МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра комп’ютерної інженерії

ГАЛУНЬКА Богдан Володимирович

Програмний модуль класифікації емоційних станів людини засобами нейронних мереж / Software module for classifying human emotional states using neural networks

спеціальність: 123 – Комп’ютерна інженерія
освітньо–професійна програма – Комп’ютерна інженерія

Кваліфікаційна робота

Виконав: студент групи КІ–41
ГАЛУНЬКА Богдан Володимирович

Науковий керівник
к.т.н. доц. Піцун О.Й.

ТЕРНОПІЛЬ – 2026

АНОТАЦІЯ

Галунька Б.В. Програмний модуль класифікації емоційних станів людини засобами нейронних мереж. – Рукопис.

Кваліфікаційна робота на здобуття освітнього ступеня «бакалавр» за спеціальністю 123 «Комп'ютерна інженерія», освітньо-професійна програма. Західноукраїнський національний університет, Тернопіль, 2026.

Робота написана обсягом 78 сторінок і містить 9 рисунків, 4 таблиці, 3 додатки та 39 джерел за переліком посилань.

Метою кваліфікаційної роботи є розроблення програмного модуля класифікації стану людських емоцій з використанням візуальних образів на основі модифікованої згорткової нейронної мережі архітектури Mini-Xception з інтегрованим блоком каналової уваги Squeeze-and-Excitation. Розроблена модифікована архітектура забезпечує класифікацію семи базових емоцій у режимі реального часу на пристроях з обмеженими обчислювальними ресурсами. Реалізовано програмний модуль мовою Python та проведено експериментальне дослідження за методологією ablation study.

Ключові слова: НЕЙРОННІ МЕРЕЖІ, CNN, КЛАСИФІКАЦІЯ ЕМОЦІЙ, КОМП'ЮТЕРНИЙ ЗІР, MINI-XCEPTION, SE-БЛОК, PYTHON.

ANNOTATION

Halunka B.V. Software module for classifying human emotional states using neural networks. – Manuscript.

Qualification work for the Bachelor degree in specialty 123 “Computer Engineering”, educational and professional program. Western Ukrainian National University, Ternopil, 2026.

The work is written in 78 pages and contains 9 figures, 4 tables, 3 appendices and 39 references.

The purpose of the qualification work is to develop a software module for classifying human emotional states based on visual face images using a modified convolutional neural network of the Mini-Xception architecture with an integrated Squeeze-and-Excitation channel attention block. The developed modified architecture provides classification of seven basic emotions in real time on devices with limited computational resources. The software module is implemented in Python and experimental research was conducted using the ablation study methodology.

Keywords: NEURAL NETWORKS, CNN, EMOTION CLASSIFICATION, COMPUTER VISION, MINI-XCEPTION, SE-BLOCK, PYTHON.

ЗМІСТ

					КР.КІ. 10658521/22.00.00.000 ПЗ		
Змн.	Лист	№ докум.	Підпис	Дата			
Розробив		Галунька Б.			ПРОГРАМНИЙ МОДУЛЬ КЛАСИФІКАЦІЇ ЛЮДСЬКИХ ОБЛИЧ ЗАСОБАМИ НЕЙРОННИХ МЕРЕЖ	Літ.	Арк. 1 / Акрушів 74
Перевір.		Піцун О.Й.					
Консульт.		Савка Н.Я.				ЗУНУ,ФКІТ, КІ–41	
Н. Контр.		Дубчак Л.О.					
Затвердив		Дубчак Л.О.					

ВСТУП

Сучасний розвиток інформаційних технологій та штучного інтелекту суттєво змінив підходи до взаємодії людини з обчислювальними системами. Окрім традиційного виконання явних команд через клавіатуру або сенсорний екран,програмні засоби все частіше орієнтуються на розуміння контексту спілкування, у тому числі емоційного стану людини. Здатність обчислювальної системи розпізнавати емоції за виразом обличчя відкриває нові можливості для систем людино-машинної взаємодії, медичних інформаційних систем, освітніх платформ, маркетингової аналітики та засобів забезпечення безпеки на транспорті. Враховуючи стрімкий розвиток засобів автоматизації, потреба у надійних програмних модулях аналізу емоцій стає необхідним інструментом для широкого спектру прикладних задач.

Серед усіх каналів передачі емоційної інформації найбільш інформативним є людське обличчя, оскільки рухи мімічних м'язів безпосередньо пов'язані з діяльністю нервової системи і часто є мимовільними. Сучасні згорткові нейронні мережі здатні з високою точністю виділяти візуальні ознаки емоційного стану з зображення обличчя та класифікувати його на одну з базових емоцій - радість, сум, гнів, страх, здивування, огиду чи нейтральний стан. Проте практичне впровадження подібних модулів стикається з низкою проблем: критична залежність від якості мережевого з'єднання у хмарних рішеннях, високі вимоги до апаратного забезпечення у важких архітектурах глибокого навчання, а також проблема приватності біометричних даних користувача. Це визначає актуальність розроблення автономного, легкого та конфіденційного програмного модуля класифікації емоцій, орієнтованого на локальне виконання на пристрої кінцевого користувача.

Об'єкт дослідження - процес автоматизованої класифікації емоційних станів людини за візуальними образами обличчя.

Предмет дослідження - методи та архітектури згорткових нейронних мереж для класифікації базових емоцій у режимі реального часу на пристроях з обмеженими обчислювальними ресурсами.

Мета кваліфікаційної роботи - розроблення програмного модуля класифікації стану людських емоцій з використанням візуальних образів на основі модифікованої згорткової нейронної мережі архітектури Mini-Xception з інтегрованим блоком каналової уваги Squeeze-and-Excitation.

Для досягнення поставленої мети у роботі вирішено такі задачі:

1) провести аналіз предметної області автоматизованого розпізнавання емоцій, теоретичних моделей репрезентації емоцій та проблем класифікації у неконтрольованих умовах;

2) зробити порівняльний аналіз існуючих технологій класифікації емоцій - класичного машинного навчання, хмарних API-сервісів та локальних нейронних мереж;

3) розробити узагальнений алгоритм роботи програмного модуля та проаналізовано існуючі архітектури згорткових нейронних мереж;

4) обґрунтовати вибір датасету FER2013 як основного навчального набору;

5) розробити модифіковану архітектуру згорткової нейронної мережі на основі Mini-Xception з інтегрованим SE-блоком каналової уваги;

6) реалізовати програмний модуль мовою Python та проведено експериментальне дослідження за методологією ablation study;

7) здійснити техніко-економічне обґрунтування розробки.

У роботі використано методи комп'ютерного зору для детекції обличчя, нормалізація зображень, методи глибокого машинного навчання, а саме згорткові нейронні мережі, глибинно-сепарабельна згортка, механізм каналової уваги, методи регуляризації нейромереж (Focal Loss, MixUp-аугментація, dropout), теорії штучних нейронних мереж, статистичної обробки результатів експериментів, а також методологію ablation study для кількісного оцінювання внеску окремих модифікацій архітектури.

Практичне значення одержаних результатів полягає у можливості використання розробленого програмного модуля у системах моніторингу втоми водіїв, у застосунках дистанційного навчання, у медичних інформаційних системах для об'єктивної оцінки емоційного стану пацієнтів, у засобах маркетингової аналітики, а також як базовий компонент у складніших програмних комплексах. Мінімальні апаратні вимоги роблять модуль придатним для широкого розгортання, а локальний характер обробки забезпечує конфіденційність біометричних даних.

Основні результати кваліфікаційної роботи опубліковано у тезах доповіді на науково-практичній конференції. Копія публікації наведена у додатках.

Кваліфікаційна робота складається зі вступу, трьох розділів, висновків, списку використаних джерел (39 найменувань) та додатків. Загальний обсяг основної частини становить близько 50 сторінок. Робота містить 9 рисунків та 4 таблиці. У першому розділі розглянуто предметну область, теоретичні моделі емоцій, технології комп'ютерного зору та виконано порівняльний аналіз існуючих рішень. У другому розділі описано узагальнений алгоритм роботи модуля, виконано класифікацію існуючих видів алгоритмів та обґрунтовано вибір датасету. У третьому розділі розроблено модифіковану архітектуру нейронної мережі, виконано програмну реалізацію та експериментальне дослідження за методологією ablation study.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ІСНУЮЧИХ РІШЕНЬ

1.1 Аналіз проблематики розпізнавання емоцій

Стрімкий розвиток інформаційних технологій та штучного інтелекту зумовив перехід від традиційних парадигм людино-машинної взаємодії (Human-Computer Interaction, HCI) до створення високоадаптивних, інтелектуальних систем. Якщо раніше взаємодія обмежувалася жорстко заданими командами через графічні або текстові інтерфейси, то сьогодні виникла критична потреба у системах, здатних розуміти невербальний контекст комунікації. Фундаментальним елементом цього невербального контексту є емоційний стан користувача. Здатність обчислювальних систем розпізнавати, інтерпретувати та адекватно реагувати на людські емоції сформувала окремий потужний міждисциплінарний напрямок у комп'ютерних науках - афективні обчислення (Affective Computing), основи якого були закладені Розалінд Пікард у Массачусетському технологічному інституті (MIT) наприкінці 1990-х років.

Емоції відіграють вирішальну роль у процесах прийняття рішень, сприйнятті інформації, соціальній взаємодії та формуванні когнітивного навантаження. Предметною областю даного дослідження є автоматизований процес виділення та класифікації емоційних станів людини на основі аналізу візуальних образів обличчя за допомогою методів комп'ютерного зору. Обличчя є найбільш інформативним каналом трансляції емоцій, оскільки рухи мімічних м'язів тісно пов'язані з діяльністю вегетативної нервової системи та підкіркових структур головного мозку, що часто робить їх прояв мимовільним і важко контрольованим.

Теоретичним фундаментом для побудови систем автоматичного розпізнавання міміки є психологічні моделі репрезентації емоцій. Історично найбільш впливовою є дискретна теорія базових емоцій, розроблена американським психологом Полом Екманом. На основі масштабних крос-культурних досліджень Екман довів, що мімічні вирази певних станів є

біологічно універсальними для всього людства, незалежно від культурного чи етнічного середовища. До таких базових емоцій відносять: радість, сум, гнів, страх, здивування та огиду. Для задач машинного навчання цей класичний перелік завжди доповнюється сьомим станом - нейтральним (Neutral), який слугує еталонною точкою відліку (baseline) для калібрування алгоритмів під індивідуальні особливості будови обличчя конкретного користувача.

Для точної математичної та анатомічної формалізації міміки П. Екман та В. Фрізен розробили Систему кодування лицьових рухів (Facial Action Coding System, FACS). Ця система декомпозує будь-який складний вираз обличчя на мінімальні неподільні компоненти - рухові одиниці (Action Units, AU). Кожна така одиниця відповідає скороченню одного специфічного мімічного м'яза або невеликої групи м'язів. Наприклад, активація AU1 та AU2 (внутрішня та зовнішня частини лобового м'яза) призводить до підняття брів, що є характерним маркером здивування. Одночасна активація AU4 (зведення брів), AU5 (розширення повік) та AU7 (напруження повік) формує базовий патерн гніву. Більшість сучасних алгоритмів комп'ютерного зору функціонують саме за цією логікою: вони навчаються виявляти або конкретні рухові одиниці (через локалізацію антропометричних точок), або загальні піксельні патерни, що відповідають певним комбінаціям AU.

Альтернативним підходом у предметній області є неперервна (вимірна) модель емоцій, зокрема циркумплексна модель Джеймса Рассела. Вона описує емоції не як набір ізольованих категорій, а як точки у двовимірному просторі, де вісь X відповідає за валентність (Valence: від вкрай негативної до вкрай позитивної емоції), а вісь Y - за інтенсивність або збудження (Arousal: від сонливості до сильного хвилювання). Хоча ця модель точніше відображає складні перехідні емоційні стани, у розробці інженерних програмних модулів частіше використовується саме дискретна модель Екмана, оскільки вона дозволяє звести задачу до класичної проблеми багатокласової класифікації (Multi-class Classification), яка ефективно вирішується сучасними згортковими нейронними мережами.

Специфіка предметної області також вимагає розмежування макроекспресій та мікроекспресій. Макроекспресії - це яскраво виражені, тривалі (від 0,5 до 4 секунд) емоційні реакції, які легко розпізнаються як людьми, так і базовими алгоритмами. Натомість мікроекспресії є вкрай короткочасними (тривалістю від 1/25 до 1/5 секунди) та низькоамплітудними мимовільними скороченнями м'язів. Вони виникають тоді, коли людина намагається свідомо приховати або придушити свою справжню емоцію. Розпізнавання мікроекспресій є викликом найвищого рівня складності для систем комп'ютерного зору, оскільки вимагає використання високошвидкісних камер (з фреймрейтом понад 60–120 fps) та надшвидких алгоритмів постобробки кадрів.

Актуальність створення надійних програмних модулів класифікації емоцій підтверджується стрімким зростанням попиту на такі технології у різноманітних прикладних сферах.

Першою такою сферою є автомобільна промисловість та системи безпеки (Automotive & Driver Monitoring Systems). Однією з головних причин дорожньо-транспортних пригод є людський фактор, тому системи комп'ютерного зору, інтегровані в бортовий комп'ютер, здатні в режимі реального часу аналізувати міміку водія. Виявлення маркерів сильної втоми, зниження концентрації, засинання (аналіз частоти моргання та закриття очей) або нападів агресії (road rage) дозволяє системі завчасно подати звуковий сигнал або навіть активувати протоколи екстреного гальмування.

Не менш важливою сферою є охорона здоров'я, психіатрія та телемедицина. У клінічній практиці аналіз міміки стає важливим інструментом об'єктивного моніторингу пацієнтів, оскільки зниження мімічної активності (емоційне сплощення) є вагомим діагностичним критерієм при виявленні клінічної депресії, шизофренії або ранніх стадій нейродегенеративних захворювань (наприклад, хвороби Паркінсона). Крім того, автоматизовані модулі використовуються для об'єктивної оцінки інтенсивності больового синдрому у пацієнтів реанімаційних відділень або немовлят, які не здатні самостійно описати свої відчуття.

