

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Система розпізнавання емоцій людини на відео
(тема)

Виконав:
студент 2 курсу, групи СШМ-22-3
Алієв К.А.
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системи штучного інтелекту
(повна назва спеціалізації)

Керівник проф. Музика К.М.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

В.О. Філатов
(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____
(повна назва)
Кафедра _____ Штучного інтелекту _____
(повна назва)
Рівень вищої освіти _____ другий (магістерський) _____
Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)
Тип програми _____ освітньо-наукова _____
(освітньо-професійна або освітньо-наукова)
Освітня програма _____ Системи штучного інтелекту _____
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«_____» _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _____ Алієву Каріму Абдулгадир огли _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Система розпізнавання емоцій людини на відео _____

затверджена наказом університету від 1 квітня 2024 р. № 260Ст

2. Термін подання студентом роботи до екзаменаційної комісії 4 червня 2024 р.

3. Вихідні дані до роботи _____ Науково-технічні публікації, дані Інтернет-джерел, книг та відомих наукових робіт _____

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі _____

2) Аналіз існуючих рішень, алгоритмів та датасетів _____

3) Опис розробленої системи та аналіз результатів _____

4) Розробка стартап проекту _____

РЕФЕРАТ

Пояснювальна записка: 77 с., 26 рис., 23 табл., 1 дод., 15 джерел.

ЕМОЦІЇ, ЕМОЦІЙНИЙ ІНТЕЛЕКТ, МАШИННЕ НАВЧАННЯ,
НЕЙРОННА МЕРЕЖА, ШТУЧНИЙ ІНТЕЛЕКТ.

Об'єкт дослідження – сучасні підходи, які використовуються у розпізнаванні емоцій людини на відео.

Предмет дослідження – здійснення аналізу математичної моделі для розпізнавання емоцій людини на відео.

Метою роботи є аналіз та розроблення системи розпізнавання емоцій людини на відео.

Методи дослідження – різні алгоритми машинних та нейронних мереж.

ABSTRACT

Master`s thesis contains: 77 pp., 26 fig., 23 tabl., 1 ann., 15 references.

**ARTIFICIAL INTELLIGENCE, EMOTIONAL INTELLIGENCE,
EMOTIONS, MACHINE LEARNING, NEURAL NETWORK.**

Object of research is modern approaches used in recognizing human emotions in video.

The subject of research is analyzing a mathematical model for recognizing human emotions in video.

The aim of the study is to analyze and develop a system for recognizing human emotions in video.

Research methods are various algorithms of machine and neural networks.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	9
Вступ.....	10
1 Аналіз предметної галузі	13
1.1 Емоційний штучний інтелект	13
1.1.1 Емоційний інтелект та його значення.....	13
1.1.2 Штучний інтелект	14
1.2 Загальний огляд проблем в задачі розпізнавання емоцій людини	14
1.2.1 Проблема набору даних.....	15
1.2.2 Похибки в розпізнанні неповністю виражених емоціях.....	16
1.2.3 Відмінність західної та східної культури	17
1.3 Висновки до розділу	19
2 Аналіз існуючих рішень, алгоритмів та датасетів	20
2.1 Існуючі рішення	20
2.1.1 EmoDetect.....	20
2.1.2 Хеома	21
2.1.3 FaceReader (Noldus)	23
2.1.4 Microsoft Oxford Project Emotion Recognition	24
2.1.5 eMotion Softwre	25
2.2 Допоміжні алгоритми в задачі розпізнавання емоцій.....	26
2.2.1 Гістограма направлених градієнтів (HOG)	26
2.2.2 Локальні бінарні шаблони (LBP).....	28
2.2.3 Метод Віоли-Джонса	29
2.2.4 Метод гнучкого порівняння на графах	32
2.2.5 Геометричний метод розпізнавання обличчя	33
2.3 Традиційні методи розпізнавання емоцій	34
2.3.1 Розріджена логістична регресія зі штрафом L1/2.....	34
2.3.2 Метод К-найближчих сусідів (KNN).....	35
2.3.3 Метод опорних векторів (SVM)	37

2.4 Існуючі датасети для розпізнавання емоцій.....	40
2.4.1 AffectNet.....	40
2.4.2 Ascertain	41
2.4.3 Dreamer.....	41
2.4.4 Extended Cohn-Kanade Dataset (CK+).....	42
2.4.5.Emotic	43
2.4.6 FER-2013	45
2.4.7 Google Facial Expression Comparison Dataset.....	46
2.4.8 K-EmoCon	46
2.4.9 Interactive Emotional Dyadic Motion Capture (IEMOCAP)	47
2.5 Висновки до розділу	49
3 Опис розробленої системи та аналіз результатів.....	50
3.1 Опис розробленого алгоритму.....	50
3.1.1 Згорткова нейронна мережа (ЗНМ).....	50
3.1.2 Глибока згорткова мережа з прямим зв'язком.....	52
3.1.3 Модифікація згорткової нейронної мережі.....	53
3.2 Обрана метрика та якісне порівняння методів.....	54
3.2.1 Обрана метрика	54
3.2.2 Результати попередніх методів.....	56
3.2.3 Результати розробленого методу	57
3.3. Опис компонентів побудованої системи	57
3.4 Висновки до розділу	58
4 Розробка стартап-проекту	59
4.1.Опис ідеї проекту	59
4.2.Технологічний аудит ідеї проекту.....	60
4.3.Аналіз ринкового потенціалу для старту стартап-проекту	60
4.4.Розробка ринкової стратегії стартап-проекту	65
4.5.Розроблення маркетингової програми стартап-проекту.....	67
4.6 Висновки до розділу	69
Висновки	70

Перелік джерел посилання	73
Додаток А Відомість кваліфікаційної роботи	75

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

Датасет – набір даних та їх розмітка (в нашому випадку набір фотографій та список емоцій до кожної з них);

ШІ – штучний інтелект;

CNN – Convolutional Neural Network – згорткова нейронна мережа;

FDCNN – Feedforward Deep Convolution Neural Network – нейронна мережа глибокої згортки зворотного зв'язку;

KNN – K-Nearest Neighbors Algorithm – метод K-найближчих сусідів;

SVM – Support Vector Machine – метод опорних векторів.

ВСТУП

Актуальність теми дослідження. В наш час, сфера Штучного інтелекту та машинного навчання вже виявляє складні та вдосконалені алгоритми. Навіть у минулому столітті машини демонстрували здатність до високого рівня в проходженні тесту Тюрінга. Сучасності часто стає вкрай складно відрізнити взаємодію з людиною чи взаємодію зі штучним інтелектом, якщо відсутні візуальні або звукові індикатори. Видатними прикладами є гуманоїдні роботи, такі як Софія. Прогнозується, що через двадцять років роботи зможуть жити поруч з людьми та активно допомагати їм у повсякденних справах, і при цьому їх важко буде відрізнити від реальних осіб.

Сучасні мобільні пристрої вже мають функції розпізнавання обличчя та голосу. Ці технології використовують алгоритми, які постійно навчаються та можуть розпізнати кілька обличч людей. Хоча зміни в людському обличчі можуть статися, програми навчаються адаптуватися та розпізнавати обличчя у різних умовах освітлення, а також при наявності змін, таких як наявність бороди чи вусів.

Основною характеристикою будь-якого інтелекту є його здатність до навчання. Залежно від того, які знання і навички він здобуде в оточенні людей, таким він буде у своєму поведінці.

Також інтелект може бути спеціально навчений певним діям, аналогічно тому, як це відбувається у випадку з дитиною. Водночас, значна кількість дослідників підкреслює, що загрозою для людства може стати бездоганне виконання завдань штучним інтелектом, що може призвести до ігнорування інтересів людей або створення умов, що несумісні з існуванням людства.

Марк Цукерберг, програміст і засновник соціальної мережі Facebook, вважає, що штучний інтелект може значно полегшити життя людей, допомагаючи вирішувати багато проблем. Наприклад, за його словами, за

допомогою штучного інтелекту люди зможуть отримувати більш якісне медичне обслуговування та діагностику захворювань, а також зменшити кількість автомобільних аварій, що є однією з головних причин смертності.

У сучасному світі технології вже допомагають зменшити забруднення навколишнього середовища, а в майбутньому можуть навіть врятувати людство від цієї проблеми. В будь-якому випадку, навіть якщо припинити розвиток штучного інтелекту неможливо, важливо надавати йому можливість вчитися емпатії, любові до людей та бажанню допомагати і захищати їх.

Метою роботи є аналіз та розроблення системи розпізнавання емоцій людини на відео.

Об'єкт дослідження – сучасні підходи, які використовуються у розпізнаванні емоцій людини на відео.

Предмет дослідження – здійснення аналізу математичної моделі для розпізнавання емоцій людини на відео.

Практичне значення отриманих результатів полягає в їхньому потенціалі для значного покращення якості життя людей у різних сферах. Зокрема, системи розпізнавання емоцій на відео можуть знайти застосування в медичній сфері, сприяючи поліпшенню діагностики та лікування пацієнтів шляхом раннього виявлення емоційних розладів, таких як депресія та тривога. Це дозволить своєчасно надавати необхідну допомогу, що може мати критичне значення для ефективного лікування. Крім того, алгоритми розпізнавання емоцій можуть бути впроваджені в системи безпеки, де вони допоможуть аналізувати поведінку людей у громадських місцях. Це дозволить виявляти потенційно небезпечні ситуації та попереджати злочини, тим самим підвищуючи рівень безпеки суспільства. В автомобільній індустрії використання таких систем може сприяти зменшенню кількості дорожньо-транспортних пригод через виявлення ознак втоми або роздратування водія і своєчасного

попередження про необхідність зробити перерву, що сприятиме безпеці на дорогах.

У сфері освіти системи розпізнавання емоцій можуть бути корисними для викладачів, допомагаючи їм краще розуміти емоційний стан студентів. Це дозволить виявляти моменти, коли студенти відчують труднощі або втрачають інтерес до навчання, і своєчасно надавати їм необхідну підтримку. У маркетингу та обслуговуванні клієнтів впровадження таких систем допоможе компаніям краще розуміти реакції клієнтів на продукти чи послуги, що дозволить удосконалювати маркетингові стратегії та підвищувати якість обслуговування. Соціальні мережі та комунікації також можуть виграти від використання систем розпізнавання емоцій, оскільки це сприятиме створенню більш персоналізованого контенту та покращенню взаємодії між користувачами.

