596043/ (дата звернення: 21.05.2026).

26. Aquino M. Fake BACS Remittance Emails Delivers Dridex Malware. Cyren Blog. 2014. URL: https://blog.cyren.com/articles/fake-bacsremittance-emails-delivers-dridex-malware.html (дата звернення: 25.05.2026).

Декларація
академічної доброчесності
здобувача освіти ВСП «Економіко-правничий фаховий коледж ЗНУ»

Я, Чопоров Микита Сергійович, здобувач освіти 4 курсу, спеціальності 121 Інженерія програмного забезпечення, освітньо-професійної програми Розробка програмного забезпечення, групи К121-22, адреса електронної пошти

- підтверджую, що написана мною кваліфікаційна робота на тему «Розробка нейронної мережі для ідентифікації людських емоцій» відповідає вимогам академічної доброчесності та не містить порушень, що визначені у ст. 42 Закону України «Про освіту», зі змістом яких ознайомлений/ознайомлена;

- заявляю, що надана мною для перевірки електронна версія роботи є ідентичною її друкованій версії;

- згоден на перевірку моєї роботи на відповідність критеріям академічної доброчесності у будь-який спосіб, у тому числі за допомогою інтернет-системи, а також на архівування моєї роботи в базі даних цієї системи.

Дата 05.06.2026 Підпис  Микита ЧОПОРОВ
Ім’я ПРІЗВИЩЕ (здобувача освіти)

Дата 05.06.2026 Підпис _____ Дмитро ДРАЄВСЬКИЙ
Ім’я ПРІЗВИЩЕ (наукового керівника)