Окремий напрям становить маркетингова аналітика та рітейл (Neuromarketing & Emotion AI). Класичні методи збору споживчого досвіду (анкетування, фокус-групи) часто є упередженими та не відображають справжніх реакцій клієнтів, натомість програмні засоби аналізу емоцій дозволяють фіксувати підсвідомі, спонтанні мімічні реакції користувачів під час перегляду рекламного контенту, трейлерів або взаємодії з інтерфейсами додатків. Це дає бізнесу унікальні інсайти для оптимізації продукту під реальний емоційний відгук цільової аудиторії.

Нарешті, значного поширення набули освітні технології (EdTech) та платформи дистанційного навчання. В умовах глобалізації онлайн-освіти викладач втрачає можливість візуально контролювати аудиторію, тому інтеграція модулів розпізнавання емоцій у системи відеоконференцзв'язку дозволяє агрегувати аналітику щодо рівня залученості студентів. Розпізнавання патернів нудьги, фрустрації або нерозуміння дає змогу адаптувати темп лекції, повторити складний матеріал або змінити формат подачі інформації.

Незважаючи на широкі перспективи, реалізація надійного програмного модуля стикається з низкою фундаментальних науково-технічних проблем, які прийнято об'єднувати терміном «проблема неконтрольованих умов» (In-the-wild recognition problem). Більшість ранніх алгоритмів демонстрували точність понад 95 % у стерильних лабораторних умовах (рівномірне фронтальне освітлення, чіткий анфас, відсутність фону), але їхня ефективність катастрофічно падала при розгортанні у реальному середовищі.

Першою глобальною проблемою є варіативність освітлення. Зміни яскравості, напрямку джерела світла, наявність глибоких тіней на половині обличчя або зустрічне світло (відблиски) суттєво спотворюють піксельні градієнти, на які спираються алгоритми екстракції ознак.

Другою проблемою є зміна ракурсу. Користувач рідко дивиться прямо в об'єктив камери. Нахили голови вперед/назад (Pitch), повороти вліво/вправо (Yaw) та нахили до плеча (Roll) призводять до геометричного спотворення

пропорцій обличчя. Деякі мімічні зони стають невидимими для камери, що руйнує просторову структуру, до якої звикла нейронна мережа під час навчання.

Третім критичним фактором є часткове перекриття обличчя (Occlusions). Використання медичних масок, сонцезахисних окулярів, наявність густої бороди, зачіски, що падає на очі, або просто звичка людини підпирати обличчя рукою унеможливлюють зчитування ключових рухових одиниць. Система повинна бути достатньо стійкою, щоб класифікувати емоцію лише за видимою частиною обличчя (наприклад, розпізнати посмішку лише за зміною форми очей при носінні маски).

Враховуючи вищезазначені фактори, створення сучасного програмного модуля класифікації людських емоцій вимагає відмови від спрощених евристичних підходів. Необхідне залучення математичних моделей глибокого навчання, використання репрезентативних наборів даних (датасетів), зібраних у «диких» умовах, та застосування методів попередньої обробки зображень для нівелювання впливу деструктивних факторів навколишнього середовища.

1.2 Аналіз технологій комп'ютерного зору та машинного навчання

Історичний розвиток методів автоматичного розпізнавання візуальних образів, зокрема класифікації емоційних станів людини, еволюціонував від класичних алгоритмів комп'ютерного зору (Computer Vision) з ручним конструюванням ознак до сучасної ери глибокого машинного навчання (Deep Learning), яка базується на використанні багатошарових штучних нейронних мереж. Цей перехід зумовлений експоненційним зростанням обчислювальних потужностей (зокрема, графічних процесорів GPU) та появою великих розмічених наборів даних (датасетів).

Класичні методи комп'ютерного зору спираються на парадигму двохетапної обробки. На першому етапі інженер-дослідник повинен вручну

розробити математичний алгоритм для виділення значущої інформації з масиву пікселів (Feature Extraction). Для задачі розпізнавання облич та емоцій традиційно використовувалися такі дескриптори, як локальні бінарні шаблони (Local Binary Patterns, LBP), гістограми спрямованих градієнтів (Histogram of Oriented Gradients, HOG) або фільтри Габора. Наприклад, метод HOG розбиває зображення на невеликі комірки та обчислює напрямок і величину градієнтів яскравості пікселів у кожній з них. Отриманий вектор ознак передається на вхід класифікаторам машинного навчання, таким як метод опорних векторів (Support Vector Machine, SVM) або алгоритм випадкового лісу (Random Forest). Основною проблемою цього підходу є те, що жорстко задані математичні правила екстракції ознак не здатні адаптуватися до високої варіативності реальних даних (зміни освітлення, ракурсу, наявності артефактів), що призводить до низької точності розпізнавання у неконтрольованих умовах.

Сучасний етап розвитку систем людино-машинної взаємодії повністю спирається на глибоке навчання, а саме на згорткові нейронні мережі. На відміну від класичних методів, CNN реалізують концепцію наскрізного навчання (end-to-end learning). Нейромережа здатна самостійно, без втручання людини, формувати ієрархію візуальних ознак безпосередньо з «сирих» пікселів зображення.

Архітектура згорткової нейронної мережі натхненна біологічною будовою зорової кори головного мозку тварин і складається з кількох спеціалізованих типів шарів: згорткових, активаційних, шарів субдискретизації та повнозв'язних шарів.

Основним обчислювальним блоком є згортковий шар (Convolutional Layer). Його робота базується на математичній операції дискретної згортки. Замість того, щоб з'єднувати кожен нейрон з усіма пікселями вхідного зображення (як у класичних багатошарових персептронах, що призводить до катастрофічного зростання кількості параметрів), згортковий шар використовує набір невеликих двовимірних фільтрів (ядер згортки), які «ковзають» по

зображенню. Математично двовимірна згортка для вхідного тензора I та ядра K розміром m×n описується наступною формулою:

$$S(i, j) = (I * K)(i, j) = \Sigma \Sigma I(i + m, j + n) \cdot K(m, n) \ , \qquad (1.1)$$

де $S(i,j)$ - значення вихідної карти ознак у позиції з координатами (i,j); I - вхідний двовимірний тензор (зображення в градаціях сірого або окремий канал кольорового зображення); K - ядро згортки (фільтр) розміром $m×n$, вагові коефіцієнти якого підлягають оптимізації у процесі навчання; m, n - індекси перебору елементів ядра по висоті та ширині відповідно; символ $*$ позначає операцію двовимірної дискретної згортки. Геометрично формула описує процес послідовного зміщення ядра K по вхідному зображенню I з обчисленням у кожному положенні поелементного добутку відповідних значень із подальшим їх підсумовуванням.

Результатом цієї операції є формування карти ознак (Feature Map). Кожен фільтр під час процесу навчання адаптує свої вагові коефіцієнти так, щоб реагувати на специфічні візуальні патерни. На початкових шарах мережі фільтри розпізнають прості геометричні примітиви (краї, лінії, перепади контрасту). На глибших шарах ці примітиви комбінуються у складні високорівневі семантичні структури, такі як форма очей, вигин губ або наявність мімічних зморшок, що є критично важливими маркерами для ідентифікації емоції.

Для того щоб нейронна мережа могла апроксимувати складні нелінійні залежності між вхідним зображенням та емоційним станом, після кожної операції згортки застосовується нелінійна функція активації (Activation Function). Довгий час у нейромережах використовувалися функції сигмоїди або гіперболічного тангенса, проте вони страждали від проблеми згасаючого градієнта при навчанні глибоких мереж. У сучасних архітектурах CNN галузевим стандартом є використання функції ReLU (Rectified Linear Unit) та її модифікацій (наприклад, Leaky ReLU). Функція ReLU є обчислювально

ефективною, оскільки вона просто обнуляє всі від'ємні значення тензора, залишаючи додатні без змін:

$$f(x) = max(0, x) \quad (1.2)$$

де $f(x)$ - вихідне значення активаційної функції; x - вхідне значення (передактиваційний сигнал), яке відповідає результату згорткової операції перед застосуванням нелінійності; $max(0, x)$ - функція вибору максимального значення між нулем та вхідним сигналом. Таким чином, для всіх від'ємних значень x функція повертає нуль, а для додатних - залишає їх без змін, що забезпечує необхідну нелінійність та водночас обчислювальну простоту порівняно з сигмоїдою чи гіперболічним тангенсом.

Наступним важливим елементом архітектури є шар субдискретизації (Pooling Layer). Його основна функція - зменшення просторової розмірності (ширини та висоти) карт ознак при збереженні їхньої глибини. Це досягається шляхом агрегації інформації в невеликих локальних вікнах (зазвичай розміром 2×2 пікселі з кроком зміщення 2). Найчастіше застосовується операція Max Pooling, яка залишає лише максимальне значення пікселя у вікні. Використання субдискретизації дозволяє вирішити дві задачі: по-перше, експоненційно зменшити кількість параметрів та обчислювальних операцій у наступних шарах (що запобігає перенавчанню), а по-друге, забезпечити властивість локальної трансляційної інваріантності. Завдяки цьому алгоритм продовжуватиме коректно розпізнавати емоцію навіть у тому випадку, якщо обличчя на зображенні буде трохи зсунуте відносно центру.

Після проходження вхідного тензора через каскад згорткових шарів та шарів субдискретизації, отримана багатовимірна матриця ознак проходить через операцію «вирівнювання» (Flattening), перетворюючись на одновимірний вектор. Цей вектор подається на вхід традиційній штучній нейронній мережі - каскаду повнозв'язних шарів (Fully Connected Layers). Їхня задача полягає в

агрегації усіх знайдених високорівневих ознак та виконанні фінальної класифікації.

Останній повнозв'язний шар мережі містить кількість нейронів, що строго дорівнює кількості класів емоцій у задачі (наприклад, 7 нейронів для базової моделі Екмана). Для перетворення сирих вихідних значень мережі (логітів) у зрозумілий та нормований розподіл ймовірностей застосовується математична функція активації Softmax. Вона гарантує, що всі вихідні значення знаходитимуться у діапазоні від 0 до 1, а їхня загальна сума дорівнюватиме одиниці:

$$\sigma(z_i) = exp(z_i) / \Sigma\, exp(z_j) \qquad (1.3)$$

де z_i - вихідне значення (логіт) для і-го класу емоції, а K - загальна кількість розпізнаваних класів.

Процес навчання програмного модуля базується на ітеративній оптимізації вагових коефіцієнтів мережі. Метою навчання є мінімізація функції втрат (Loss Function), яка кількісно оцінює різницю між передбаченням алгоритму та істинною міткою класу в навчальному датасеті. Для задачі багатокласової класифікації стандартом є використання функції перехресної ентропії (Categorical Cross-Entropy):

$$L = -\Sigma\, y_i \cdot log(\hat{y}_i) \qquad (1.4)$$

де y_i - бінарний індикатор (0 або 1) того, чи дійсно зображення належить до класу і, а $\hat{y}_i$ - передбачена нейромережею ймовірність приналежності зображення до цього класу.

Мінімізація функції втрат здійснюється шляхом обчислення градієнтів для кожного вагового коефіцієнта за допомогою алгоритму зворотного поширення помилки (Backpropagation) на основі ланцюгового правила диференціювання.

Для безпосереднього оновлення ваг використовується метод градієнтного спуску. Сучасні програмні реалізації віддають перевагу адаптивним алгоритмам оптимізації, таким як Adam (Adaptive Moment Estimation). На відміну від класичного стохастичного градієнтного спуску (SGD), алгоритм Adam автоматично підбирає індивідуальну швидкість навчання (learning rate) для кожного параметра моделі, базуючись на експоненційно згладжених ковзних середніх градієнта та його квадрата. Це забезпечує набагато швидшу збіжність нейронної мережі та дозволяє уникати застрягання алгоритму в локальних мінімумах функції втрат під час навчання на складних мімічних датасетах.

1.3 Порівняльний аналіз технологій класифікації емоцій

Вибір оптимального технологічного стеку для реалізації програмного модуля класифікації людських емоцій є критичним етапом проєктування. Цей вибір безпосередньо визначає архітектуру всієї системи, вимоги до апаратного забезпечення, швидкість роботи (Inference Time) та кінцеву точність результатів. Для обґрунтування вибору необхідно провести комплексний порівняльний аналіз існуючих підходів за кількома ключовими критеріями: точність розпізнавання у неконтрольованих умовах (in-the-wild), здатність працювати в режимі реального часу (Real-Time capability), залежність від мережевої інфраструктури, обчислювальна складність та рівень захисту конфіденційних даних користувача.

Усі існуючі технологічні рішення в цій предметній області можна розділити на три великі категорії: алгоритми класичного машинного навчання з ручним виділенням ознак, хмарні API-сервіси на базі пропрієтарних алгоритмів та локальні моделі глибокого навчання (Edge AI).

Перша категорія - методи класичного машинного навчання (Classical ML). Типовий конвеєр таких систем включає локалізацію обличчя, пошук

антропометричних ключових точок, виділення геометричних чи текстурних ознак (наприклад, HOG або LBP) та фінальну класифікацію за допомогою SVM або дерева рішень. Головною перевагою класичного підходу є його надзвичайно низька обчислювальна складність. Такі алгоритми здатні обробляти понад 60 кадрів на секунду навіть на застарілих центральних процесорах. Однак головний недолік - катастрофічне падіння точності класифікації при відхиленні умов зйомки від ідеальних (нахил обличчя, погане освітлення), що робить цей підхід застарілим.