Таким чином, розроблення системи розпізнавання емоцій людини на відео має широкий спектр застосувань, здатних значно покращити якість життя, сприяти розвитку технологій та забезпечувати безпеку і комфорт у повсякденному житті.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Емоційний штучний інтелект

1.1.1 Емоційний інтелект та його значення

Міжособистісне спілкування людей – це не лише словесні вислови, але й невербальні вияви, такі як рухи, жести, вираз обличчя та тон голосу. Ці невербальні засоби використовуються для вираження емоцій, передачі настрою та забезпечення ефективного спілкування.

Емоційний інтелект (EQ) включає в себе розуміння власних і чужих емоцій, вміння виражати їх конструктивно та розуміти почуття та переживання оточуючих. Від давніх часів, коли люди ще не володіли мовою, вони здатні були сприймати емоції один одного та використовувати їх у повсякденному житті.

У сучасному світі велика потреба в аналізі людських емоцій, оскільки ми переживаємо новий етап розвитку людства. Роботи вже стали невід'ємною частиною нашого життя і продовжують активно проникати в різні сфери людського існування. Даний застосунок може бути використаний як у рекламних кампаніях, так і в галузі робототехніки.

Вчені виявили, що лише 7% інформації у людських комунікаціях передається вербально (мовою), 38% – за допомогою модуляцій голосу, і 55% – через вираз обличчя. Розпізнавання та класифікація емоцій стали важливим напрямком для численних досліджень у галузі комп'ютерного зору.

З психологічної точки зору відомо, що всі людські емоції можна класифікувати на 6 базових категорій: гнів, відраза, страх, радість, смуток, здивування, а також нейтральний стан. Даний проект є спробою навчити комп'ютер аналізувати міміку людини та класифікувати емоції в одну з цих семи категорій.

1.1.2 Штучний інтелект

Штучний інтелект (AI) є здатністю системи обробляти, використовувати та удосконалювати набуті знання та навички. Зараз найбільш ефективно застосування штучного інтелекту спостерігається в таких галузях:

– електронна комерція (онлайн продажі): інструменти ШІ допомагають створювати індивідуалізовані пропозиції на основі історії переглядів та вподобань, що сприяє зміцненню відносин з клієнтами та збільшенню їхньої вірності до бренду;

– навігаційна сфера: в даний момент ШІ активно використовується Uber та численними логістичними компаніями для підвищення операційної ефективності, аналізу трафіку та оптимізації маршрутів;

– робототехніка: сучасні роботи, хоч і не обдаровані інтелектом, подібним до людського, все ж виконують різноманітні функції та завдання, вони наділені здатністю «думати». Роботи виконують функції збирання транспортних засобів Tesla на заводі та можна навчати студентів з Японії англійської мови. За допомогою роботів здійснюють переміщення лікарняних вантажів, фабричних та складських вантажів, прибирають офісні приміщення та очищають велике обладнання, серед іншого. У майбутньому вони впораються з ще більшою кількістю завдань. Логічним розвитком робототехніки є вдосконалення емоційного інтелекту щодо людини та спроба наділити роботів емпатією.

1.2 Загальний огляд проблем в задачі розпізнавання емоцій людини

Упродовж останніх років, науковці, які працюють над розпізнаванням емоцій людини, стикалися з рядом проблем, які ускладнюють створення успішної моделі.

1.2.1 Проблема набору даних

Основною проблемою, з якою зіштовхуються фахівці з обробки даних, є відсутність якісного набору даних. Набір даних повинен відповідати наступним критеріям якості для будь-якої задачі машинного або глибокого навчання:

- зображення найвищої якості: прикладні зображення, які використовують для «навчання» моделей повинні бути максимальної якості, чіткими та з високим розширенням для кращого аналізу та розпізнання;

- великий вміст набору даних: найкращі моделі, що існують на сьогодні, були натреновані на великому обсязі даних (датасети, які містять понад 150 тисяч зображень);

- варіативність набору даних: для ефективного тренування моделі важливо використовувати набір даних з різноманітними характеристиками. Це дозволяє моделі навчитися розпізнавати та адаптуватися до різних ситуацій, забезпечуючи готовність до обробки нових даних у майбутньому. Різноманітність може включати змінні параметри, такі як різні контексти фотографій, кути зору, різноманітність національностей людей на зображеннях, їх вік та наявність інших об'єктів;

- системність набору даних: при аналізі проблеми класифікації зображень важливо врахувати не лише різноманітність даних, але й їх збалансованість. Це означає, що кількість зображень, що належать до кожного класу, повинна бути приблизно однаковою, забезпечуючи адекватне представлення кожного класу в наборі даних. Такий підхід допомагає уникнути перекосів у навчанні моделі та забезпечує більш об'єктивні результати класифікації

Правильний поділ даних на навчальну, варіаційну та тестову вибірку: під час тренування моделей зображень стандартно використовують три різні набори даних: навчальний, варіаційний та

тестовий. Навчальний набір використовується безпосередньо під час навчання моделі, варіаційний служить для налаштування гіперпараметрів та перевірки її ефективності на різних етапах, а тестовий – для оцінки її точності на нових даних. Зважаючи на велику важливість різноманітності та збалансованості даних у цих наборах, формування їх відповідно до цих критеріїв є викликом, оскільки вони впливають на ефективність та адаптивність моделі до різноманітних сценаріїв.

1.2.2 Похибки у розпізнанні недостатньо виражених емоцій

Більшість наборів даних, розроблених для вирішення завдання щодо здійснення розпізнавання штучних емоцій на фотографіях людей. В таких випадках людина намагається максимально передати необхідну емоцію у повному обсязі (рисунок 1.1). Щодо реального життя, то люди рідко проявляють максимальний спектр власних емоцій (рисунок 1.2).



Рисунок 1.1 – Програмна модель вираження емоцій



Рисунок 1.2 – Приклад посмішки в реальному житті

Як можна побачити, ця різниця є досить значною, що ускладнює наше завдання.

1.2.3 Відмінність західної та східної культури

Східна культура охоплює вірування, звичаї та традиції, що відрізняють людей східної частини світу, включаючи Далекий Схід, Західну, Середню, Північну та Південну Азію. Заснована на різних релігійних системах, таких як іслам, буддизм, конфуціанство, індуїзм, дзен, даосизм, східна культура акцентує увагу на духовних аспектах, досліджуючи внутрішній світ людини та віддаючи перевагу ідеї нескінченної циклічної природи Всесвіту. Заснована на практиці медитації та розвитку внутрішнього потенціалу, східна культура підтримує контроль над емоціями та душевним здоров'ям. Колективізм та активна участь у спільноті є ключовими цінностями східної культури, яка підкреслює важливість соціальної взаємодії та відчуття спільності серед людей.

Західна культура визначається як спадщина традицій, етичних цінностей, систем вірувань, звичаїв, артефактів. Технологій, що впливають на життя, переконання, ідеї людей із західної частини світу. Її корені

сягають до різних етнічних і мовних груп, зокрема кельтських, германських, еллінських, латинських, єврейських та слов'янських, і починаються в Європі. Заснована в основному на християнстві, західна культура сприймає людину як частину божественного порядку і життя у служінні Богу. Від Стародавнього Риму до Стародавньої Греції та до сучасних часів, вона продовжується на фоні християнства, впливала на дослідження Просвітництва та відкриття науки, а також поширювалася у всьому світі внаслідок міграції, глобалізації людей у проміжок часу між 16-м та 20-м століттями.

Справді, однією з основних проблем сучасності у розпізнанні емоцій є різниця у вираженні їх між західною, східною культурами. У таких культурах, як західна, люди часто відкрито виражають свої емоції, не приховуючи їх (рисунок 1.3). Для східної культури притаманними є стриманість в емоціях та мінімалізм (рисунок 1.4).



Рисунок 1.3 – Європейська модель

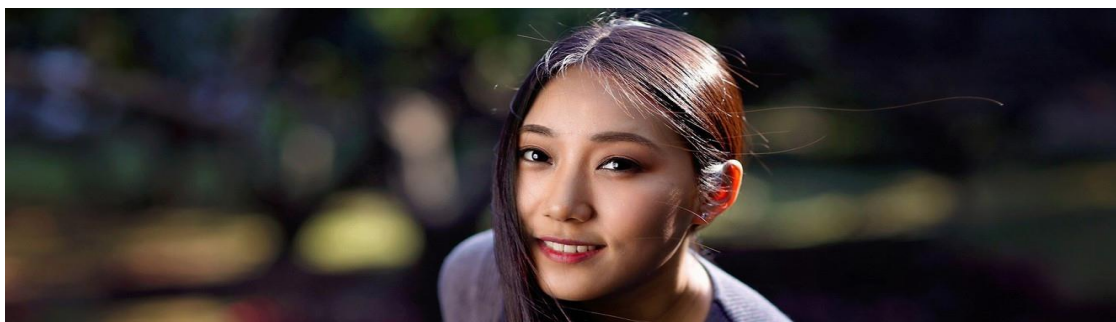


Рисунок 1.4 – Східна модель посмішки

Отже, як ми бачимо, люди зі Сходу та Заходу виражають однакові емоції зовсім по-різному. Це призводить до того, що математичні моделі можуть помилково сприймати посмішку як нейтральну емоцію, і навпаки.

1.3 Висновки до розділу

Отже, в першому розділі даної роботи проводилось вивчення інших проектів, спрямованих на розпізнавання емоцій, а також аналіз проблем, які вони зазнали під час своєї реалізації. Основна увага була зосереджена на ретельному дослідженні вже існуючих підходів та технічних рішень у цій галузі. Головними проблемами, які були ідентифіковані під час цього аналізу, були наступні:

- проблема набору даних: виявлено, що доступні набори даних для розпізнавання емоцій можуть бути недостатньо об'ємними або непридатними за якістю. Це ускладнює навчання моделей розпізнавання емоцій та може призвести до неадекватної роботи алгоритмів;

- неповністю виражені емоції: зазначено, що в реальних умовах емоції часто виражаються не повністю або неоднозначно, що ускладнює їх розпізнавання комп'ютерними системами. Ця проблема може виникнути через різноманіття індивідуальних виразів обличчя, поставів тіла та інших факторів;

- різниця між західною і східною культурами: виявлено, що способи вираження емоцій можуть суттєво відрізнятися в західних і східних культурах, що може призвести до помилкового розпізнавання емоцій комп'ютерними системами, які базуються на стандартах західної культури.