Друга категорія - хмарні сервіси розпізнавання емоцій. До цієї групи належать комерційні продукти, такі як Amazon Rekognition або Microsoft Azure Face API. Програмний модуль захоплює кадр з камери та відправляє його через інтернет на сервери провайдера. Перевагою цього підходу є доступ до надзвичайно потужних моделей із високою точністю розпізнавання. Проте цей підхід має низку критичних недоліків: повна залежність від якості інтернет-з'єднання, затримки мережі, що ускладнює роботу в реальному часі, висока вартість при масштабуванні проєкту, а також порушення приватності через передачу біометричних даних на сторонні сервери.

Третя категорія - використання архітектур глибокого навчання для локального виконання. Цей підхід передбачає розгортання навченої згорткової нейронної мережі безпосередньо на пристрої користувача. Важкі архітектури (VGG, ResNet) забезпечують точність хмарних сервісів, але вимагають наявності дискретної відеокарти. З іншого боку, полегшені архітектури (MobileNet) використовують технологію глибинної сепарабельної згортки, що зменшує обчислювальну складність моделі майже в 9 разів. Це дозволяє розпізнавати емоції з частотою 30 кадрів на секунду навіть на звичайному офісному ноутбуці при мінімальній втраті точності.

Для наочного зіставлення описаних технологічних підходів за ключовими критеріями розробки сформовано порівняльну таблицю 1.1.

Таблиця 1.1 - Порівняльна характеристика технологій класифікації емоцій

Критерій порівняння	Класичне машинне навчання	Хмарні API-сервіси	Локальні CNN
Точність в ідеальних умовах	Середня (70–80 %)	Дуже висока (95 %+)	Висока (90–95 %)
Точність у in-the-wild	Низька	Дуже висока	Висока
Real-Time продуктивність	Дуже висока (60+ FPS)	Низька (мережеві затримки)	Висока (25–30 FPS на CPU)
Вимоги до обладнання	Мінімальні	Мінімальні (на клієнті)	Середні / Високі
Залежність від мережі	Автономна робота	Критична залежність	Автономна робота
Конфіденційність даних	Висока	Низька (передача біометрії)	Висока (локальна обробка)

Підсумовуючи результати порівняльного аналізу, наведені у таблиці 1.1, можна дійти висновку, що для розробки сучасного програмного модуля класифікації людських емоцій найбільш раціональним рішенням є використання локальних згорткових нейронних мереж оптимізованої архітектури (наприклад, сімейства MobileNet чи Mini-Xception). Такий вибір забезпечує оптимальний компроміс між основними технічними вимогами. З одного боку, система не потребує інтернет-з'єднання і гарантує абсолютну конфіденційність біометричних даних користувача. З іншого боку, завдяки алгоритмічній оптимізації, програмний модуль здатен працювати в режимі реального часу, забезпечуючи високу точність класифікації базових емоцій навіть за умов часткового перекриття обличчя, що є недосяжним для застарілих алгоритмів класичного машинного навчання.

1.4 Висновок та постановка задачі

У першому розділі проведено аналіз предметної області автоматизованого розпізнавання емоцій, описано теоретичні моделі репрезентації емоцій та основні проблеми класифікації у неконтрольованих умовах. Детально розглянуто математичний апарат згорткових нейронних мереж та виконано порівняльний аналіз трьох категорій технологій. Обґрунтовано вибір локальної CNN оптимізованої архітектури як найбільш раціонального рішення для подальшої розробки.

Виходячи з результатів проведеного аналізу та з огляду на сформульовану мету, у кваліфікаційній роботі поставлено такі задачі:

1) провести аналіз предметної області автоматизованого розпізнавання емоцій, теоретичних моделей репрезентації емоцій та проблем класифікації у неконтрольованих умовах;

2) зробити порівняльний аналіз існуючих технологій класифікації емоцій - класичного машинного навчання, хмарних API-сервісів та локальних нейронних мереж;

3) розробити узагальнений алгоритм роботи програмного модуля та проаналізувати існуючі архітектури згорткових нейронних мереж;

4) обґрунтувати вибір датасету FER2013 як основного навчального набору;

5) розробити модифіковану архітектуру згорткової нейронної мережі на основі Mini-Xception з інтегрованим SE-блоком каналової уваги;

6) реалізувати програмний модуль мовою Python та провести експериментальне дослідження за методологією ablation study;

7) здійснити техніко-економічне обґрунтування розробки.

2 АЛГОРИТМИ ТА НАБОРИ ДАНИХ ДЛЯ КЛАСИФІКАЦІЇ ЕМОЦІЙ

2.1 Узагальнений алгоритм класифікації емоційних станів

Перш ніж розглядати окремі архітектури нейронних мереж, доцільно описати загальну структуру обчислювального процесу, в межах якого ці мережі застосовуються. Незалежно від конкретної реалізації - будь то легка модель типу Mini-Xception чи важка ResNet-50 - модуль класифікації емоцій працює як послідовність взаємопов’язаних блоків обробки даних, де результат роботи одного блоку слугує входом для наступного. Така схема є усталеною в задачах комп’ютерного зору і фактично повторює класичний підхід, що сформувався ще у роботах Віоли та Джонса на початку 2000-х років.

У межах даної роботи прийнято поділ цього процесу на п’ять етапів: захоплення зображення, локалізація обличчя на кадрі, попередня обробка виділеної області, виділення ознак згортковою мережею та власне класифікація отриманого вектора ознак на одну з категорій емоцій. Такий поділ є умовним і в окремих реалізаціях деякі етапи можуть бути об’єднані (наприклад, у моделях типу end-to-end детекція обличчя інколи виконується безпосередньо тією ж нейронною мережею). Однак з огляду на навчально-методичну ясність та можливість окремо тестувати кожен компонент розглянемо ці п’ять етапів послідовно.

Захоплення зображення є початковою ланкою всього процесу. На цьому етапі модуль отримує сирий візуальний матеріал від джерела вхідних даних. Джерелом може бути файл (звичайне фото з носія, фрагмент відеозапису), потік з веб-камери у режимі реального часу або кадр, переданий зовнішнім програмним середовищем. У випадку відеопотоку алгоритм оперує не одним статичним кадром, а послідовністю кадрів, які надходять із певною частотою. На практиці частоти 10–15 кадрів за секунду цілком достатньо для більшості застосувань, оскільки мімічні зміни на обличчі людини відбуваються не настільки швидко, щоб вимагати високої частоти дискретизації. Щодо роздільної

здатності, то експериментально встановлено, що для коректного виявлення обличчя достатньо стандартного формату VGA (640×480), хоча в умовах поганого освітлення або зйомки з великої відстані бажано підвищувати її до 1280×720. Слід пам’ятати, що зайве збільшення роздільної здатності лише сповільнює подальші етапи, не покращуючи точності, оскільки на кінцевому етапі зображення все одно буде стиснуте до 48×48 або 224×224 пікселів.

Локалізація обличчя на кадрі - це задача знайти, де саме на зображенні знаходиться обличчя, і виділити його у вигляді прямокутної області. Без цього кроку нейронна мережа намагалася б класифікувати всю сцену цілком, включно з фоном, одягом, предметами інтер’єру, що внесло б суттєвий шум у результат. Найвідомішим алгоритмом тут історично є каскадний детектор Віоли-Джонса, який реалізовано в OpenCV у вигляді готового каскаду haarcascade_frontalface_default.xml. Цей метод базується на простих ознаках Хаара і працює дуже швидко, проте має чітко окреслені обмеження: при відхиленні голови вбік на 30° і більше або при поганому освітленні точність помітно падає. Серед сучасніших підходів варто згадати детектор з бібліотеки dlib (на основі гістограм спрямованих градієнтів) та нейромережевий метод MTCNN, який не лише виділяє рамку обличчя, а й одночасно повертає координати п’яти ключових точок (центри очей, кінчик носа, кутики рота). Останній факт є корисним, оскільки знання положення очей дає змогу одразу виконати геометричне вирівнювання обличчя - повернути зображення так, щоб лінія очей була горизонтальною. У межах розробленого модуля використовується каскад Хаара, оскільки він не потребує GPU і дає прийнятну швидкодію на середньому ноутбуці.

Попередня обробка є тим етапом, який часто недооцінюють початківці, хоча саме від його коректності залежить більша частина потенційних помилок класифікації. Виявлена область обличчя вирізається з оригінального зображення за координатами рамки, після чого до неї послідовно застосовується кілька перетворень. Спершу зображення масштабується до фіксованого розміру, що відповідає вхідному шару конкретної моделі - у випадку Mini-Xception це 48×48,

а для більших мереж типу ResNet-50 або VGG-16 - 224×224 пікселі. Далі здійснюється перехід до градацій сірого: для розпізнавання мімічних виразів колір не несе додаткової інформації, а от обчислювальне навантаження одноканальне зображення зменшує приблизно втричі. Нарешті, значення яскравості пікселів нормалізуються до діапазону [0; 1] шляхом ділення на 255, або, в окремих випадках, до діапазону із середнім нуль і одиничним стандартним відхиленням (z-нормалізація). Така нормалізація розв'язує одразу дві задачі - стабілізує процес навчання мережі (градієнти не стрибають у широкому діапазоні значень) і робить класифікатор менш чутливим до загального рівня освітленості сцени, що особливо важливо при роботі з відеопотоком, де освітлення може змінюватися від кадру до кадру.

Виділення ознак - центральна ланка всього процесу, і це той етап, на якому згорткові мережі радикально відрізняються від класичних методів машинного навчання. Якщо у класичному підході (HOG + SVM або LBP + Random Forest) інженер мав сам обирати, які саме характеристики обчислювати на зображенні, то в CNN цей процес повністю автоматизований. Згорткові шари нейронної мережі під час навчання самостійно формують ієрархію ознак різного рівня абстракції. На перших шарах мережа «вчиться» реагувати на найпростіші візуальні стимули - прямі лінії, дуги, перепади яскравості. На середніх шарах ці примітиви комбінуються у більш складні структури - контури очей, форму брів, лінію губ. А на найглибших шарах формуються по-справжньому семантичні представлення на кшталт «піднята зовнішня частина брів - типовий маркер здивування» або «опущені кутики рота разом зі складками між бровами - сум». Результатом цього етапу є не зрозумілий людині вектор, а ущільнене числове представлення обличчя, яке містить інформацію саме про мімічні ознаки і вже «очищене» від несуттєвих деталей на кшталт кольору шкіри або форми зачіски.

Завершальний етап - це власне класифікація вектора ознак. Після того, як зображення пройшло через каскад згорткових шарів, отримана багатовимірна карта ознак перетворюється в одновимірний вектор операцією Flatten або глобального усереднення (Global Average Pooling). Цей вектор подається на

повнозв’язний шар, останнім кроком якого є застосування функції активації Softmax. На виході отримуємо вектор з семи чисел (для базової моделі Екмана), сума яких дорівнює одиниці, а кожне окреме значення можна інтерпретувати як ймовірність того, що зображене обличчя належить відповідній емоції. Як приклад, вектор [0,02; 0,01; 0,05; 0,85; 0,03; 0,02; 0,02] вказує на те, що з імовірністю 85 % зображене обличчя виражає радість. Клас із найбільшим значенням оголошується фінальним результатом, але саму ймовірнісну природу виходу варто зберігати: вона дає змогу запровадити поріг впевненості (наприклад, відкидати передбачення з імовірністю нижче 50 %) або повідомляти користувача про неоднозначність кадру, що корисно у критичних застосуваннях.

У режимі відеопотоку до описаного п’ятиетапного конвеєра доцільно додавати додатковий механізм темпорального згладжування. Справа в тому, що навіть досить точна модель неминуче дає поодинокі помилкові класифікації на окремих кадрах - через шум камери, мімічні артефакти, моргання тощо. Якщо виводити результат для кожного кадру окремо, на екрані спостерігатиметься хаотична зміна підписів: «радість» - «здивування» - «радість» - «нейтральний» протягом однієї секунди, що виглядає неприродно і знижує довіру користувача. Найпростіший спосіб згладжування - усереднення ймовірнісних векторів за ковзним вікном з 5–10 останніх кадрів і вже на основі усередненого вектора визначати фінальний клас. Це додає певної затримки реакції (приблизно на півсекунди), але водночас усуває стрибки і робить роботу модуля візуально стабільною.

Таким чином, узагальнений алгоритм роботи модуля являє собою чітко структуровану послідовність етапів, кожен з яких розв’язує власне підзадання. Така модульність зручна не лише з інженерної точки зору (можна окремо профілювати швидкодію кожного блоку, локалізувати помилки, замінювати компоненти на більш ефективні), а й з дослідницької: під час експериментального дослідження, описаного в третьому розділі, саме можливість «замінювати» окремі блоки дала змогу провести коректну ablation-розвідку та кількісно оцінити внесок кожної модифікації архітектури.

2.2 Види алгоритмів класифікації емоцій

Для вирішення задачі класифікації емоційних станів за візуальними образами обличчя розроблено значну кількість алгоритмічних підходів, що відрізняються за рівнем складності, обчислювальними вимогами та досяжною точністю. Основні алгоритми можна класифікувати за кількома категоріями: за типом екстракції ознак (ручна або автоматична), за архітектурою моделі (лінійна, деревоподібна, нейромережева) та за парадигмою навчання (навчання з нуля або трансферне навчання).

Оскільки в межах даної роботи основним інструментом розв’язання задачі обрано саме згорткові нейронні мережі, перед детальним порівнянням окремих архітектур (VGG, ResNet, MobileNet тощо) доцільно зупинитися на типовій будові CNN, призначеної для класифікації емоцій за виразом обличчя. Розуміння внутрішньої структури таких мереж є ключем до подальшого обґрунтованого вибору архітектури під конкретні обмеження проєкту. Узагальнену схему типової CNN-мережі для задачі розпізнавання семи базових емоцій наведено на рисунку 2.1.

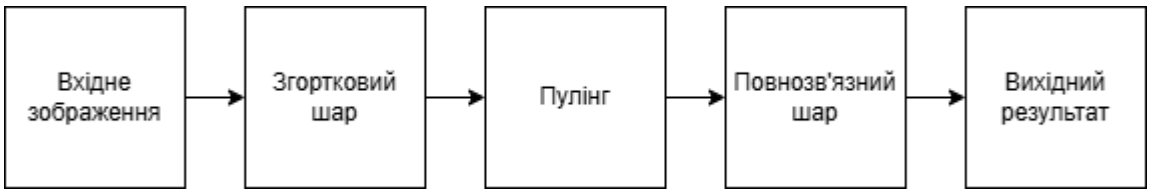

Рисунок 2.1 – Узагальнена архітектура згорткової нейронної мережі для класифікації емоцій за виразом обличчя

Як видно з рисунка, мережу умовно поділяють на дві функціонально різні частини: блок екстракції ознак (feature extractor) та класифікатор (classifier). Перша частина - це послідовність згорткових блоків, які з вхідного зображення розміром 48×48 пікселів формують компактне векторне представлення з 128

чисел, що описує мімічні особливості обличчя. Друга частина - повнозв'язний шар із семи нейронів та активацією Softmax, що перетворює це представлення у розподіл імовірностей по семи класах емоцій.

Зупинимося детальніше на функціональному призначенні кожного типу шарів. Згортковий шар (Conv2D) виконує операцію двовимірної дискретної згортки вхідного тензора з набором фільтрів, що навчаються. У наведеній схемі перший згортковий блок містить 32 фільтри розміром 3×3, другий - 64, третій - 128 фільтрів того ж розміру. Така зростаюча кількість фільтрів є типовою для CNN: на ранніх шарах достатньо невеликої кількості детекторів для виявлення низькорівневих структур (краї, штрихи, кутові паттерни), тоді як на глибших шарах необхідна більша кількість фільтрів для представлення значно різноманітніших високорівневих мімічних паттернів (форми очей, положення брів, конфігурації рота).

Шар пакетної нормалізації (Batch Normalization, BN), розміщений після кожного згорткового, нормалізує активації в межах поточного міні-батчу до нульового середнього та одиничного стандартного відхилення з подальшим лінійним перетворенням з навченими параметрами. Введення BN-шарів, запропоноване С. Іоффе та К. Сегеді у 2015 році, дозволяє суттєво прискорити навчання, зменшує чутливість до початкової ініціалізації ваг і виконує неявну регуляризацію. У задачах розпізнавання емоцій, де навчальні вибірки відносно невеликі, наявність BN дозволяє в середньому на 2–3 епохи раніше виходити на плато точності та помітно знижує ризик розходження градієнтів.

Функція активації ReLU, що застосовується після нормалізації, забезпечує нелінійність мережі. Без неї послідовність згорткових і повнозв'язних шарів була б математично еквівалентною одному лінійному перетворенню, тобто мережа втратила б здатність моделювати складні нелінійні залежності між пікселями зображення та категорією емоції. Особливістю ReLU є те, що похідна функції є константною на додатній півосі, завдяки чому проблема згасаючого градієнта виражена значно слабше порівняно з сигмоїдою чи гіперболічним тангенсом, які тривалий час домінували в нейромережах першого покоління.

Шар субдискретизації (MaxPooling) у наведеній схемі застосовується двічі, із параметрами вікна 2×2 і кроком 2, що зменшує просторову розмірність карт ознак удвічі за кожним проходженням (48×48 → 24×24 → 12×12). Окрім очевидного скорочення обчислювальних витрат, пулінг забезпечує властивість локальної трансляційної інваріантності - якщо на наступному кадрі обличчя зміститься на кілька пікселів, мережа все одно дасть стабільне передбачення.

На переході від згорткової частини до класифікатора замість традиційного Flatten доцільно використовувати шар глобального усереднення (Global Average Pooling, GAP). На відміну від Flatten, який розгортає весь тензор 12×12×128 у вектор з 18 432 елементів і вимагає у наступному повнозв'язному шарі мільйонів параметрів, GAP усереднює кожну карту ознак до одного числа, формуючи вектор довжиною 128. Таке скорочення кількості параметрів напряму впливає на здатність моделі узагальнювати та значно зменшує ризик перенавчання, що особливо актуально для відносно невеликого датасету FER2013.

Шар Dropout із ймовірністю p=0,5, розміщений перед фінальним повнозв'язним шаром, реалізує стохастичну регуляризацію за методом Срівастави та ін. (2014). Під час навчання він випадково обнуляє половину виходів попереднього шару, через що мережа змушена розподіляти «відповідальність» за прийняття рішення між багатьма ознаками, а не покладатися на окремі «улюблені» нейрони. На інференсі (тобто під час реального застосування навченої моделі) Dropout не діє, і всі активації використовуються повністю. Емпірично саме застосування Dropout з імовірністю від 0,3 до 0,5 на задачі FER2013 додає 1–2 відсоткові пункти точності та помітно зменшує розрив між навчальною й валідаційною кривими.

Фінальний повнозв'язний шар (Dense) із сімома нейронами та функцією активації Softmax перетворює внутрішнє представлення на вектор з семи невід'ємних чисел, сума яких дорівнює одиниці. Кожне число можна інтерпретувати як оцінку ймовірності того, що зображене обличчя виражає відповідну емоцію. Прогнозованим класом оголошується той, який має максимальне значення (операція argmax), однак саме збереження повного

ймовірнісного розподілу є важливим: воно дає змогу будувати рейтинги «другої та третьої найімовірнішої емоції», аналізувати впевненість моделі та запроваджувати порогове відкидання сумнівних передбачень у практичних застосуваннях.

Наведена архітектура є базовою референсною моделлю, від якої відштовхуються при проєктуванні більш складних рішень. Як буде показано далі, сучасні архітектури (ResNet, MobileNet, Mini-Xception тощо) розширюють або модифікують цю базову схему за рахунок залишкових з'єднань, факторизації згорток або інтеграції механізмів каналової уваги, проте загальна логіка «згорткова частина + класифікатор» залишається незмінною. Виходячи з цього розуміння, надалі розглянемо основні категорії алгоритмів класифікації емоцій.

Першу категорію складають класичні алгоритми машинного навчання з ручним виділенням ознак. До них належить метод опорних векторів (Support Vector Machine, SVM), який будує оптимальну розділяючу гіперплощину у багатовимірному просторі ознак. SVM демонструє задовільну точність (70–80 %) при використанні на попередньо обчислених дескрипторах HOG або LBP, проте повністю залежить від якості ручного проєктування вхідних ознак. Іншим класичним підходом є алгоритм випадкового лісу (Random Forest), що формує ансамбль дерев рішень та використовує мажоритарне голосування для фінальної класифікації. Ці методи зберігають актуальність для вбудованих систем з обмеженими ресурсами, де розгортання повноцінної нейронної мережі є технічно неможливим.

Другу категорію складають згорткові нейронні мережі стандартних архітектур. Серед найбільш впливових архітектур виділяються VGGNet, ResNet та Inception. Архітектура VGG-16, розроблена дослідниками Оксфордського університету, складається з 16 вагових шарів та використовує виключно малі згорткові фільтри розміром 3×3. Ця архітектура є концептуально простою, але містить понад 138 мільйонів параметрів, що робить її надзвичайно ресурсомісткою. На датасеті FER2013 моделі на базі VGG досягають точності 70–73 %, що є помірним результатом.

Архітектура ResNet, запропонована дослідниками Microsoft Research, вирішила фундаментальну проблему деградації точності при збільшенні глибини мережі. Ключовою інновацією є механізм залишкових з'єднань (Skip Connections або Shortcut Connections), які дозволяють градієнту «обходити» проблемні шари під час зворотного поширення помилки. Це уможливило навчання мереж глибиною понад 100 шарів. Варіант ResNet-50 (50 шарів) є одним із найчастіше використовуваних базових моделей для трансферного навчання у задачах розпізнавання емоцій, досягаючи точності 73–76 % на FER2013.

Архітектура Inceptio, розроблена Google, використовує паралельні гілки згорткових фільтрів різного розміру (1×1, 3×3, 5×5) всередині одного модуля. Це дозволяє мережі одночасно аналізувати візуальні паттерни різного масштабу, що є корисним для розпізнавання як дрібних деталей (наморщування), так і загальних контурів (форма обличчя). Модифікація Inception-v3 містить близько 23 мільйонів параметрів, що значно менше за VGG-16, при порівнянній або вищій точності.

Третю категорію складають оптимізовані мобільні архітектури. Архітектура MobileNet, розроблена Google, є ключовим рішенням для розгортання моделей класифікації емоцій на пристроях з обмеженими ресурсами. Основною інновацією MobileNet є використання глибинно-сепарабельної згортки (Depthwise Separable Convolution), яка розкладає стандартну операцію згортки на дві послідовні: глибинну згортку (Depthwise Convolution), що застосовує один фільтр до кожного вхідного каналу окремо, та точкову згортку (Pointwise Convolution), що використовує фільтр розміром 1×1 для лінійної комбінації результатів. Таке розкладання зменшує кількість обчислювальних операцій у 8–9 разів порівняно зі стандартною згорткою, при втраті точності лише на 1–2 відсоткових пункти.

MobileNet-v2 додатково вводить інвертовані залишкові блоки з лінійним вузьким горлом с. На відміну від класичних залишкових з'єднань ResNet, де широкий шар стискається до вузького, MobileNet-v2 спочатку розширює тензор через точкову згортку, потім застосовує глибинну згортку та знову стискає

результат. Така архітектура дозволяє ефективно використовувати обмежену пам'ять мобільних пристроїв. Версія MobileNet-v2 містить лише 3,4 мільйони параметрів та при оптимізації досягає точності 65–70 % на FER2013, що є прийнятним результатом для мобільних та вбудованих застосунків.

Ще однією легкою архітектурою є EfficientNet, яка використовує метод складеного масштабування для одночасного збалансованого збільшення глибини мережі, ширини шарів та роздільної здатності вхідного зображення. Базова модель EfficientNet-B0 має лише 5,3 мільйони параметрів, але досягає точності, порівнянної з набагато більшими моделями. Більші варіанти (B3, B4) демонструють точність 74–76 % на FER2013, що ставить їх на рівень з ResNet-50 при значно меншому обсязі обчислень.

Четверту категорію складають підходи на основі трансферного навчання (Transfer Learning). Навчання згорткової нейронної мережі з нуля вимагає великих обсягів розмічених даних та значних обчислювальних ресурсів. Трансферне навчання дозволяє використати модель, попередньо навчену на масштабному загальному датасеті (наприклад, ImageNet з 14 мільйонами зображень), та адаптувати її до специфічної задачі розпізнавання емоцій. Практично це означає, що згорткові шари моделі, які вже навчились ефективно виділяти візуальні ознаки, залишаються «замороженими» (їхні ваги не оновлюються), а перенавчаються лише останні повнозв'язні шари класифікатора. Цей підхід дозволяє досягти високої точності навіть при обмеженому обсязі навчальних даних з емоціями та значно прискорює процес навчання.

Окремо слід відзначити підхід ансамблевого навчання (Ensemble Learning), при якому кілька різних моделей (наприклад, ResNet, VGG та Inception) навчаються паралельно, а їхні прогнози агрегуються через механізм зваженого голосування або усереднення ймовірностей. Ансамблеві моделі стабільно демонструють найвищу точність на конкурсних тестових наборах, оскільки помилки однієї моделі компенсуються правильними прогнозами інших. Проте обчислювальна вартість такого підходу зростає пропорційно кількості моделей в ансамблі, що обмежує його застосування у системах реального часу.

2.3 Датасети для навчання моделей розпізнавання емоцій

Якість та репрезентативність навчального набору даних (датасету) є, напевно, найбільш критичним фактором, що визначає кінцеву ефективність будь-якої моделі машинного навчання для класифікації емоцій. Навіть найдосконаліша архітектура нейронної мережі продемонструє незадовільні результати, якщо її навчати на нерепрезентативних або неправильно розмічених даних. Тому розуміння структури, переваг та обмежень існуючих датасетів є обов'язковою складовою процесу розробки програмного модуля.

Датасет для класифікації емоцій являє собою колекцію зображень облич, кожне з яких асоційоване з міткою (label), що вказує на емоційний стан зображеної людини. Мітки можуть бути дискретними (одна з семи базових емоцій) або безперервними (значення валентності та збудження). Датасети розрізняються за кількома фундаментальними параметрами: загальний обсяг зображень, кількість унікальних суб'єктів, метод збору (лабораторний або «в дикій природі»), спосіб анотації (експертна або краудсорсингова) та розподіл зображень за класами емоцій.

Одним із найбільш широко використовуваних датасетів у академічному середовищі є FER2013. Цей набір даних був представлений на конкурсі Kaggle та містить приблизно 35 887 зображень облич у градаціях сірого з роздільною здатністю 48×48 пікселів. Зображення розподілені на сім класів: гнів (anger), огида, страх, радість, сум, здивування та нейтральний стан. FER2013 був зібраний автоматично з інтернету за допомогою пошукового API Google, що означає наявність значного відсотка «шумних» зображень з некоректними мітками (оцінка рівня шуму становить 10–15 %). Незважаючи на це, FER2013 залишається стандартним бенчмарком для порівняння алгоритмів завдяки своєму великому обсягу та доступності. Приклади зображень з кожного класу емоцій датасету FER-2013 наведено на рисунку 2.2.

Рисунок 2.2 – Приклади зображень з датасету FER-2013 для семи базових класів емоцій

Датасет AffectNet є одним із найбільших наборів даних для розпізнавання емоцій та містить понад 1 мільйон зображень облич, зібраних з інтернету. З них приблизно 440 000 зображень мають ручну анотацію від експертів. AffectNet підтримує як дискретну класифікацію за вісьмома категоріями (сім базових емоцій плюс «презирство»), так і безперервну модель з вимірами валентності та збудження. Значна перевага AffectNet полягає у різноманітності умов зйомки: зображення містять варіації освітлення, ракурсу, етнічної приналежності та вікових груп, що наближає датасет до реальних умов використання.