Аналіз цих проблем є важливим кроком для подальшої розробки ефективних та надійних систем розпізнавання емоцій, оскільки він дозволяє зазначити ключові виклики та недоліки існуючих підходів і шляхи їх подолання.

2 АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ, АЛГОРИТМІВ ТА ДАТАСЕТІВ

2.1 Існуючі рішення

На сучасному ринку існує різноманіття рішень для розпізнавання емоцій, які варіюються за методологією, технологіями та застосуванням.

2.1.1 EmoDetect



Рисунок 2.1 – Логотип продукту EmoDetect

Відповідно до інформації, наданої на веб-сайті [1], ця програма включена до Єдиного реєстру програм для баз даних та обчислювальних машин. Цей продукт має кілька переваг, включаючи:

- використання системи кодування лицевих рухів (FACS Action Units) П. Екмана для розрахунку та класифікації рухових одиниць;
- класифікація емоцій за допомогою трьох незалежних класифікаторів: нейронної мережі, рішень на основі правил та вагової суми ознак;
- способи визначення різних виразів обличчя, наприклад нейтрального, розпізнавання різних людських емоцій, таких як, радість, сум, здивування, злість, страх, огида;

– адаптивність алгоритму до специфічних особливостей кожної людини;

– систематизація динамічних змін емоційної інтенсивності користувача з плином часу у графічному відображенні;

– легка інтеграція з модулем через API для максимальної зручності.

Проте під час тестування цього продукту виявлено помітну проблему: відносно низька точність розпізнавання емоцій.

2.1.2 Хеома

Подібно до попереднього продукту, програма Хеома (рисунок 2.2) також аналізує фотографії та класифікує сім базових емоційних станів людини: здивування, радість, сум, страх, злість, огида [2].



Рисунок 2.2 – Логотип Хеома

Кожна з цих емоцій може проявлятися на обличчі з різною інтенсивністю, і для кожної емоції програма присвоює відсоткове значення від нуля до ста, відображаючи ступінь її вираження (рисунок 2.3).

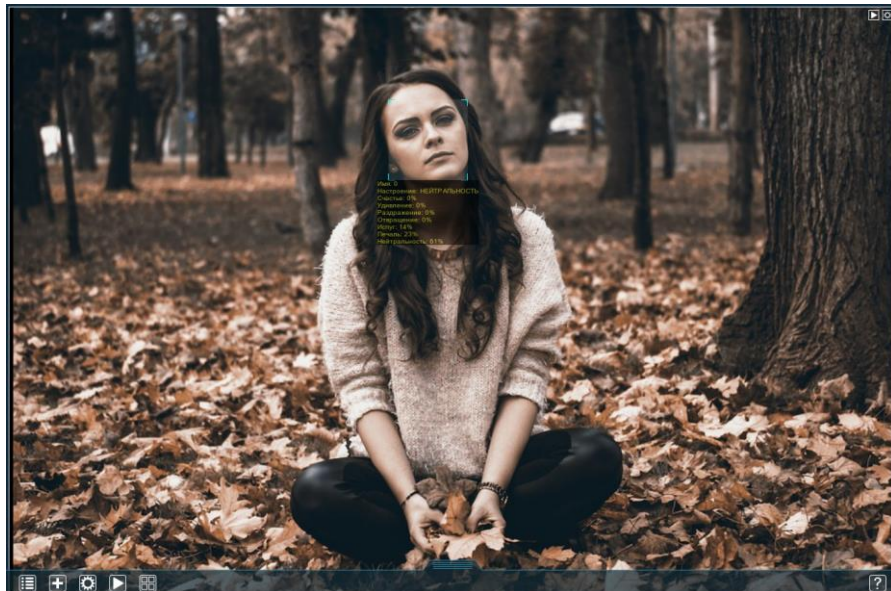


Рисунок 2.3 – Приклад роботи програми Хеона

Ось кілька переваг даної програми:

- гнучкість у роботі з обличчям: обличчя не обов'язково має бути в анфас, що дозволяє робити аналіз навіть при різних кутах зйомки;
- різноманітні тарифні плани: програма пропонує різні варіанти використання, що дозволяє користувачам обирати підходящий тарифний план відповідно до їхніх потреб;
- якісно продуманий інтерфейс: програма має зручний та легкий у використанні інтерфейс, що спрощує процес роботи з нею;
- налаштування сценаріїв розпізнавання емоцій: користувач може налаштовувати різноманітні сценарії розпізнавання емоцій відповідно до вимог та потреб свого проекту;
- можливість збереження даних у CSV-файл: програма дозволяє зручно зберігати дані у форматі CSV, що полегшує подальшу обробку та аналіз результатів.

До недоліків програми відноситься:

- відсутність інформації щодо точності моделі: програма не надає достатньої інформації про точність своєї моделі, що ускладнює оцінку її ефективності та надійності;

– не можливість розпізнавання обличчя з нахилом камери більше 15 градусів: обличчя, які зображені на фотографіях під кутом більше 15 градусів, не розпізнаються програмою, що може обмежити її використання в деяких сценаріях та ускладнити її застосування у реальних умовах.

2.1.3 FaceReader (Noldus)

Алгоритм роботи FaceReader складається з таких етапів:

- виявлення осіб: використовує алгоритм глибокого навчання для виявлення осіб на зображеннях;
- моделювання обличчя: створює детальну штучну модель обличчя, використовуючи набір майже п'ятсот ключових точок, що дозволяє здійснювати точні аналізи та розпізнавати різні вирази обличчя;
- класифікація осіб: застосовує штучні нейронні мережі для класифікації виразів обличчя та емоцій, визначаючи їх типи та інтенсивність.

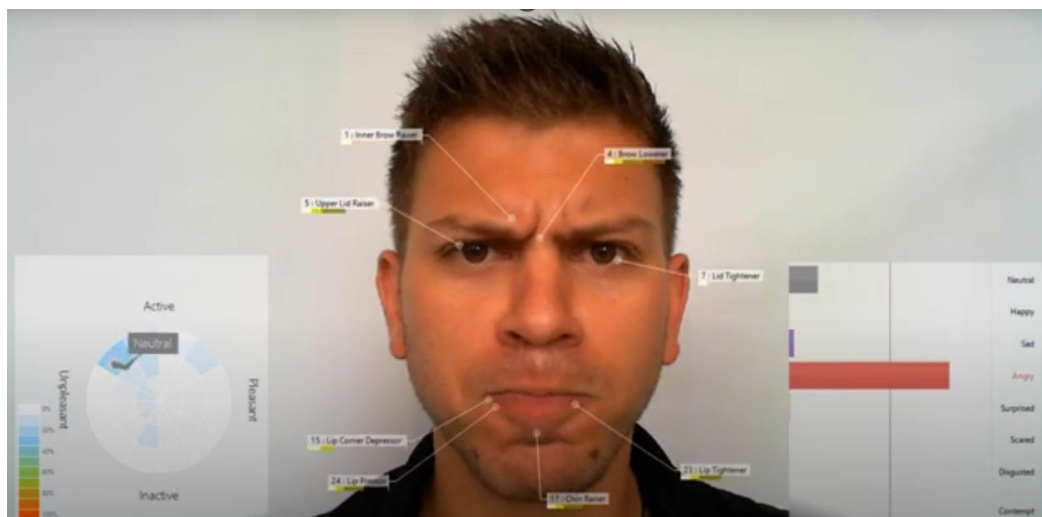


Рисунок 2.4 – Приклад роботи програми FaceReader

Даний алгоритм виявляє різноманітні параметри, що включають базові вирази обличчя, вислови, які виражають настрій, орієнтацію голови

та напрямок погляду. Він також оцінює валентність та рівень збудження, а також зафіксовані дії та реакції. Крім того, він аналізує частоту скорочень серця та і варіабельність серцевого ритму для подальшої реєстрації патернів у поведінці споживача.

Переваги цього алгоритму включають:

- зручний інтерфейс, що сприяє комфортному використанню;
- багатий функціонал, що дозволяє проводити різноманітний аналіз та отримувати різні типи даних;
- високий відсоток правильно розпізнаних емоцій, що гарантує надійні результати;
- можливість використання як у режимі реального часу, так і з обробкою завантажених відео.

Проте для даного продукту характерні наступні недоліки:

- низький відсоток правильного розпізнавання емоцій у дітей;
- висока ціна, що може бути обтяжливим фактором для деяких користувачів.

2.1.4 Microsoft Oxford Project Emotion Recognition

Microsoft Oxford Project Emotion Recognition [3] спрямований на аналіз емоцій, відкриває можливості для створення спеціальних систем, що спрямовані на розпізнавання різних емоційних станів, наприклад, гнів, страх, зневагу, огиду, щастя, смуток, здивування, нейтральність на основі типового виразу обличчя, що відображає ці емоції (рисунок 2.5).

Бета-версія цього продукту включає такі функції:

- розпізнавання особи за голосом, щоб вивчити її унікальні риси;
- інтеграція сервісу в власні розробки;
- привабливі тарифи та можливість безкоштовного тестового періоду.