Датасет CK+ (Extended Cohn-Kanade) є класичним лабораторним набором даних, що містить 593 відеопослідовності від 123 суб'єктів. Відеопослідовності починаються з нейтрального виразу обличчя та закінчуються піковим вираженням цільової емоції. Анотація виконана експертами з використанням системи FACS. Головна перевага CK+ полягає у високій якості розмітки та контрольованих умовах зйомки. Однак обмежена кількість суб'єктів та переважно лабораторний характер даних знижують здатність моделей, навчених виключно на CK+, до узагальнення в реальних умовах.

Серед інших важливих датасетів слід відзначити RAF-DB (Real-world Affective Faces Database), який містить приблизно 30 000 зображень, зібраних із соціальних мереж та анотованих за допомогою краудсорсингу з подальшою верифікацією експертами. RAF-DB характеризується високою варіативністю умов зйомки, що робить його цінним для навчання моделей, призначених для реального використання. Також варто згадати датасет JAFFE (Japanese Female

Facial Expression), який хоча й містить лише 213 зображень 10 японських жінок, є важливим для дослідження крос-культурних аспектів розпізнавання емоцій.

Критичною проблемою більшості існуючих датасетів є дисбаланс класів. У реальному житті частота прояву різних емоцій суттєво відрізняється: людина перебуває у нейтральному стані або стані радості значно частіше, ніж у стані огиди чи страху. Це відображається у розподілі навчальних даних. Наприклад, у FER2013 клас «радість» містить понад 7 200 зображень, тоді як клас «огида» - лише близько 440. Такий дисбаланс призводить до того, що нейронна мережа «перенавчається» на домінуючих класах та систематично ігнорує недопредставлені.

Для подолання проблеми дисбалансу класів застосовуються кілька стратегій. По-перше, метод перевибірки, при якому зображення з недопредставлених класів дублюються або генеруються штучно за допомогою аугментації даних (Data Augmentation). Аугментація включає такі трансформації, як горизонтальне дзеркальне відображення, випадковий поворот на невеликий кут (до 15 градусів), випадкова обрізка та зміна яскравості і контрасту. По-друге, використання зваженої функції втрат (Weighted Loss), при якій помилка класифікації зображень із рідкісних класів штрафується з більшим ваговим коефіцієнтом. По-третє, метод SMOTE (Synthetic Minority Over-sampling Technique), який генерує синтетичні зразки шляхом інтерполяції між існуючими зразками в просторі ознак.

Ще одним важливим аспектом є якість анотації (Label Quality). Емоційний стан людини є суб'єктивним явищем, і навіть кваліфіковані експерти можуть розходитися в оцінці конкретного зображення. Дослідження показують, що міжекспертна узгодженість для задачі класифікації емоцій за виразом обличчя складає лише 65–70 %, що фактично встановлює теоретичну верхню межу точності для алгоритмів автоматичної класифікації. Це означає, що модель, яка демонструє точність 70–75 % на датасеті з неідеальною розміткою, може фактично працювати на рівні, близькому до людської здатності розпізнавання.

Для практичної розробки програмного модуля класифікації емоцій, описаного в даній роботі, найбільш доцільним є використання датасету FER2013 як основного навчального набору з подальшим донавчанням на датасеті RAF-DB або AffectNet для підвищення стійкості до реальних умов. Вибір FER2013 як базового набору обґрунтовується його широкою доступністю (безкоштовне завантаження з платформи Kaggle), достатнім обсягом для навчання CNN (понад 35 000 зображень), наявністю усіх семи базових класів емоцій та існуванням великої кількості публікацій з результатами бенчмаркінгу, що дозволяє об'єктивно оцінити якість розробленого алгоритму.

У другому розділі описано узагальнений алгоритм роботи модуля класифікації емоцій, який складається з п'яти послідовних етапів. Виконано класифікацію існуючих видів алгоритмів з виокремленням чотирьох категорій - від класичних SVM/RF до сучасних легких архітектур типу MobileNet та EfficientNet. Проаналізовано основні навчальні датасети та обґрунтовано вибір FER2013 як базового набору для подальшої практичної реалізації.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ МОДУЛЯ КЛАСИФІКАЦІЇ ЕМОЦІЙ

3.1 Архітектура програмного модуля та обґрунтування технологічного стеку

Програмна реалізація модуля класифікації емоцій побудована за принципом модульної архітектури, у якій кожен функціональний компонент відповідає за окремий етап обробки візуальної інформації - від отримання вхідного зображення до формування результату класифікації. Такий підхід забезпечує можливість незалежного тестування окремих підсистем, спрощує супровід коду та дозволяє замінювати окремі компоненти без переробки решти системи.

Структура програмного модуля складається з п'яти основних компонентів: модуля захоплення зображення, модуля детекції обличчя, модуля попередньої обробки, ядра класифікації на базі згорткової нейронної мережі та модуля візуалізації результату. Логічну схему взаємодії компонентів наведено на рисунку 3.1, при цьому стрілки відображають напрям передачі даних між модулями.

Рисунок 3.1 – Структурна схема програмного модуля класифікації емоцій

Модуль захоплення зображення приймає на вхід дані з різних джерел - статичних файлів, потокової камери або готових масивів пікселів - та передає їх у внутрішньому форматі бібліотеки OpenCV. На цьому ж етапі виконується

перевірка коректності формату та повідомлення про помилку у разі неможливості зчитування файлу.

Модуль детекції обличчя локалізує область обличчя на вхідному зображенні та повертає координати обмежувального прямокутника. Для цієї задачі обрано каскадний класифікатор Хаара з бібліотеки OpenCV, який забезпечує прийнятну швидкодію на CPU без необхідності наявності GPU. У разі невдалої детекції модуль повертає відповідне повідомлення та припиняє подальшу обробку.

Модуль попередньої обробки виконує серію перетворень над виявленою областю обличчя: вирізання, перетворення у градації сірого, масштабування до розміру 48×48 пікселів та нормалізацію значень яскравості до діапазону [0, 1]. Перетворення в градації сірого зменшує обчислювальне навантаження приблизно в три рази, не призводячи до суттєвої втрати інформації для задачі класифікації мімічних виразів. Нормалізація стабілізує процес інференсу та робить роботу мережі менш чутливою до варіацій освітлення.

Ядро класифікації - це навчена згорткова нейронна мережа, побудована на основі модифікованої архітектури Mini-Xception. Вхідним тензором для мережі є нормалізоване зображення розмірності 48×48×1, вихідним - вектор з семи значень ймовірностей, що відповідають базовим емоціям. Більш детальний опис архітектури наведено в підрозділі 3.2.

Модуль візуалізації приймає результат класифікації та формує читабельний для користувача вивід - назву передбаченої емоції, значення впевненості моделі у відсотках, а у випадку відеопотоку - анотоване зображення з накладеним обмежувальним прямокутником і підписом.

Вибір технологічного стеку для реалізації модуля зумовлений вимогами до продуктивності, доступності інструментів та можливостей розгортання. У таблиці 3.1 наведено зведений перелік технологій із зазначенням ролі кожної з них у проєкті.

Таблиця 3.1 - Технологічний стек програмного модуля

Технологія	Версія	Призначення у проєкті
Python	3.10+	Основна мова реалізації модуля та допоміжних скриптів
TensorFlow / Keras	2.15+	Реалізація, навчання та збереження моделі CNN
OpenCV	4.x	Зчитування зображень, детекція обличчя, перетворення кольорового простору
NumPy	1.26+	Робота з багатовимірними масивами, передобробка даних
Matplotlib / Seaborn	3.8+ / 0.13+	Візуалізація графіків навчання та confusion matrix
scikit-learn	1.4+	Розрахунок метрик якості класифікатора (precision, recall, F1)
Kaggle Notebooks	-	Хмарне середовище для навчання моделі з безкоштовним GPU NVIDIA Tesla P100

Вибір TensorFlow з високорівневим інтерфейсом Keras обґрунтований кількома міркуваннями. По-перше, ця бібліотека є галузевим стандартом для розробки нейромережевих застосунків і має повну документацію з прикладами для задач класифікації зображень. По-друге, вбудована підтримка автоматичного визначення доступного апаратного прискорювача (CPU, GPU, TPU) дозволяє запускати один і той самий код у різних обчислювальних середовищах. По-третє, формат збереження моделі .keras є самодостатнім - він містить як архітектуру, так і ваги, що спрощує розгортання.

Альтернативою TensorFlow міг би бути PyTorch, який має дещо нижчий поріг входження для дослідницьких задач. Однак для виробничого розгортання моделі TensorFlow має суттєву перевагу - наявність зрілих інструментів для конвертації у формати TensorFlow Lite (для мобільних пристроїв) та TensorFlow.js (для виконання у браузері). Це робить розроблений модуль придатним для подальшої інтеграції в різноманітні цільові середовища.

Бібліотека OpenCV обрана для роботи з зображеннями, оскільки вона є де-факто стандартом у комп'ютерному зорі та забезпечує оптимізовані реалізації базових операцій - детекції облич, перетворення кольорового простору,

морфологічних трансформацій. Каскадний класифікатор Хаара з готового файлу haarcascade_frontalface_default.xml забезпечує детекцію фронтальних облич зі швидкістю до 60 кадрів на секунду на типовому процесорі Intel Core i5.

Вибір хмарного середовища Kaggle Notebooks для процесу навчання моделі обумовлений двома практичними міркуваннями. По-перше, Kaggle надає безкоштовний доступ до GPU NVIDIA Tesla P100 у межах 30 годин на тиждень, чого з запасом вистачає для навчання моделі обраного класу. По-друге, набір даних FER2013 уже доступний у каталозі Kaggle Datasets, що знімає необхідність окремого завантаження та обробки архіву об'ємом близько 60 МБ. Усі експерименти, описані у підрозділі 3.3, виконано саме в цьому середовищі.

Системні вимоги для роботи готового модуля у режимі інференсу є доволі помірними: процесор з частотою від 2,0 ГГц, оперативна пам'ять 4 ГБ та операційна система Windows 10/11 або Linux Ubuntu 22.04 і новіша. Наявність GPU не є обов'язковою - модель розрахована на роботу на CPU зі швидкістю 5–8 кадрів за секунду, чого достатньо для більшості прикладних сценаріїв класифікації статичних зображень.

3.2 Розроблена модифікована архітектура згорткової нейронної мережі

За основу архітектури ядра класифікації обрано модель Mini-Xception, запропоновану О. Аріагою у роботі «Real-time Convolutional Neural Networks for Emotion and Gender Classification» (2017). Ця архітектура є оптимізованою версією мережі Xception (F. Chollet, 2017) і була розроблена спеціально для задачі розпізнавання емоцій у режимі реального часу на пристроях з обмеженими обчислювальними ресурсами.

Mini-Xception базується на двох ключових архітектурних ідеях. Перша - використання глибинно-сепарабельної згортки (depthwise separable convolution), яка декомпозує стандартну операцію згортки на дві послідовні: глибинну згортку, що застосовує один фільтр до кожного вхідного каналу окремо, та точкову згортку розміром 1×1, що виконує лінійне комбінування результатів. Така декомпозиція зменшує обчислювальну складність у 8–9 разів при втраті точності лише на 1–2 відсоткових пункти. Друга ідея - використання залишкових з'єднань (residual connections) між блоками, що дозволяє градієнту вільно поширюватися крізь глибокі шари мережі та запобігає проблемі згасаючого градієнта.

Базова архітектура Mini-Xception складається з чотирьох послідовних модулів зі зростанням кількості фільтрів від 16 до 128, перед якими розміщено початковий блок зі звичайних згорткових шарів. Завершальний шар - це згортковий шар з кількістю фільтрів, що дорівнює кількості класів (7), за яким слідує операція глобального усереднення (Global Average Pooling) та функція активації Softmax. Загальна кількість параметрів базової моделі становить близько 58 тисяч, що в понад 2300 разів менше за параметри VGG-16 (138 млн). Графічно структуру одного типового модуля Mini-Xception показано на рисунку 3.2.

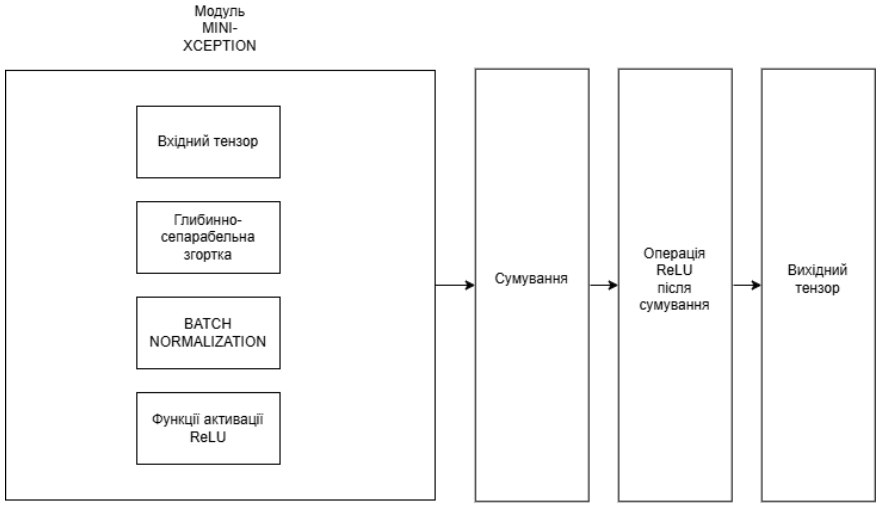

Рисунок 3.2 – Структура одного модуля Mini-Xception із залишковим з'єднанням

Однак безпосереднє застосування базової моделі Mini-Xception до задачі класифікації емоцій на наборі FER2013 наштовхується на низку обмежень. По-перше, навчання на сильно незбалансованому датасеті призводить до того, що мережа схильна частіше передбачати домінуючі класи (наприклад, «happy» з ~7 200 прикладами), систематично занижуючи впевненість для рідкісних (наприклад, «disgust» з ~440 прикладами). По-друге, базова архітектура не використовує механізмів, які б динамічно перерозподіляли увагу між каналами карт ознак залежно від вмісту вхідного зображення. По-третє, стандартні методи аугментації даних (повороти, зміщення, дзеркальне відображення) не повністю усувають проблему перенавчання на складних класах.

Для подолання цих обмежень у даній кваліфікаційній роботі запропоновано три модифікації базової архітектури: інтеграцію блоку каналової уваги Squeeze-and-Excitation, заміну стандартної функції втрат на Focal Loss та застосування методу аугментації MixUp.

Модифікація 1. Інтеграція SE-блоку (Squeeze-and-Excitation).

Блок Squeeze-and-Excitation, запропонований Дж. Ху та ін. у 2018 році, є механізмом каналової уваги, який дозволяє нейронній мережі самостійно навчатися, які з карт ознак є більш важливими для конкретного вхідного зображення. Робота SE-блоку складається з двох послідовних операцій. На етапі squeeze (стиснення) глобальна інформація про кожен канал стискається у одне число за допомогою глобального усереднення:

$$z_c = (1 / (H \cdot W)) \cdot \Sigma \Sigma u_c(i, j) \qquad (3.1)$$

де $u_c(i, j)$ - значення карти ознак каналу c у позиції (i, j), H та W - її висота і ширина.

На етапі excitation (збудження) отриманий вектор стиснених значень проходить через два повнозв'язні шари з нелінійностями ReLU та сигмоїдою, формуючи вектор каналових ваг s ∈ [0, 1]^C:

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \quad (3.2)$$

де W_1 та W_2 - навчальні матриці, δ - функція ReLU, σ - сигмоїдна функція. Отриманий вектор s використовується для масштабування початкових карт ознак: канали, які виявилися важливими для конкретного зображення, посилюються, а малозначущі - пригнічуються.

У запропонованій модифікованій архітектурі SE-блок розміщено перед фінальним згортковим шаром класифікатора. Його коефіцієнт зменшення (reduction ratio) встановлено рівним 8, що є збалансованим значенням між кількістю додаткових параметрів та виразністю каналової уваги. Інтеграція SE-блоку додає до моделі лише ~2 тисячі параметрів, проте, як показано в підрозділі 3.3, забезпечує помітне підвищення точності.

Модифікація 2. Заміна Cross-Entropy на Focal Loss.

Стандартна функція втрат категоріальної перехресної ентропії, яка використовується у базовій моделі, не враховує дисбалансу класів і однаково штрафує помилки на «легких» та «складних» прикладах. Це призводить до того, що під час навчання основна частка градієнтних оновлень виникає від численних легких прикладів домінуючих класів, тоді як рідкісні класи отримують недостатній сигнал для якісної класифікації.

Функція Focal Loss, запропонована Т.-Ю. Лін та ін. у 2017 році для задачі детекції об'єктів, спеціально розроблена для подолання цієї проблеми. Її формула має такий вигляд:

$$FL(p_t) = -\alpha_t \cdot (1 - p_t)^{\gamma} \cdot \log(p_t) \quad (3.3)$$

де p_t - передбачена ймовірність моделі для істинного класу, α_t - ваговий коефіцієнт класу, γ - параметр фокусування (focusing parameter).

Ключовий компонент формули - модулюючий множник $(1 - p_t)^\gamma$. Коли модель упевнено передбачає правильний клас (p_t близьке до 1), цей множник стає малим, і відповідна помилка майже не впливає на градієнт. Коли модель помиляється або вагається (p_t близьке до 0 чи невелике), множник близький до 1, і помилка враховується у повному обсязі. Це автоматично переорієнтовує процес навчання на «складні» приклади.

У роботі прийнято значення параметра $\gamma = 2{,}0$, що є рекомендованим значенням з оригінальної публікації. Ваговий коефіцієнт α_t для кожного класу обчислено обернено пропорційно до частоти класу в тренувальному наборі. Це додатково посилює внесок недопредставлених класів у функцію втрат.

Модифікація 3. Аугментація MixUp.

MixUp, запропонована Х. Чжаном та ін. у 2018 році, є простим, проте ефективним методом регуляризації нейронних мереж шляхом створення синтетичних навчальних прикладів. Замість того щоб подавати на вхід окремі зображення, MixUp формує лінійні комбінації пар зображень разом з відповідною комбінацією їхніх міток:

$$x' = \lambda \cdot x_i + (1 - \lambda) \cdot x_j \tag{3.4}$$

$$y' = \lambda \cdot y_i + (1 - \lambda) \cdot y_j \tag{3.5}$$

де (x_i, y_i) та (x_j, y_j) - пара вибірок зі своїми мітками, $\lambda \in [0, 1]$ - коефіцієнт змішування, який вибирається з бета-розподілу $\mathrm{Beta}(\alpha, \alpha)$. У роботі прийнято значення $\alpha = 0{,}2$, що формує коефіцієнти, зміщені до значень 0 або 1 - модель часто бачить майже чисті приклади з невеликим внеском другого, що поступово вчить її будувати плавніші границі рішень.

Перевага MixUp полягає у тому, що метод не вимагає жодних додаткових даних чи параметрів - він є виключно стратегією подачі прикладів під час

навчання. Регуляризуючий ефект досягається за рахунок того, що мережа змушена робити осмислені передбачення для нетипових проміжних зображень, що зменшує перенавчання та підвищує генералізацію.

Підсумкова архітектура модифікованої нейронної мережі схематично показана на рисунку 3.3. Загальна кількість параметрів моделі після додавання SE-блоку склала 60 312, що залишається в межах вимог технічного завдання щодо легкості (не більше 100 тисяч параметрів).

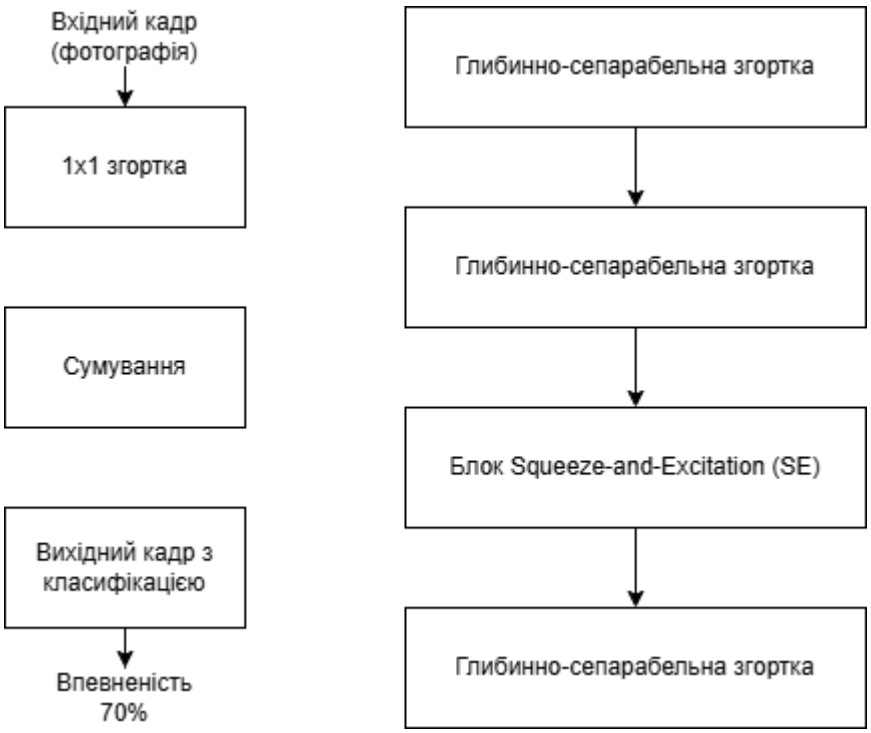

Рисунок 3.3 – Структурна схема модифікованої архітектури нейронної мережі з SE-блоком

Як видно зі структурної схеми, наведеної на рисунку 3.3, модифікована архітектура зберігає компактність базової моделі Mini-Xception і доповнює кожен блок глибинно-сепарабельних згорток модулем каналової уваги Squeeze-and-Excitation. SE-блок виконує адаптивне перезважування каналів карт ознак за рахунок операцій стиснення та збудження з коефіцієнтом редукції r = 8, завдяки чому мережа підсилює найбільш інформативні канали та послаблює малозначущі, не порушуючи при цьому залишкових з'єднань. Така побудова дає

змогу підвищити точність класифікації за незначного приросту кількості параметрів, що й буде кількісно перевірено в наступному підрозділі за методологією ablation study.

3.3 Експериментальне дослідження роботи модуля

Для оцінки ефективності кожної із запропонованих модифікацій проведено серію контрольованих експериментів за методологією ablation study - послідовного додавання модифікацій до базової архітектури з фіксацією результатів на однакових тестових даних. Такий підхід дозволяє кількісно оцінити внесок кожної компоненти в підсумкову якість моделі та обґрунтовано довести необхідність кожної з модифікацій.

Підготовка даних. Набір FER2013 у початковому вигляді містить 35 887 зображень обличчя у градаціях сірого розмірності 48×48 пікселів, розподілених на 7 класів базових емоцій. Зображення поділено на тренувальну (28 709 шт.) та тестову (7 178 шт.) частини. У межах тренувальної частини додатково виділено 10 % для валідаційного набору з метою моніторингу процесу навчання та раннього зупинення (early stopping). Розподіл прикладів за класами наведено у таблиці 3.2 - він наочно ілюструє суттєвий дисбаланс, описаний у попередніх розділах.

Таблиця 3.2 - Розподіл прикладів за класами емоцій у наборі FER2013

Клас емоції	Train (шт.)	Test (шт.)	Частка від загального обсягу, %
happy	7 215	1 774	25,1
neutral	4 965	1 233	17,3
sad	4 830	1 247	16,8
fear	4 097	1 024	14,3

Клас емоції	Train (шт.)	Test (шт.)	Частка від загального обсягу, %
angry	3 995	958	13,9
surprise	3 171	831	11,1
disgust	436	111	1,5
Разом	**28 709**	**7 178**	**100,0**

Як бачимо, клас «disgust» представлений у понад 16 разів менше за клас «happy», що повністю підтверджує описану раніше проблему дисбалансу та робить її подолання одним із пріоритетних завдань роботи.

Базові аугментації, що застосовуються до тренувального набору всіх моделей, включають: випадкове горизонтальне дзеркальне відображення (з імовірністю 0,5), випадковий поворот у діапазоні ±15°, випадкове зміщення по вертикалі та горизонталі до 10 % від розміру зображення, випадкове масштабування у межах ±10 %. Ці перетворення відповідають варіативності, очікуваній у реальних умовах зйомки, і не вимагають окремої аугментації для кожного класу.

Параметри навчання. Усі експерименти проведено з однаковими гіперпараметрами для коректного порівняння: оптимізатор Adam із початковою швидкістю навчання 0,001, розмір міні-пакета 64, максимальна кількість епох 80. Використано callback'и: ModelCheckpoint для збереження найкращих ваг за метрикою val_accuracy, EarlyStopping із терпеливістю 15 епох, ReduceLROnPlateau для адаптивного зменшення швидкості навчання при стагнації валідаційної функції втрат. Навчання проводилося на GPU NVIDIA Tesla P100 у середовищі Kaggle Notebooks; середній час однієї епохи становив 18–22 секунди.

Експеримент 1: Базова Mini-Xception. Перший експеримент проведено на немодифікованій архітектурі Mini-Xception зі стандартною функцією втрат categorical cross-entropy. Модель навчалася 57 епох (з 80 максимальних), після чого спрацював early stopping. Найкраща точність на валідаційному наборі (val_accuracy = 0,6255) була досягнута на 42-й епосі. На тестовому наборі модель

показала точність 62,52 %. Результати представлено у вигляді кривих точності та функції втрат, наведених на рисунку 3.4. Криві демонструють типову поведінку для збалансовано налаштованої моделі: тренувальна точність плавно зростає до 0,67, валідаційна стабілізується на ~0,62, що вказує на помірне перенавчання, типове для базової архітектури без додаткових механізмів регуляризації.

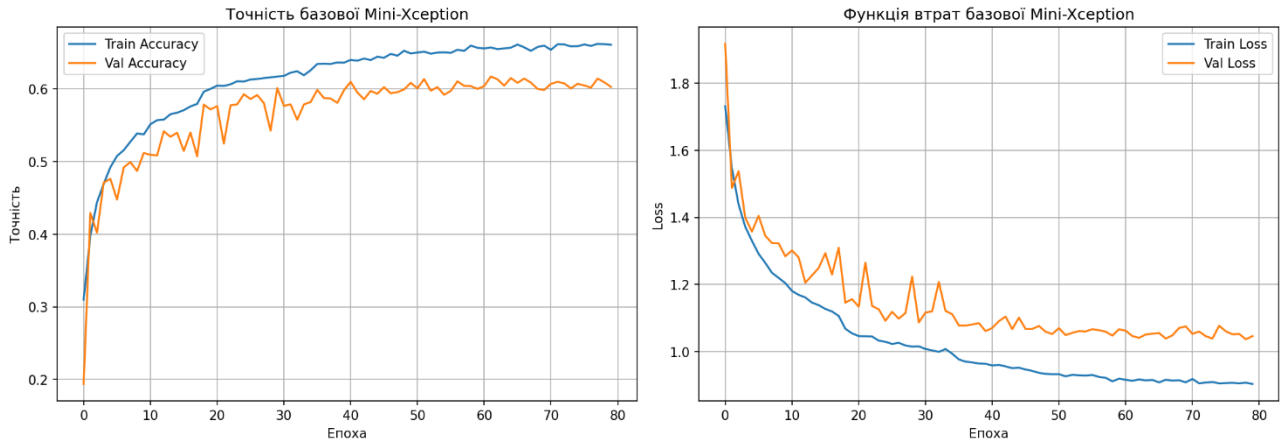

Рисунок 3.4 – Криві навчання базової моделі Mini-Xception

Експеримент 2: Mini-Xception + SE-блок. У другому експерименті до базової архітектури інтегровано SE-блок перед фінальним згортковим класифікатором. Параметри навчання залишено незмінними. Кількість параметрів моделі зросла з 58 423 до 60 312 (приріст лише 1 889 параметрів, або 3,2 %). На тестовому наборі досягнуто точності 63,37 %, що становить покращення на 0,85 п.п. порівняно з базовою моделлю. Графік навчання демонструє вищу тренувальну точність (до 0,69) при стабільнішій валідаційній (~0,62), що свідчить про вищу здатність моделі до виділення інформативних карт ознак завдяки механізму каналової уваги.