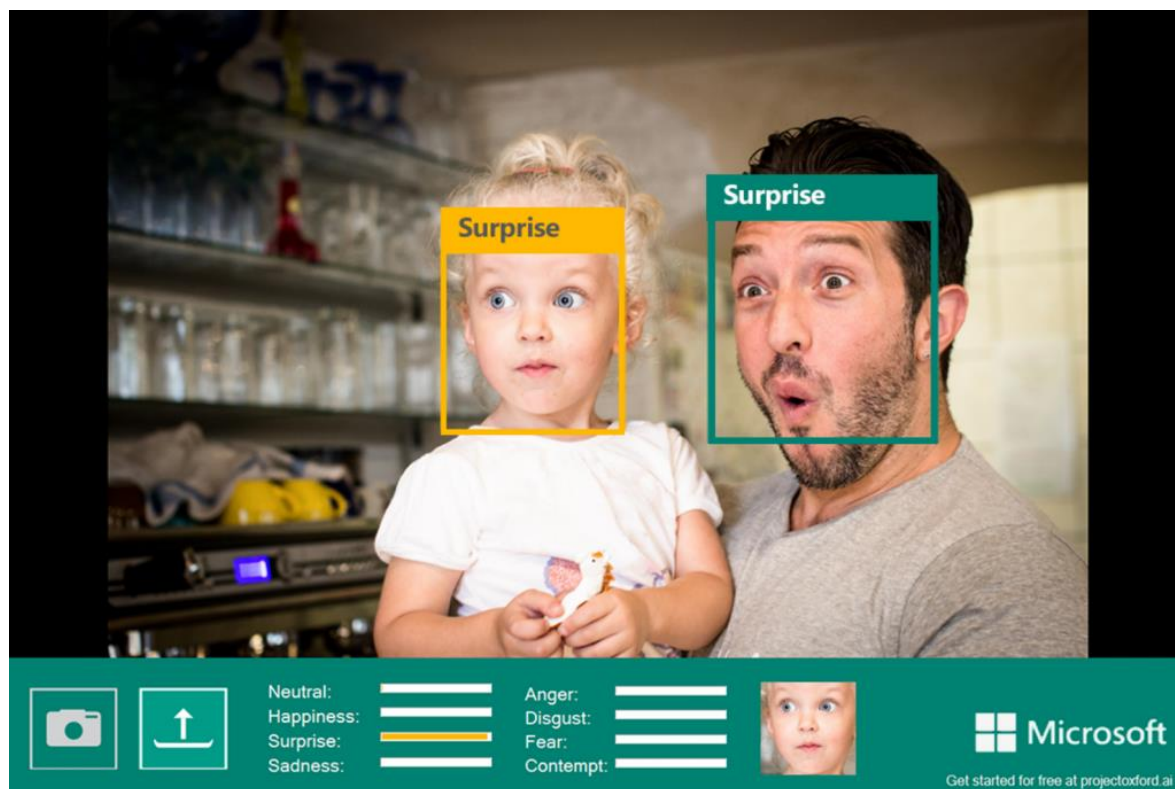


Рисунок 2.5 – Приклад роботи програми Microsoft Oxford Project Emotion Recognition

Проте, серед недоліків системи можна відзначити:

- система не працює належним чином з фотографіями, на яких обличчя людей знаходяться під значним кутом, що призводить до недостатньо точного аналізу;
- висока ймовірність низької точності розпізнавання емоцій;
- обмеженість системи у здатності аналізувати та розпізнавати лише статичні зображення, що може обмежувати її застосування в реальному часі.

2.1.5 eMotion Software

До функцій eMotion Software [4] належить створення тривимірної моделі обличчя, під час цього здійснюється поєднання основних областей

обличчя губи, брови та очі. Однією з особливостей є те, що компанія Unilever колись інтегрувала його та використала для збільшення продажей автоматом з морозивом. Ця ідея полягала в тому, що люди посміхаючись автомату, отримували безкоштовне морозиво.

Переваги вищезазначеного об'єкта:

- наявність розпізнавального механізму базових емоцій людини;
- пробна (безкоштовна) версія;
- простий для використання інтерфейс.

Недоліками програми можна вважати загальну застарілість технологій та низьку точність розпізнавання.

2.2 Допоміжні алгоритми в задачі розпізнавання емоцій

2.2.1 Гістограма направлених градієнтів (HOG)

Цей алгоритм являє собою дескриптор ознак, який використовується у комп'ютерному зорі та обробці зображень для розпізнавання об'єктів. Метод аналізує напрямки градієнтів у локальних точках зображення. Він схожий на гістограму орієнтованих границь та SIFT дескриптор, а також враховує значення форми. Однак відрізняється тим, що обчислення проводиться у тісній сітці та застосовується локальна нормалізація контрасту.

Дослідження Далала та Тріггса виявило, що HOG дескриптор має кілька ключових переваг у порівнянні з іншими дескрипторами. По-перше, він працює з локальними клітинами, що робить його нечутливим до геометричних та фотометричних змін, за винятком орієнтації об'єкта. Це дозволяє використовувати його для аналізу областей різних розмірів. З цих причин HOG особливо ефективний у виявленні людей на зображеннях.

Алгоритм HOG працює покроково і включає наступні частини:

– обчислення градієнту: один із простіших методів – використання 1-D різницевої маски у горизонтальному та вертикальному напрямках. Проте, у своїй роботі Далал та Тріггс використовували більш складні маски, зокрема, 3x3 Оператор Собеля. Вони також провели експерименти з застосуванням Розмиття Гауса до різницевих масок. Виявилося, що на практиці алгоритм працює краще без розмиття;

– групування напрямків: у наступному кроці формується набір гістограм напрямлених градієнтів. Кожен піксель у комірці додає свій внесок у гістограму, враховуючи величину та напрямок градієнту. Комірки можуть мати різну форму – прямокутну або круглу, і кількість каналів гістограми може розділятися рівномірно між 0 і 180 або між 0 і 360, залежно від того, чи обчислюється беззнаковий чи знаковий градієнт. Дослідники Далал та Тріггс встановили, що в їхній роботі з виявленням людей найкращі результати демонструє беззнаковий градієнт та гістограми з 9 каналами. Для визначення ваги пікселя може використовуватися або величина градієнта, або функція її визначення;

– дескриптори блоків – це метод, який використовується для опису патернів чи особливостей у конкретних областях зображення. Вони розглядають області зображення як блоки та використовують специфічні характеристики чи властивості цих блоків для створення опису, який може бути використаний для подальшого аналізу чи визначення подібності між зображеннями. Дескриптори блоків можуть включати в себе такі характеристики, як градієнти, текстурні особливості, або інші структурні деталі;

– нормування блоків – це процес стандартизації розмірів чи характеристик блоків у зображенні з метою забезпечення однакового масштабу чи діапазону значень. Це важлива операція в обробці зображень та комп'ютерному зорі, оскільки дозволяє забезпечити однакові умови для аналізу різних областей зображення. Нормування блоків може включати в

себе процеси, такі як зсув, масштабування, або зміну контрасту для досягнення бажаних характеристик у блоках зображення;

– метод опорних векторів: завершальним етапом у процесі визначення образів з використанням гістограми напрямлених градієнтів є введення дескрипторів у систему класифікації, засновану на навчанні з учителем. Метод опорних векторів (SVM) є класифікатором, який визначає оптимальну гіперплощину як вирішальну функцію. Після тренування класифікатора він може приймати рішення щодо присутності об'єкта, наприклад, людини, на додаткових тестових зображеннях.

2.2.2 Локальні бінарні шаблони (LBP)

Цей елементарний оператор, що використовується для аналізу текстур у галузі комп'ютерного зору, був вперше описаний у 1994 році. Локальні бінарні шаблони (ЛБШ) представляють собою опис границі пікселя зображення у вигляді двійкової форми. Пізніше виявилось, що в поєднанні з гістограмою напрямлених градієнтів ЛБШ значно підвищує продуктивність розпізнавання на певних наборах даних. У 2015 році було проведено порівняльний аналіз кількох модифікацій оригінального ЛБШ у контексті відокремлення фону.

Алгоритм дій:

– розділити вікно пошуку на осередки рівні комірки (наприклад, 16x16 пікселів для кожної комірки);

– для кожного пікселя у комірці необхідно порівняти його з кожним з його восьми сусідів, рухаючись за годинниковою або проти годинниковою стрілкою;

– якщо значення центрального пікселя перевищує значення будь-якого з його сусідів, тоді центральному пікселю присвоюється значення «0», в іншому випадку – «1». Таким чином, отримуємо восьми розрядне

двійкове число, яке зазвичай перетворюється у ціле десяткове число для подальшого аналізу;

- розраховується гистограма в комірці з урахуванням частоти появи кожного «числа» (тобто всіх комбінацій пікселів, які більші або менші за центральний). Гистограма може бути представлена у вигляді 256-розмірного вектора ознак;

- виконується нормалізація гистограми (необов'язково);

- нормалізовані гистограми всіх комірок. Це дозволяє отримати вектор ознак для всього вікна. На цій стадії, алгоритм завершує своє плінне танцювання. Отриманий вектор ознак лежить перед нами як неосяжна картина, готовий для подальшого творчого втілення за допомогою методів машинного навчання, таких як SVM, Extreme Learning Machine та інші мистецькі алгоритми класифікації зображень. Ці унікальні класифікатори можуть розкривати свою естетичну магію у системах розпізнавання обличчя або у глибокому та непередбачуваному аналізі текстур. А нижче розкривається власна схема роботи цього алгоритму (рисунок 3.1).

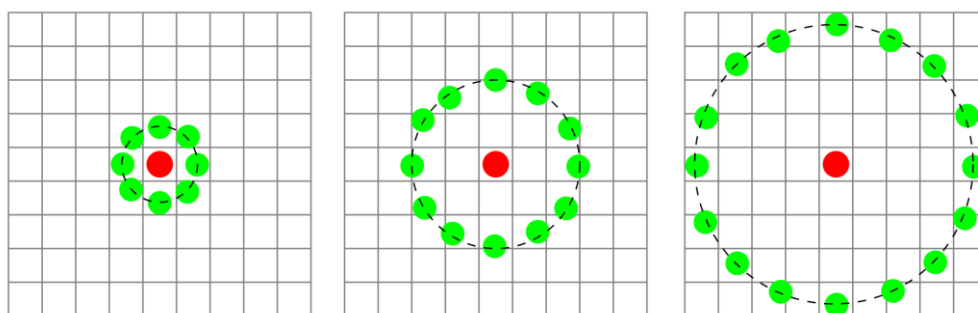


Рисунок 2.6 – Текстура та обчислення локального бінарного шаблону (ЛБШ)

2.2.3 Метод Віоли-Джонса

Цей алгоритм, розроблений Полом Віолою та Майклом Джонсом у 2001 році, призначений для виявлення об'єктів на зображеннях у реальному часі. Хоча його основною метою є розпізнавання обличчя, але його функціональність розширюється на виявлення різних об'єктів.