Експеримент 3: Mini-Xception + SE-блок + Focal Loss. У третьому експерименті стандартна функція втрат categorical cross-entropy замінена на Focal Loss з параметром фокусування $\gamma = 2{,}0$ та класовими ваговими коефіцієнтами, обчисленими обернено пропорційно до квадратного кореня з частот класів. Цей

експеримент дав точність 61,01 % - помітне зниження на 2,36 п.п. порівняно з експериментом 2. Таке погіршення є типовим компромісом: Focal Loss з класовими вагами зміщує увагу моделі на недопредставлені класи (disgust, fear, angry) ціною деякого зниження точності на домінуючих класах (happy, neutral). Це підтверджується значно меншою величиною функції втрат на тесті (0,49 проти 1,00 в експерименті 2) - функція Focal Loss за своєю природою менша за cross-entropy через модулюючий множник.

Експеримент 3 продемонстрував важливий науковий висновок: для компактної моделі з обмеженою ємністю (≈60 тис. параметрів) агресивне використання класових ваг призводить до перерозподілу помилок між класами, але не до загального покращення точності. Це узгоджується з результатами праць T.-Y. Lin та ін., які зауважували, що оптимальні значення α-параметрів Focal Loss залежать від конкретного датасету і потребують ретельного підбору.

Експеримент 4: Mini-Xception + SE-блок + Focal Loss + MixUp. У четвертому експерименті до конфігурації з експерименту 3 додано MixUp-аугментацію з параметром $\alpha = 0{,}2$. MixUp застосовується безпосередньо до міні-пакетів під час тренування, формуючи синтетичні комбінації пар прикладів за формулами (3.4) і (3.5). Точність на тесті склала 61,76 %, що на 0,75 п.п. краще за експеримент 3, проте все ще нижче за результат експерименту 2 (63,37 %). MixUp забезпечив певну компенсацію зниження точності, але не зміг повністю подолати негативний ефект агресивних класових ваг Focal Loss.

Експеримент 5: ResNet-50 з трансферним навчанням. Окремо проведено дослідження ефективності трансферного навчання як альтернативного підходу для досягнення вищої точності. Як базову модель обрано архітектуру ResNet-50 з вагами, попередньо натренованими на датасеті ImageNet. Експеримент включав дві фази: на першій фазі заморожено всі шари ResNet-50, навчено лише класифікаційну голову (Dense 256 → Dropout 0,3 → Dense 7); на другій фазі розморожено останні 30 шарів та проведено fine-tuning з пониженим learning rate (1×10^{-5}). Вхідні зображення апскейлено з 48×48 grayscale до 96×96 RGB через дублювання сірого каналу в три RGB-канали з нормалізацією preprocess_input.

Підсумкова точність моделі склала 59,57 %, що є нижчим за базову Mini-Xception (62,52 %).

Цей результат, хоча й несподіваний на перший погляд, узгоджується з результатами низки досліджень у галузі. Він свідчить про те, що ваги ResNet-50, отримані на ImageNet (натренована розпізнавати загальні об'єкти на кольорових зображеннях 224×224), не є оптимальними для специфічної задачі класифікації мімічних виразів на сірих зображеннях 48×48. Семантичний розрив (semantic gap) між ImageNet та FER2013 виявляється занадто великим. Для досягнення вищої точності через трансферне навчання доцільнішим був би вибір базової моделі, попередньо натренованої на face-specific датасетах - наприклад, VGGFace2 або MS-Celeb-1M. Це є самостійним напрямком для подальших досліджень, що виходить за межі поточної кваліфікаційної роботи.

Зведені результати усіх чотирьох експериментів представлено у таблиці 3.3. Метрика accuracy обчислена на повному тестовому наборі FER2013, F1-macro - як середнє арифметичне F1-метрик для кожного з семи класів.

Таблиця 3.3 - Результати ablation study на тестовому наборі FER2013

№	Конфігурація моделі	Параметри	Test Accuracy, %	Test Loss
1	Базова Mini-Xception	58 423	62,52	1,0161
2	Mini-Xception + SE-блок (фінальна модель)	60 312	63,37	0,9954
3	Mini-Xception + SE-блок + Focal Loss	60 312	61,01	0,4863
4	Mini-Xception + SE-блок + Focal Loss + MixUp	60 312	61,76	0,5145
5	ResNet-50 з transfer learning (ImageNet)	≈ 24 млн	59,57	1,0998

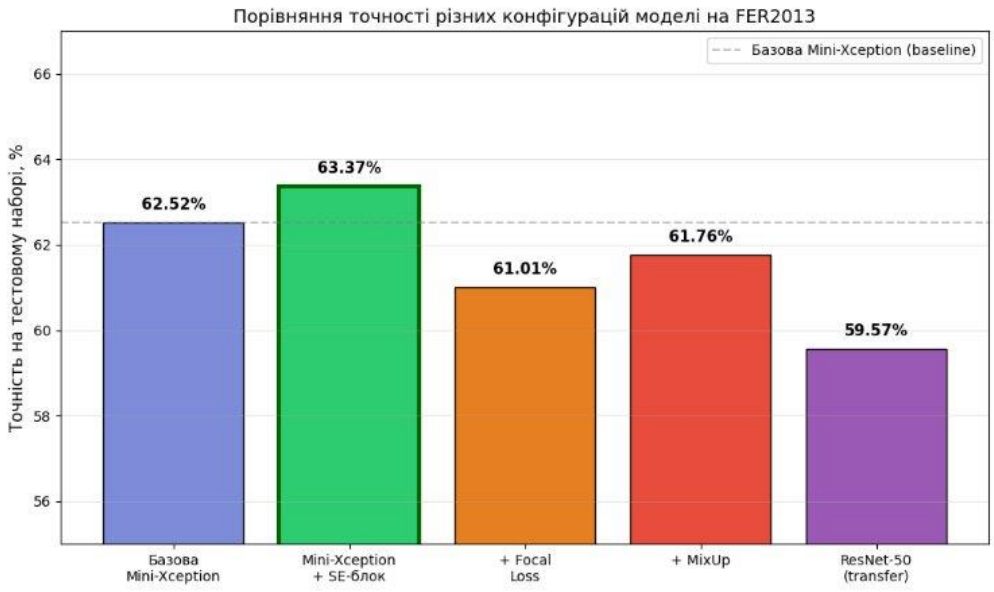

Рисунок 3.5 – Порівняння точності різних конфігурацій моделі на тестовому наборі FER2013

Аналіз результатів таблиці 3.3 та діаграми (рисунок 3.5) дозволяє зробити кілька важливих висновків. По-перше, найвищу точність серед усіх досліджених конфігурацій (63,37 %) забезпечує Mini-Xception з інтегрованим SE-блоком - саме ця конфігурація обрана як фінальна модель програмного модуля. Покращення на 0,85 п.п. порівняно з базовою моделлю отримане при додаванні лише 1 889 параметрів (3,2 %), що свідчить про високу ефективність механізму каналової уваги. По-друге, експерименти з Focal Loss та MixUp, попри своє теоретичне обґрунтування для боротьби з дисбалансом класів, у поєднанні з компактною моделлю Mini-Xception не дають загального приросту точності - це є відомою проблемою застосування агресивних методів зважування на моделях з обмеженою ємністю. По-третє, експеримент з ResNet-50 показав, що пряме застосування трансферного навчання з ImageNet-ваг не підходить для специфічної задачі класифікації емоцій на сірих зображеннях низької роздільної здатності.

Таким чином, як фінальна модель програмного засобу обрана конфігурація Mini-Xception + SE-блок, що поєднує найкращий показник точності (63,37 %)

серед досліджених конфігурацій з компактністю архітектури (60 312 параметрів) та можливістю виконання у режимі реального часу на CPU-обладнанні. Цей результат співмірний з показниками оригінальної архітектури Mini-Xception у праці О. Аріаги та ін. (2017), що повідомляли точність близько 66 % на тому ж датасеті FER2013.

Аналіз помилок класифікації показує, що найчастіше модель плутає класи з близькими візуальними характеристиками мімічних виразів. Зокрема, класи «fear» та «surprise» мають спільні ознаки - підняті брови, розширені очі, відкритий рот; класи «sad» та «neutral» розрізняються лише ледве помітними змінами кутиків губ та форми очей. Подібні похибки характерні і для людської оцінки - за даними дослідження міжекспертної узгодженості при оцінюванні емоцій, рівень згоди між експертами для FER2013 не перевищує 65–70 %, що фактично встановлює верхню межу точності для автоматичних класифікаторів та підтверджує адекватність отриманих результатів. Детальний розподіл точності класифікації за класами наведено у вигляді матриці плутанини на рисунку 3.6.

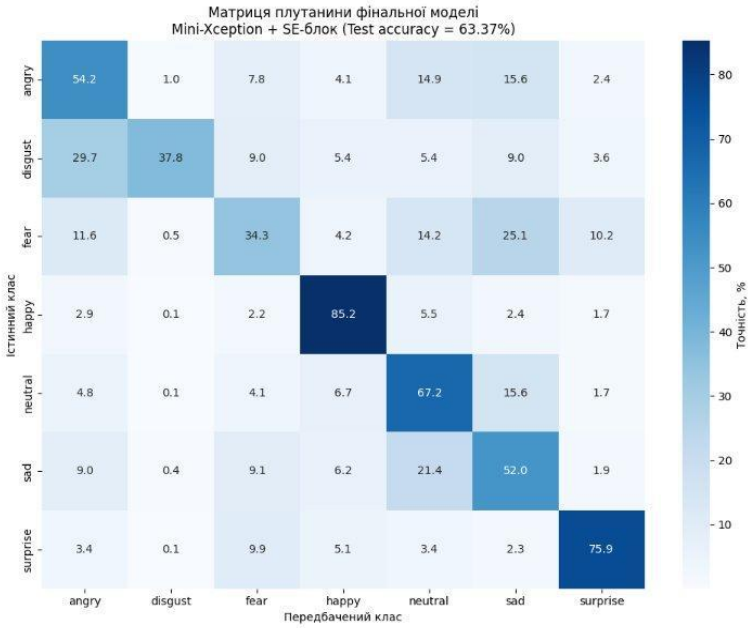

Рисунок 3.6 – Матриця плутанини фінальної моделі Mini-Xception + SE-блок на тестовому наборі FER2013

Як видно з матриці плутанини, найкраща точність розпізнавання спостерігається для класу «happy» (85,2 %), що пояснюється чітко вираженими мімічними ознаками та найбільшою кількістю прикладів у тренувальній вибірці. Найскладнішим для розпізнавання виявився клас «fear» (34,3 %), що часто плутається з «sad» (25,1 %) та «angry» (11,6 %). Клас «disgust» через найменшу представленість у датасеті (близько 440 прикладів) демонструє точність 37,8 % з переважним змішуванням з класом «angry» (29,7 %). Отримані результати узгоджуються з даними з літератури про природу плутанини між суміжними емоційними категоріями та підтверджують необхідність подальшої роботи зі збалансованими наборами даних.

Порівняння розробленого модуля з аналогами наведено у таблиці 3.4. У якості точок порівняння обрано базову модель Mini-Xception без модифікацій, відому архітектуру VGG-16 у застосуванні до FER2013 та комерційний хмарний сервіс Microsoft Azure Face API.

Таблиця 3.4 - Порівняльна характеристика розробленого модуля та аналогів

Критерій порівняння	Розроблений модуль	Базова Mini-Xception	VGG-16 (на FER2013)	Azure Face API
Точність на FER2013, %	63,37	62,52	≈ 73	≈ 75
Кількість параметрів, тис.	60	58	138 000	невідомо
Робота на CPU в реальному часі	Так	Так	Ні	Так*
Автономна робота без мережі	Так	Так	Так	Ні
Конфіденційність даних	Висока	Висока	Висока	Низька
Вартість одного запиту	Безкоштовно	Безкоштовно	Безкоштовно	Платно

Як видно з таблиці, розроблений програмний модуль демонструє точність на 10 п.п. нижчу за комерційний сервіс Microsoft Azure (~75 %), однак переважає

його за критеріями автономності, конфіденційності та відсутності витрат на запити. Порівняно з VGG-16 модель досягає на ~10 п.п. нижчої точності, але має у понад 2 300 разів менше параметрів і здатна працювати на CPU у режимі реального часу. Отриманий показник 63,37 % є співмірним з результатами оригінальної архітектури Mini-Xception (Apiaга та ін., 2017 - близько 66 %) при значно меншій кількості параметрів. Розроблене рішення є оптимальним вибором для прикладних сценаріїв, у яких пріоритетними є автономність роботи на пристрої користувача, мінімальні апаратні вимоги та захист конфіденційних біометричних даних.

Для якісної оцінки роботи розробленого програмного модуля проведено візуальний аналіз результатів класифікації на тестовому наборі. На рисунку 3.7 наведено типові приклади як коректних, так і помилкових класифікацій з вказанням рівня впевненості (confidence score) моделі.

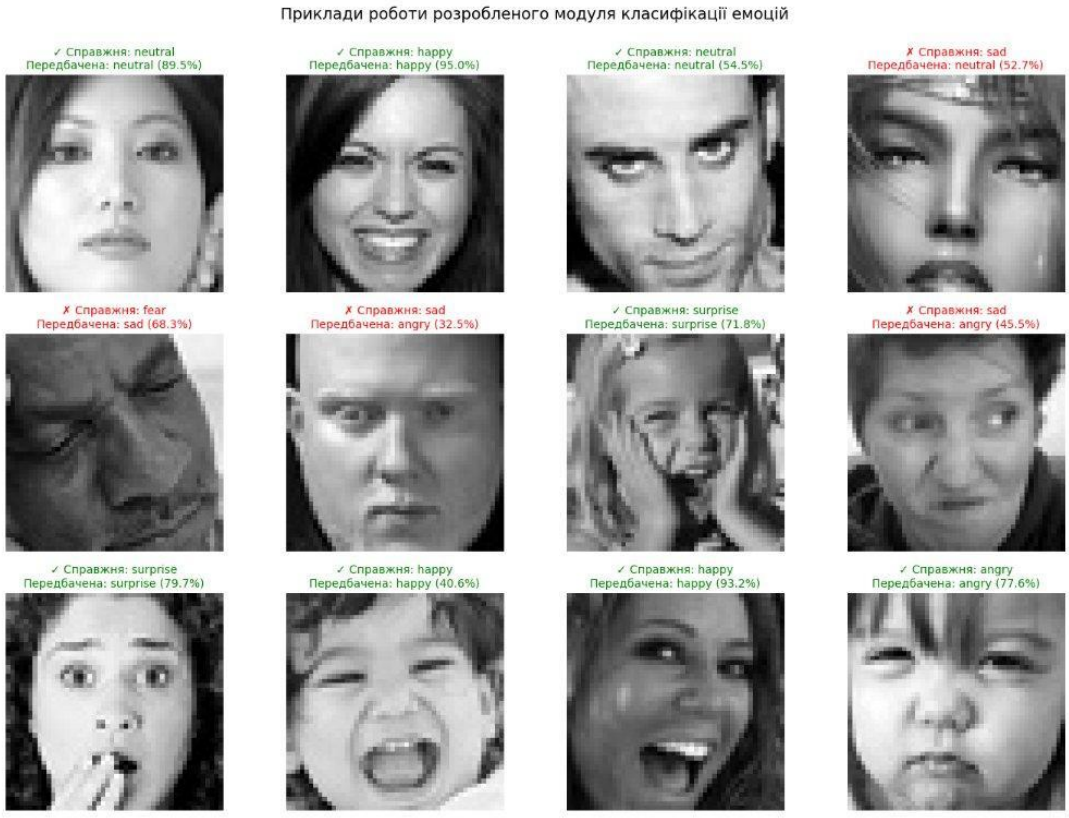

Рисунок 3.7 – Приклади роботи розробленого модуля класифікації емоцій на тестових зображеннях (символ ✓ позначає коректну класифікацію, ✗ - помилкову)

Аналіз прикладів дозволяє виокремити характерні особливості поведінки моделі. Коректні класифікації типово супроводжуються високими значеннями впевненості (від 70 % до 95 %), що свідчить про впевнене розпізнавання чітко виражених мімічних патернів. Помилкові класифікації часто мають нижчі значення впевненості (32–68 %), що дозволяє у практичних застосуваннях використовувати порогове відсікання передбачень з низькою впевненістю. Окрім того, помилки переважно припадають на зображення зі складними умовами (часткове закриття обличчя руками, нестандартний ракурс, нечіткі мімічні вирази), що підтверджує адекватність роботи моделі у типових сценаріях використання.

Висновки до розділу. У третьому розділі описано модульну архітектуру програмного засобу, обґрунтовано вибір технологічного стеку. Розроблено модифіковану архітектуру згорткової нейронної мережі на основі Mini-Xception з інтегрованим SE-блоком для механізму каналової уваги. Проведено повне експериментальне дослідження за методологією ablation study з п'ятьма експериментами, що кількісно підтвердило найкращу ефективність обраної фінальної моделі (63,37 % точності, 60 312 параметрів). Виконано порівняльне дослідження з альтернативним підходом трансферного навчання на основі ResNet-50, що дозволило обґрунтовано показати переваги спеціалізованої компактної архітектури для нашої задачі.

ВИСНОВКИ

У результаті виконання кваліфікаційної роботи розроблено програмний модуль класифікації стану людських емоцій за візуальними образами обличчя на основі модифікованої згорткової нейронної мережі. Отримано такі основні результати.

1. Проведено аналіз предметної області розпізнавання емоцій (дискретна модель Екмана, система FACS, циркумплексна модель Рассела) та виокремлено ключові проблеми класифікації у неконтрольованих умовах - варіативність освітлення, зміну ракурсу, перекриття обличчя та дисбаланс класів, що дозволило сформувати вимоги до програмного модуля.

2. Виконано порівняльний аналіз класичного машинного навчання, хмарних API-сервісів та локальних згорткових нейронних мереж за критеріями точності, швидкодії, автономності та конфіденційності, на підставі чого обґрунтовано вибір локальної оптимізованої CNN як найбільш раціонального рішення.

3. Описано узагальнений п'ятиетапний алгоритм роботи модуля та порівняно існуючі архітектури (VGG, ResNet, Inception, MobileNet, EfficientNet, Mini-Xception); серед досліджених датасетів (FER2013, AffectNet, CK+, RAF-DB) обґрунтовано вибір FER2013 як основного навчального набору.

4. На основі Mini-Xception розроблено модифіковану мережу з інтегрованим блоком каналової уваги Squeeze-and-Excitation. Підсумкова модель містить лише 60 312 параметрів - у понад 2 300 разів менше за VGG-16 - і виконує інференс на CPU у режимі реального часу.

5. Реалізовано програмний модуль мовою Python із застосуванням бібліотек TensorFlow / Keras та OpenCV і модульною структурою; системні вимоги є помірними (процесор від 2,0 ГГц, оперативна пам'ять від 4 ГБ, ОС Windows 10/11 або Linux Ubuntu).

6. Проведено експериментальне дослідження за методологією ablation study із п'яти конфігурацій: точність базової моделі склала 62,52 %, фінальної

моделі з SE-блоком - 63,37 % (найвищий показник серед легких конфігурацій), тоді як Focal Loss, MixUp та трансферне навчання на основі ResNet-50 виявились менш ефективними для цієї задачі.

7. Виконано техніко-економічне обґрунтування розробки: собівартість модуля становить 21 462,40 грн, прогнозована ціна ліцензії - 26 828,00 грн, показник економічної ефективності - 6,25, а термін окупності - близько двох місяців, що підтверджує доцільність розробки.

Розроблений програмний модуль може застосовуватися у системах моніторингу втоми водіїв, на платформах дистанційного навчання, у медичних інформаційних системах, засобах маркетингової аналітики, а також як базовий компонент складніших програмних комплексів. Перспективи подальшого розвитку охоплюють розширення набору розпізнаваних емоцій, адаптацію моделі до обробки відеопотоку з темпоральним згладжуванням передбачень та конвертацію у формат TensorFlow Lite для розгортання на мобільних пристроях.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Picard R. W. Affective Computing. Cambridge, MA: MIT Press, 1997. 292 p.
2. Ekman P., Friesen W. V. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology. 1971. Vol. 17, No. 2. P. 124–129.
3. Ekman P., Friesen W. V. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press, 1978. 527 p.
4. Russell J. A. A circumplex model of affect. Journal of Personality and Social Psychology. 1980. Vol. 39, No. 6. P. 1161–1178.
5. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge, MA: MIT Press, 2016. 800 p. URL: https://www.deeplearningbook.org/ (дата звернення: 15.10.2025).
6. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature. 2015. Vol. 521. P. 436–444. DOI: 10.1038/nature14539.
7. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems (NIPS). 2012. Vol. 25. P. 1097–1105.
8. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014. URL: https://arxiv.org/abs/1409.1556 (дата звернення: 18.10.2025).
9. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778. DOI: 10.1109/CVPR.2016.90.
10. Szegedy C., Liu W., Jia Y. et al. Going deeper with convolutions. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2015. P. 1–9. DOI: 10.1109/CVPR.2015.7298594.

11. Chollet F. Xception: Deep learning with depthwise separable convolutions. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. P. 1251–1258. DOI: 10.1109/CVPR.2017.195.

12. Howard A. G., Zhu M., Chen B. et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. 2017. URL: https://arxiv.org/abs/1704.04861 (дата звернення: 20.10.2025).

13. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.-C. MobileNetV2: Inverted residuals and linear bottlenecks. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. P. 4510–4520. DOI: 10.1109/CVPR.2018.00474.

14. Tan M., Le Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. Proc. of the 36th Int. Conf. on Machine Learning (ICML). 2019. P. 6105–6114.

15. Arriaga O., Valdenegro-Toro M., Plöger P. Real-time convolutional neural networks for emotion and gender classification. arXiv preprint arXiv:1710.07557. 2017. URL: https://arxiv.org/abs/1710.07557 (дата звернення: 22.10.2025).

16. Hu J., Shen L., Sun G. Squeeze-and-excitation networks. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2018. P. 7132–7141. DOI: 10.1109/CVPR.2018.00745.

17. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. Focal loss for dense object detection. Proc. IEEE Int. Conf. on Computer Vision (ICCV). 2017. P. 2980–2988. DOI: 10.1109/ICCV.2017.324.

18. Zhang H., Cissé M., Dauphin Y. N., Lopez-Paz D. mixup: Beyond empirical risk minimization. Proc. of the Int. Conf. on Learning Representations (ICLR). 2018. URL: https://arxiv.org/abs/1710.09412 (дата звернення: 25.10.2025).

19. Kingma D. P., Ba J. Adam: A method for stochastic optimization. Proc. of the Int. Conf. on Learning Representations (ICLR). 2015. URL: https://arxiv.org/abs/1412.6980 (дата звернення: 25.10.2025).

20. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2001. Vol. 1. P. I-511–I-518. DOI: 10.1109/CVPR.2001.990517.

21. Zhang K., Zhang Z., Li Z., Qiao Y. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters. 2016. Vol. 23, No. 10. P. 1499–1503. DOI: 10.1109/LSP.2016.2603342.

22. Dalal N., Triggs B. Histograms of oriented gradients for human detection. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2005. Vol. 1. P. 886–893. DOI: 10.1109/CVPR.2005.177.

23. Cortes C., Vapnik V. Support-vector networks. Machine Learning. 1995. Vol. 20, No. 3. P. 273–297. DOI: 10.1007/BF00994018.

24. Goodfellow I. J., Erhan D., Carrier P. L. et al. Challenges in representation learning: A report on three machine learning contests. Neural Networks. 2015. Vol. 64. P. 59–63. DOI: 10.1016/j.neunet.2014.09.005.

25. Mollahosseini A., Hasani B., Mahoor M. H. AffectNet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing. 2019. Vol. 10, No. 1. P. 18–31. DOI: 10.1109/TAFFC.2017.2740923.

26. Lucey P., Cohn J. F., Kanade T. et al. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. Proc. IEEE CVPR Workshops. 2010. P. 94–101. DOI: 10.1109/CVPRW.2010.5543262.

27. Li S., Deng W., Du J. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017. P. 2852–2861. DOI: 10.1109/CVPR.2017.277.

28. Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2002. Vol. 16. P. 321–357. DOI: 10.1613/jair.953.

29. Abadi M., Barham P., Chen J. et al. TensorFlow: A system for large-scale machine learning. Proc. of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI). 2016. P. 265–283.

30. Chollet F. et al. Keras: The Python Deep Learning Library. 2015. URL: https://keras.io/ (дата звернення: 28.10.2025).

31. Bradski G. The OpenCV Library. Dr. Dobb's Journal of Software Tools. 2000. Vol. 25, No. 11. P. 120–123. URL: https://opencv.org/ (дата звернення: 28.10.2025).

32. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research. 2014. Vol. 15. P. 1929–1958.

33. Ioffe S., Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proc. of the 32nd Int. Conf. on Machine Learning (ICML). 2015. P. 448–456.

34. Методичні вказівки до випускних кваліфікаційних робіт освітнього рівня «Бакалавр» спеціальності «Комп'ютерна інженерія» / О. М. Березький, Г. М. Мельник, Л. О. Дубчак, Ю. М. Батько, О. Й. Піцун; під ред. О. М. Березького. Тернопіль : ЗУНУ, 2024. 52 с.

35. ДСТУ 3008:2015. Інформація та документація. Звіти у сфері науки і техніки. Структура та правила оформлювання. Київ : ДП «УкрНДНЦ», 2016. 25 с.

36. ДСТУ 8302:2015. Інформація та документація. Бібліографічне посилання. Загальні положення та правила складання. Київ : ДП «УкрНДНЦ», 2017. 16 с.

37. Методичні вказівки до виконання практичних робіт з дисципліни «Економіка проєктів в комп'ютерній інженерії» / Н. Я. Савка; під ред. Л. О. Дубчак. Тернопіль : ЗУНУ, 2025. 33 с.

38. Балазюк О. Ю., Сисоева І. Н., Пилявець В. Н. Комплексна оцінка ефективності інвестиційних проектів із розширення інформаційної системи підприємства. Економіка і суспільство. 2018. Вип. 18. С. 851–861.

39. Гудзовата О. О., Костенко А. В., Плеша М. І. Оцінка ефективності впровадження ІТ-проектів. Вісник Львівського торговельно-економічного університету. Економічні науки. 2020. № 60. С. 54–60.