Принцип роботи алгоритму полягає у наступному:

- використання зображень у формі інтегралів для швидкого обчислення потрібних об'єктів;
- застосування ознак Хаара для пошуку відповідних об'єктів, зокрема осіб та їхніх рис;
- використання методу бустингу для вибору найбільш підходящих ознак для об'єкта, що розглядається на даному фрагменті зображення;
- подання всіх ознак на вхід класифікатора, який робить висновок про наявність об'єкта або його відсутність;
- використання каскадів ознак для швидкого відкидання вікон, в яких не знайдено обличчя.

Цікавим відзначити, що навчання класифікаторів для даного алгоритму є довгим та затратним процесом, але сам процес пошуку обличчя відбувається дуже швидко. Цей фактор був вирішальним у включенні алгоритму в різноманітні дослідження, де ключовим було виявлення людини на зображенні.

Крім того, цей детектор характеризується дуже низькою ймовірністю помилкового виявлення обличчя. Він ефективно розпізнає риси обличчя під невеликим кутом, до приблизно 30 градусів. Однак, при нахилі більше 30 градусів, відсоток виявлень різко зменшується. Це обмеження не дозволяє алгоритму ефективно детектувати обличчя людини, яка дивиться під довільним кутом, що ускладнює його застосування в сучасних системах з урахуванням зростаючих потреб.

Давайте розглянемо цей метод більш детально. Алгоритм узагальнено шукає обличчя та риси особи за допомогою принципу сканування вікна.

Узагальнено, метою виявлення обличь та їх рис на цифровому зображенні є наступне:

– вхідним об'єктом є зображення, на якому потрібно виявити об'єкти.

Зображення подається у вигляді двовимірної матриці пікселів розміром $w \cdot h$, де кожен піксель має своє значення:

– від 0 до 255 для чорно-білих зображень;

– від 0 до 2553 для кольорових зображень (компоненти R, G, B).

Під час виконання завдання алгоритм має виявити особи та їх характеристики, а потім відзначити їх на зображенні. Для досягнення цього пошук виконується в активній зоні зображення за допомогою прямокутних ознак, які описують виявлені особи та їх характеристики (рисунок 2.7). Кожен прямокутник має наступні параметри: координати центру (x , y), ширину (w), висоту (h) та кут нахилу (α) відносно вертикальної осі зображення.

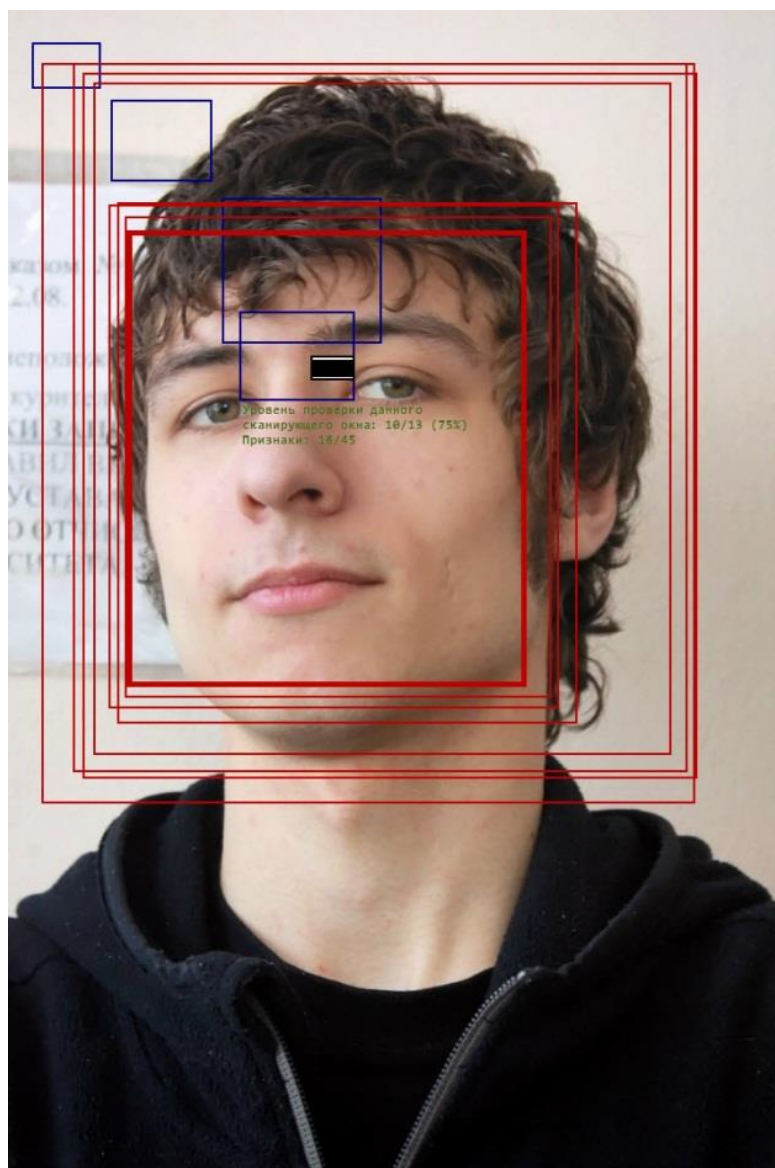


Рисунок 2.7 – Приклад роботи алгоритму Віюлі-Джонса

2.2.4 Метод гнучкого порівняння на графах

Основна ідея цього методу полягає у порівнянні графів, які відображають особливості обличчя особи на зображенні. Деякі дослідження зазначають ефективність розпізнавання на рівні 95–97%, навіть якщо на зображенні присутні різні емоційні вирази або коливання ракурсу до 15 градусів. Однак для порівняння вхідного зображення особи з 87 стандартними еталонами вимагається близько 25 секунд при використанні паралельного обчислювального обладнання. Наприклад структур графів для розпізнавання осіб, такі як регулярна решітка, граф на

основі антропометричних точок обличчя та деформації графа у формі регулярної решітки, представлені на рисунку 2.8.

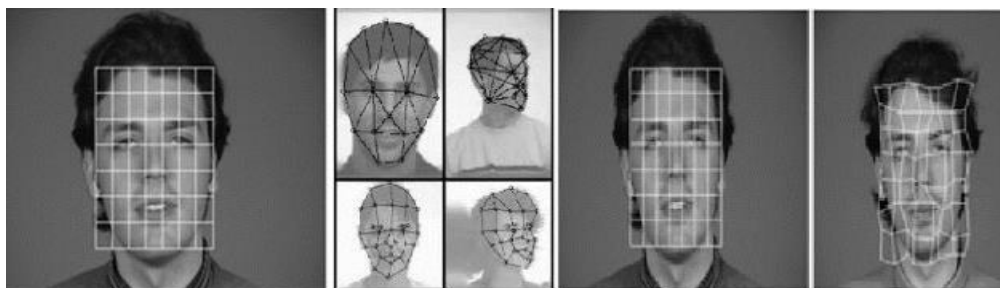


Рисунок 2.8 – Ілюстрація структури графа для ідентифікації осіб може бути такою: а) розташування точок у вигляді регулярної решітки; б) граф, побудований на основі антропометричних точок обличчя; в) граф, який уявлено у формі регулярної решітки з врахуванням його деформаційних характеристик

Один з недоліків цього підходу полягає у його обмеженій ефективності при збереженні нових еталонів. Незважаючи на це, головною перевагою є мінімальна чутливість до рівня освітленості обличчя та зміни кута зору. Однак саме цей метод виявляє меншу надійність розпізнавання порівняно з технологіями, що використовують нейромережі.

2.2.5 Геометричний метод розпізнавання обличчя

Цей алгоритм є одним з перших методів, які застосовуються для розпізнавання обличчя. У таких методах визначається набір ключових точок або областей на обличчі особи, і потім створюється набір ознак. Ключовими точками можуть бути куточки очей, губ, кінчик носа, центр ока і так далі. Цей підхід встановлює вимоги до зображень. На рисунку 2.9 наведено приклад побудови геометричних ліній на обличчі.

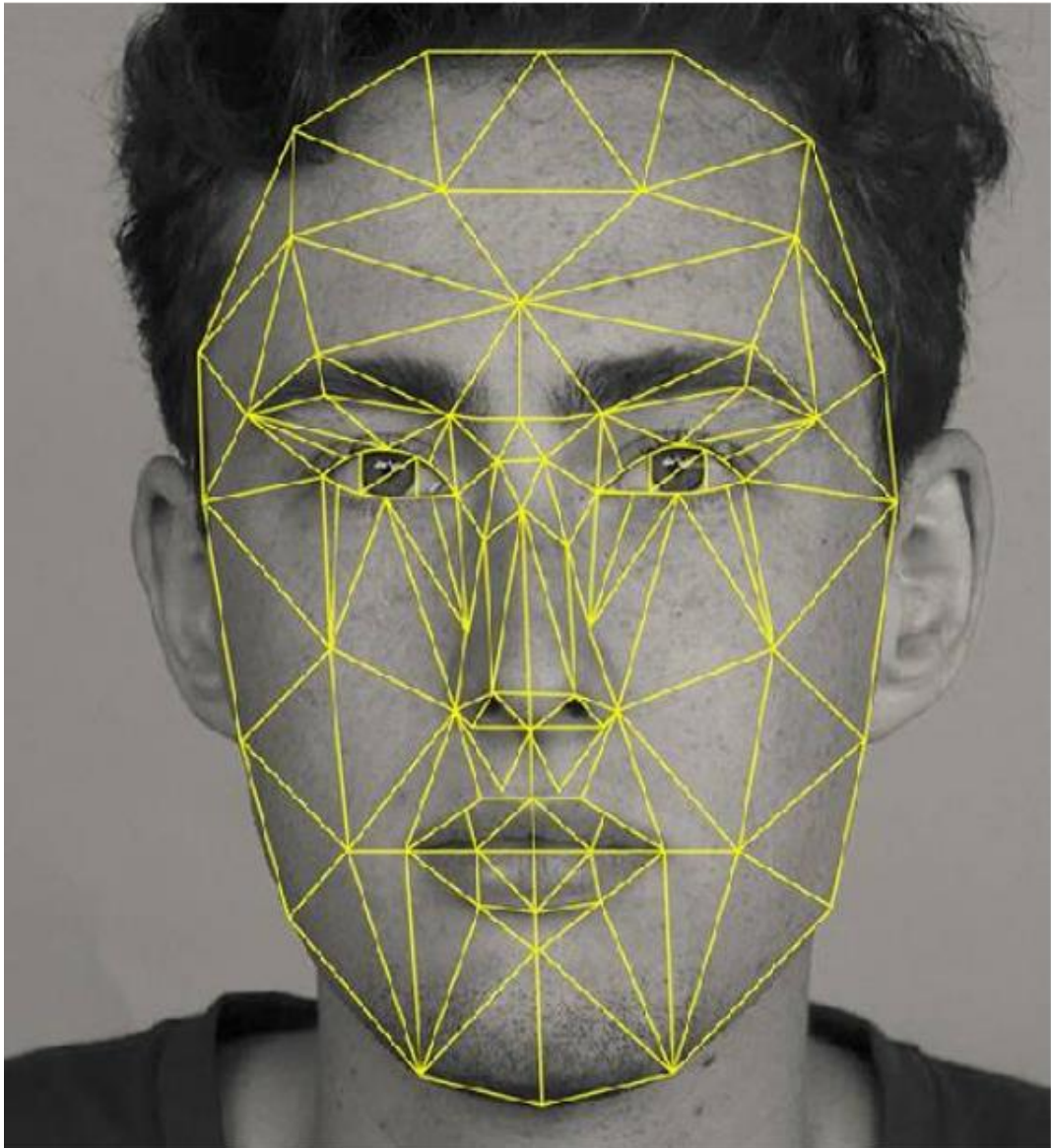


Рисунок 2.9 – Побудова геометричних ліній

Переваги цього методу включають:

- використання доступного та економічного обладнання;
- при належному обладнанні можливе розпізнавання зображень, отриманих з великих відстаней.

Недоліки включають:

- низьку статистичну достовірність;
- високі вимоги до освітлення;
- потребу у фронтальному зображенні особи з мінімальними відхиленнями;

– недостатню увагу до можливих змін у міміці обличчя.

2.3 Традиційні методи розпізнавання емоцій

2.3.1 Розріджена логістична регресія зі штрафом L1/2

Логістична регресія (рисунок 2.10) – це метод статистичної регресії, який застосовується в випадках, коли залежна змінна є бінарною, тобто може приймати лише два значення (0 або 1). Вона знаходить широке застосування у класифікації, встановлюючи порогове значення, яке допомагає визначити, до якої з двох категорій належить об'єкт.

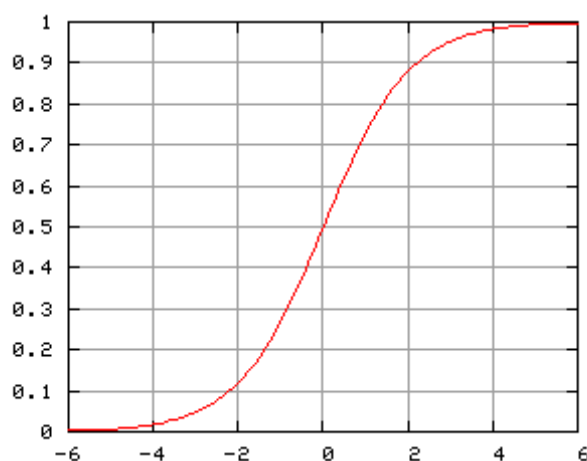


Рисунок 2.10 – Логістична функція

У випадку розпізнавання емоцій, логістична регресія використовується для прогнозування ймовірності наявності певної емоції на основі вхідних даних. Наприклад, модель може використовувати певні ознаки обличчя або голосу як вхідні дані і передбачати ймовірність того, що особа виражає певну емоцію, таку як радість, смуток або здивування. Це дозволяє визначити, наскільки імовірно виникнення кожної емоції для конкретної ситуації чи об'єкта.

2.3.2 Метод К-найближчих сусідів (KNN)

Під час вирішення проблеми класифікації, алгоритм KNN [5] передбачає такі етапи:

- а) завантаження даних для подальшого аналізу;
- б) ініціалізація параметру K на вибрану кількість сусідів;
- в) для кожного прикладу у наборі даних:

- обчислення відстані між прикладом запиту та поточним прикладом за даними;

- додавання відстані та індексу прикладу до впорядкованої колекції;

- сортування впорядкованої колекції відстаней та індексів у порядку зростання за відстанями;

- вибір перших K записів з відсортованої колекції;

- отримання міток для обраних K записів;

- повернення режиму міток K .

Важливим елементом цього алгоритму є правильний вибір значення K (рисунок 2.11). Для досягнення цієї мети запускають алгоритм KNN кілька разів з різними значеннями K і вибирають те, яке зменшує кількість помилок, зберігаючи точність прогнозування при наданні даних, які раніше були використані.

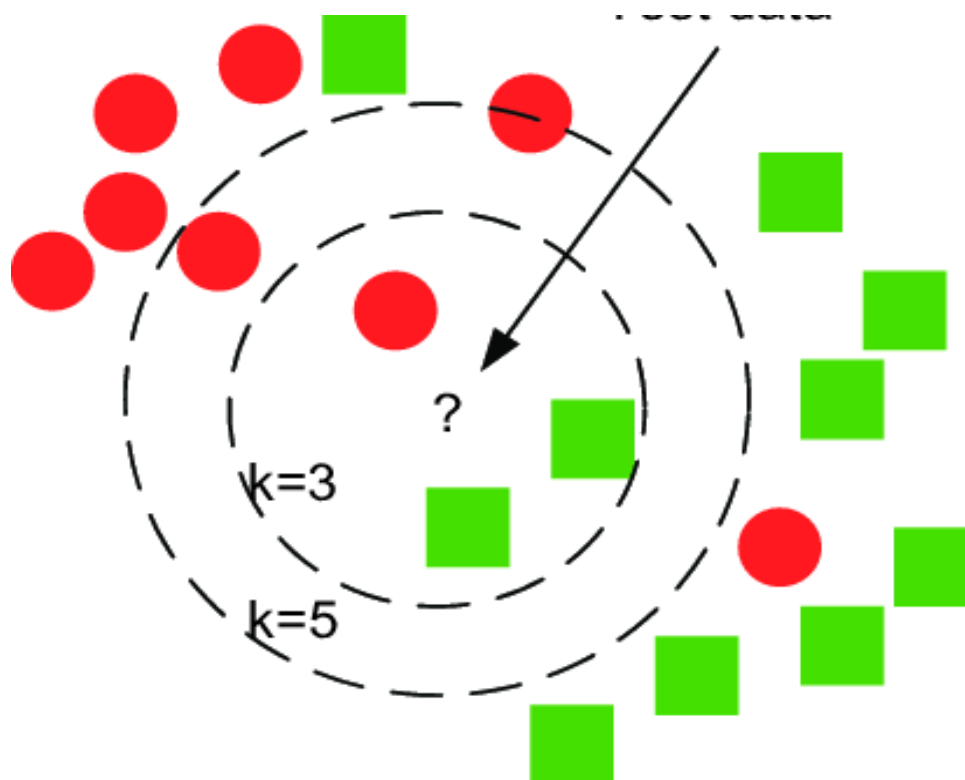


Рисунок 2.11 – Вибір параметру K в алгоритмі KNN

Ось декілька важливих моментів, які варто враховувати при виборі значення K :

– при зменшенні значення K до 1 наші прогнози стають менш стійкими. Уявіть, що $K=1$, і у нас є точка запиту, оточена кількома червоними і одним зеленим пунктом. Навіть якщо більшість точок червоні, KNN може помилково прогнозувати, що точка запиту зелена через єдиний найближчий зелений пункт. Можна припустити, що точка запиту, ймовірно, має червоний колір, але через те, що значення K дорівнює 1, метод найближчих сусідів (KNN) невірно передбачає, що точка запиту має зелений колір;

– навпаки, при збільшенні значення K прогнози стають більш стійкими завдяки голосуванню/середньому значенню більшості, що зазвичай призводить до більш точних прогнозів (до певної межі). Проте зі збільшенням K також зростає ймовірність помилок. Важливо зупинитися вчасно, щоб уникнути занадто великого значення K ;

– у випадках, коли ми використовуємо голосування більшості (наприклад, у задачі класифікації), зазвичай варто обирати непарне значення K , щоб мати можливість визначити вибір більшості.

При класифікації за допомогою методу K -найближчих сусідів (KNN) алгоритм спирається на голосування більшості для визначення, чи належить спостереження до певного K -подібного екземпляра. Для цього використовується евклідова відстань у вимірному просторі. Оскільки KNN є непараметричним методом класифікації, він присвоює мітки раніше не вибірковим точкам, що часто призводить до зниження ефективності зі збільшенням обсягу даних; тому особливо важливо розглядати алгоритми редукції ознак. Ефективність повністю залежить від значення K , тому модель була розроблена для оцінки різних значень K шляхом ітеративної перехресної перевірки від 2 до 50. Цей підхід був запропонований, оскільки вибір правильного значення K може бути складним теоретично без використання вичерпного пошуку за допомогою витратних методів.

Переваги даного алгоритму:

- простота та легкість реалізації;
- не потребує побудови моделі, налаштування параметрів або додаткових припущень;
- універсальність. Може використовуватися для класифікації, регресії та пошуку (як буде розглянуто в наступному розділі).

Недоліки даного алгоритму:

- збільшення кількості прикладів або предикторів/незалежних змінних призводить до помітного сповільнення алгоритму.

2.3.3 Метод опорних векторів (SVM)

SVM Метод опорних векторів – це основні моделі лінійної класифікації, гіперплощини для розділення класів даних. При застосуванні до вибірки даних з каналів ЕЕГ, зразки проектується у простір, де числові

осі представляють різні ознаки. Категорії даних розділяються настільки широко, наскільки можливо. Під час навчання моделі вектор w та зміщення b оцінюються розв'язанням квадратного рівняння. Таким чином, SVM може бути реалізований за поліноміальний час, що робить його задачею p -повною.

Опорні вектори просто представляють координати окремих спостережень. Класифікатор SVM представляє собою межу, яка найкраще розділяє два класи (гіперплощина/лінія) (рисунок 2.12).

SVM може бути двох типів:

– лінійний SVM застосовується для аналізу даних, які можна розділити за допомогою прямої лінії на два класи. Такі дані класифікуються як лінійно роздільні, і для їхнього аналізу використовується лінійний класифікатор SVM;

– нелінійний SVM використовується для дослідження наборів даних, які не можна розділити за допомогою прямої лінії. Такі дані вважаються нелінійно роздільними, і для їхнього аналізу застосовується нелінійний класифікатор SVM.

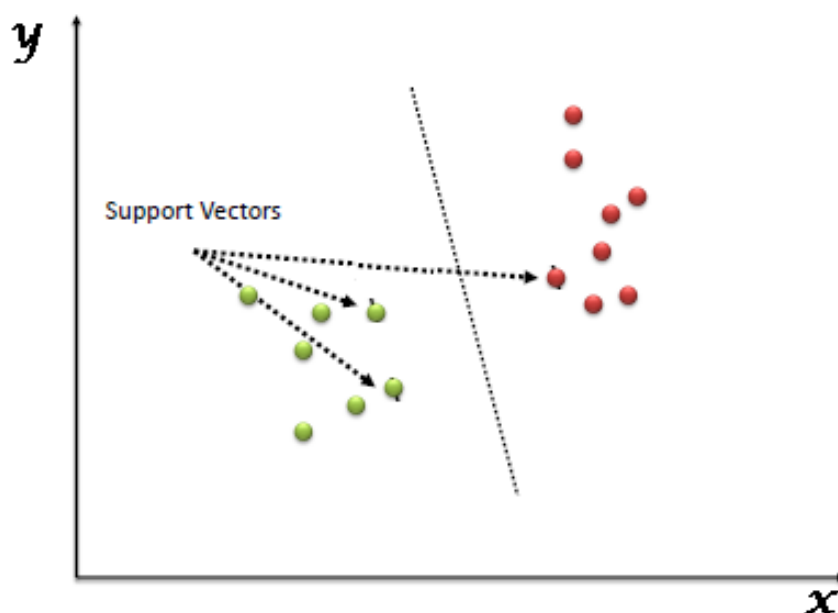


Рисунок 2.12 – Гіперплощина, що розділяє точки двох класів

Давайте поглибимося у деталі роботи кожного із них.

Лінійний SVM: для кращого розуміння алгоритму SVM можна навести приклад. Припустимо, що маємо набір даних з двома класами (зеленим і синім) та двома ознаками x_1 і x_2 (рисунок 2.13). Нам необхідний класифікатор, який зможе віднести пару координат (x_1, x_2) до зеленого або синього кольору.

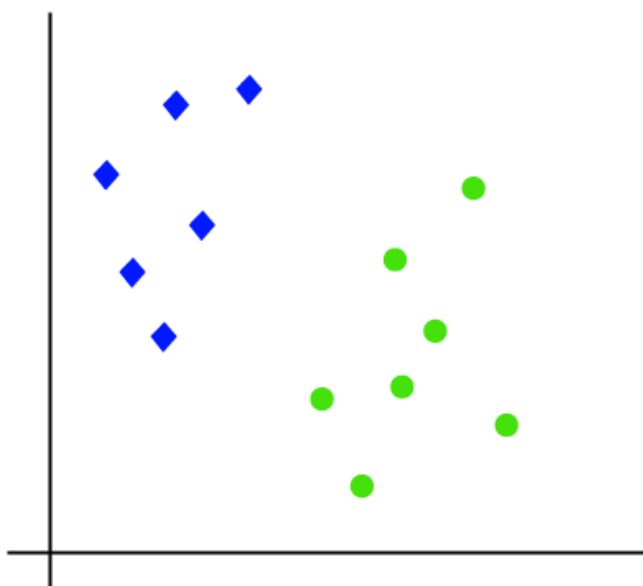


Рисунок 2.13 – Лінійний SVM

Отже, оскільки ми працюємо у двовимірному просторі, можна легко виділити ці два класи, використовуючи просту пряму лінію. Проте може існувати кілька можливих лінійних розділових меж для цих класів.

Алгоритм SVM допомагає знайти оптимальну лінію або площину розділення, яку ми називаємо гіперплощиною. Він робить це шляхом пошуку найближчих точок до лінії з обох класів, які ми називаємо опорними векторами. Відстань між цими векторами і гіперплощиною визначається як «запас», і завданням SVM є максимізувати цей запас. Гіперплощина з максимальним запасом вважається оптимальною.

Нелінійний SVM. У випадку лінійно роздільних даних, ми можемо просто використати пряму лінію для їх розділення (рисунок 2.14). Проте, якщо дані не можна розділити прямою лінією, нам потрібно шукати більш складні шляхи.

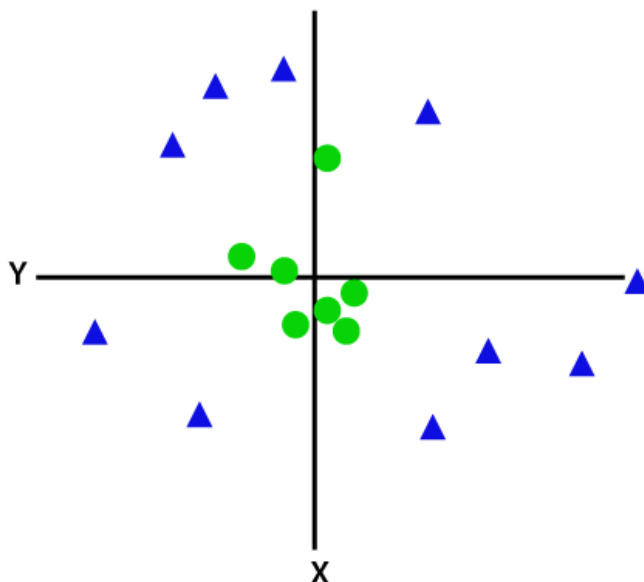


Рисунок 2.14 – Нелінійний SVM

Отже, для розділення цих даних потрібно додатковий вимір. У випадку лінійних даних ми використовували два виміри – x і y . Для нелінійних даних додаємо останній, третій рівень – z . Його можна розрахувати як квадрат суми квадратів x та y ($z=x^2+y^2$).

2.4 Існуючі датасети для розпізнавання емоцій

2.4.1 AffectNet

AffectNet [7] – це база даних виразів обличчя, що складається з фотографій, отриманих з реального життя та анотованих вручну. Набір даних включає понад мільйон зображень обличь, які були зібрані з

Інтернету за допомогою запитів до основних пошукових систем. Використано ключові слова пов'язаних з емоціями, які були представлені у шести різних мовах.

Приблизно половина зібраних зображень (близько 440 тисяч) була анотована вручну для виявлення семи дискретних виразів обличчя (за категоріальною моделлю) та для вимірювання інтенсивності валентності та збудження (за розмірною моделлю). AffectNet вважається найбільшою базою даних виразів обличчя, валентності та збудження, зібраною у реальних умовах, що надає можливість проведення досліджень щодо автоматичного розпізнавання виразів обличчя за допомогою двох різних моделей емоцій.

Наразі тестовий набір даних ще не був випущений. У найближчому майбутньому розробники цього набору даних планують організувати виклик на AffectNet, і тестовий набір буде використаний для оцінки методів та алгоритмів учасників.

2.4.2 Ascertain

Ascertain [8] – представляє собою комплексну базу даних, розроблену для аналізу неявних аспектів особистості та їх впливу, і може використовуватися для вивчення характеристик особистості та емоційних станів на основі фізіологічних реакцій. Набір даних включає інформацію про 5 основних шкал особистості та емоційні самооцінки 58 учасників, а також одночасно записані електрокардіограми (ЕКГ), електроенцефалограми (ЕЕГ), гальванічні реакції шкіри (GSR).

2.4.3 Dreamer

Dreamer [9] – це мультимодальний архів даних, який включає записи сигналів електроенцефалограми (ЕЕГ) та електрокардіограми (ЕКГ),

отримані під час виклику емоційно-афективного стану за допомогою аудіовізуальних стимулів. Інформація була зібрана від двадцяти трьох учасників, які також оцінювали свій афективний стан за кожним стимулом щодо валентності, збудження та домінування.

Усі сигнали були зафіксовані за допомогою доступного за ціною та стандартного портативного обладнання, що надає можливість використання ефективних обчислювальних методів у повсякденних ситуаціях. Для отримання сигналів ЕЕГ використовувалася бездротова гарнітура ЕЕГ Emotiv EPOC, а для ЕКГ – датчик ЕКГ Shimmer2.

Отримані результати класифікації за валентністю, збудженням та домінуванням на основі цієї бази даних можуть бути порівняні з результатами, отриманими для інших баз даних, що використовують пристрої медичного класу, що не є портативними та є дорогими.

Запропонована база даних доступна для загального використання з метою надання дослідникам можливості провести більш докладну оцінку придатності цих пристроїв захоплення для програм розпізнавання афектів.

2.4.4 Extended Cohn-Kanade Dataset (СК+)

База даних Кон-Канаде (СК) [10] з метою сприяння дослідженням автоматичного розпізнавання індивідуальних виразів обличчя була створена у 2000 році. Протягом часу СК стала відомою як один з фундаментальних тестових наборів даних для розробки та оцінки алгоритмів.

Хоча коди AU (Action Units) були добре перевірені, мітки емоцій виявилися неадекватними, оскільки вони відображали запитання, а не реально виявлені вирази обличчя.

Показник ефективності, який міг би оцінити нові алгоритми в порівнянні з існуючими – відсутній.

Відсутність стандартних протоколів для використання загальних баз даних у дослідженнях.

У зв'язку з цим, база даних СК використовувалася як для виявлення Action Units (AU), так і для аналізу емоційних станів, незважаючи на те, що мітки для останніх не були достатньо перевірені.

Співставлення з алгоритмами тестування відсутнє, а використання випадкових підмножин вихідних даних ускладнювало мета-аналіз. З метою вирішення цих та інших проблем, була розроблена розширена база даних Sohn-Kanade (СК+). Кількість послідовностей зросла на 22%, а кількість вивчених осіб — на 27%.

Основні результати використання моделей активного зовнішнього вигляду (AAMs) та класифікатора опорних векторів (SVM) були представлені з використанням перехресної перевірки суб'єктів, без урахування як AU, так і емоційних міток для вхідних даних. Мітки емоцій та Action Units. Розширені дані зображення стали доступними у 2010 році. Кожна послідовність у базі даних була повністю закодована за допомогою Facial Action Coding System (FACS), а емоційні мітки пройшли перегляд і були підтвержені. Крім того, до бази даних були включені послідовності без позицій для різних типів посмішок, а також пов'язані з ними метадані.

2.4.5 Emotic

Анотації в наборі даних EMOTIC (рисунок 2.15) об'єднують два методи представлення емоцій: дискретні категорії та безперервні виміри [11].

Цей набір даних також був анотований на основі реальних життєвих ситуацій. У ньому визначено 26 дискретних категорій:

– спокій: включає гарний настрій та внутрішній спокій, відсутність тривоги, присутність позитивних думок або почуттів, відчуття задоволеності;



Рисунок 2.15 – Приклади розмітки датасету Emotic

- прихильність: охоплює ніжні почуття, любов та турботу;
- повага: відображає почуття поваги та захоплення, а також вдячність;
- очікування: відображає стан очікування на майбутнє, надіючись або готуючись до можливих подій;
- залученість: звертання уваги на щось, поглиблення в чомусь, допитливість, зацікавленість;
- впевненість: відчуття впевненості, переконаність у сприятливому результаті, стимульованість, гордість;
- щастя: відчуття задоволення, насолода або розвага;
- задоволення: насолодження від почуттів;
- хвилювання: ентузіазм, стимулювання, енергія;
- раптове відкриття чогось несподіваного;
- симпатія: співчуття, підтримка, жалість;
- сумнів/розгубленість: складність прийняття рішення, роздуми про різні варіанти;
- відчуття незацікавленості в основній події оточення, байдужість, нудьга, відволікання;
- втома: відчуття втоми, сонливість;
- почуття сорому, збентеження або провини;
- сильне прагнення до чогось, туга, ревності, заздрість, бажання;
- відчуття засудження, несхвалення, презирство, ворожість;
- відраза: огиду, неприязнь, відразу, відчуття ненависті;

- роздратування, нетерплячість, розчарування;
- гнів: сильне незадоволення або гнів, гнів, обурення;
- відчуття фізичного або емоційного поранення, вразливість, чутливість;
- смуток: відчуття нещастя, суму, розчарування або збентеження;
- неспокій: нервозність, тривога, смуток, тривожність, напруженість, стрес;
- страх, підозра, жах;
- фізичні страждання;
- страждання: емоційне або психологічне занепокоєне, почуття скорботи, переживання болю.

У цьому наборі даних також присутні безперервні категорії:

- валентність: цей параметр відображає, наскільки позитивною або приємною є емоція. Він варіюється від негативного до позитивного спектру;
- збудження: цей показник вимірює рівень активності людини. Він коливається від спокійного/неактивного до збудженого/готового до дій;
- домінування: цей показник визначає рівень контролю особи над ситуацією. Він змінюється від підпорядкованого/не впливового до домінуючого/впливового.

2.4.6 FER-2013

Набір даних FER-2013 містить три тисячі п'ятсот позначених зображень у наборі для розробки, двадцять вісім тисяч позначених зображень у навчальному наборі та три тисячі п'ятсот зображень у тестовому наборі. Даний набір даних був створений шляхом збору зображень з Google для кожної емоції та її синонімів.

Кожне зображення у наборі FER-2013 класифікується як одна з семи емоцій: гнів, радість, здивування, сум, страх, огида (рисунок 2.16). Щастя є

найпоширенішою емоцією, забезпечуючи основний рівень для випадкових здогадів на рівні 24,4%.



Рисунок 2.16 – Приклади фотографій з датасету FER-2013

2.4.7 Google Facial Expression Comparison Dataset

Даний набір даних є обширною колекцією трійок зображень облич, які супроводжуються анотаціями, що вказують, які два обличчя в кожній трійці найбільш схожі за виразом обличчя. Цей набір даних має істотні відмінності від інших наборів даних щодо виразів, які зазвичай сконцентровані на дискретній класифікації емоцій або виявленні конкретних дій, призначений для підтримки дослідників, що працюють у сферах, пов'язаних з аналізом виразу обличчя, таких як пошук зображень за виразом, створення загального зображення альбому на основі синтезу виразу, класифікації емоцій, та інші суміжні напрямки [13].

2.4.8 K-EmoCon

К-EmoCon [14] є мультимодальним набором даних з обширними анотаціями щодо неперервних емоцій під час природних розмов. У ньому включені мультимодальні вимірювання, такі як аудіовізуальні записи, ЕЕГ та периферичні фізіологічні сигнали, отримані в ході 16 сеансів приблизно 10-хвилинних дебатів з соціальних питань.

Набір даних відрізняється від попередніх тим, що містить анотації емоцій з усіх трьох точок зору: учасників дискусій, їхніх партнерів у дебатах та зовнішніх спостерігачів. Експерти, кожні 5 секунд, аналізували емоційні вияви у відеозаписах дебатів, зосереджуючись на аспектах збудження-валентності та 18 додаткових категорійних емоцій. К-EmoCon є першим доступним для загального використання набором даних з емоцій, який враховує багато аспектів оцінки емоцій під час соціальних взаємодій.

2.4.9 Interactive Emotional Dyadic Motion Capture (IEMOCAP)

Interactive Emotional Dyadic Motion Capture (IEMOCAP) [15] вважається активною мультимодальною базою даних, яка містить різноманітні динамічні дані, зібрані в лабораторії SAIL в USC. Вона містить близько 12 годин аудіовізуального контенту, який включає відео, аудіо, записи обличчя та текстові транскрипції. Ця база даних складається з діадичних сеансів, під час яких актори виконують імпровізації або сценарії, спеціально створені для виклику емоційних відгуків.

IEMOCAP анотується різними анотаторами за допомогою категоріальних міток гніву, смутку, щастя, байдужості, а також за розмірними мітками, такими як валентність, активація та домінування. Завдяки детальній інформації про рух, інтерактивному налаштуванню для виклику реальних емоцій та обсягу бази даних, цей корпус стає важливим додатком до існуючих джерел інформації у галузі вивчення та моделювання мультимодального та виразного людського спілкування.

Ця база даних може бути розділена на три основні області досліджень:

– розпізнавання та аналіз виразу емоцій: ця область вивчає методи та алгоритми для виявлення та розпізнавання емоційних виразів у відео-, аудіо- та текстових даних. Мета – розробка систем, які можуть автоматично ідентифікувати емоції, виражені людьми;

– аналіз діадичних взаємодій людини: ця область охоплює вивчення взаємодій між людьми у діадах, наприклад, спілкування в парах. Досліджується взаємодія, комунікація та емоційні прояви у різних сценаріях та ситуаціях;

– проектування чутливих до емоцій інтерфейсів та віртуальних агентів: ця область фокусується на розробці інтерфейсів та віртуальних агентів, які можуть реагувати на емоції користувачів. Мета – створення систем, які можуть взаємодіяти з людьми, розуміючи їхні емоції та адаптуючи свою поведінку відповідно.

2.5 Висновки до розділу

На сьогоднішній день існує значна кількість продуктів, які намагалися вирішити проблему розпізнавання емоцій у людини. Незважаючи на широкий асортимент додаткових функцій та привабливе оформлення, більшість з них має свої недоліки, що не дозволяє вважати їх еталоном та стверджувати, що проблема вже вирішена.

Основні недоліки таких продуктів включають наступне:

– низький рівень точності розпізнавання;
– неспроможність розпізнати емоції у дітей;
– значне погіршення точності при нахилі камери під кутом більше 15 градусів.

Ці обмеження стимулюють розробку нового, більш якісного продукту, який зможе вирішити ці проблеми більш ефективно.

До того ж, у цьому розділі були розглянуті основні набори даних, що стосуються розпізнавання емоцій людини на основі фотографій. Як вже було зазначено, більшість з них не відповідає вимогам якості, тому навіть в теорії не можуть бути основою для розробки професійної моделі для розпізнавання емоцій.

Проте серед усіх датасетів найбільш якісним виявився останній, IEMOCAP. Саме його було обрано для побудови моделі, оскільки він найбільш відповідає вимогам якості та достовірності.

3 ОПИС РОЗРОБЛЕНОЇ СИСТЕМИ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

3.1 Опис розробленого алгоритму

У попередньому розділі ми прийняли рішення використовувати датасет IEMOSCAP. У цьому розділі ми докладніше розглянемо обраний метод та його модифікацію, яка значно підвищила точність моделі. Потім ми розглянемо побудовану систему. Але передусім давайте поглибимось у знання того, що означає згорткова нейронна мережа.

3.1.1 Згорткова нейронна мережа (ЗНМ)

Згорткова нейронна мережа є ключовим інструментом для класифікації та розпізнавання об'єктів та осіб на фотографіях, а також для розпізнавання мови. Вона має безліч застосувань, таких як Deep Convolutional Neural Network (DCNN), Region-CNN (R-CNN), Fully Convolutional Neural Networks (FCNN), Mask R-CNN та інші.

Згорткові нейронні мережі використовуються як варіанти багатошарових перцептронів, спроектованих таким чином, щоб потребувати мінімальної обробки. Вони мають спільні ваги та характеристики інваріантності відносно паралельного зсуву. Згорткові нейронні мережі потребують порівняно невеликої попередньої обробки, порівняно з іншими алгоритмами класифікації зображень. Це означає, що мережа навчається за допомогою фільтрів, які в традиційних алгоритмах доводиться розробляти вручну.

Ця автономність у створенні ознак без попереднього знання та втручання людини є ключовою перевагою. Згорткова нейронна мережа складається зі вхідних і вихідних шарів, а також з декількох прихованих шарів.

Зазвичай, в прихованих шарах згорткової нейронної мережі включаються згорткові шари, шари пулінгу, повнозв'язні шари та шари нормалізації. Згорткові шари застосовують операцію згортки до вхідних даних та передають результат наступному шару (рисунок 3.1). Ця операція емулює реакцію окремого нейрону на зоровий стимул.

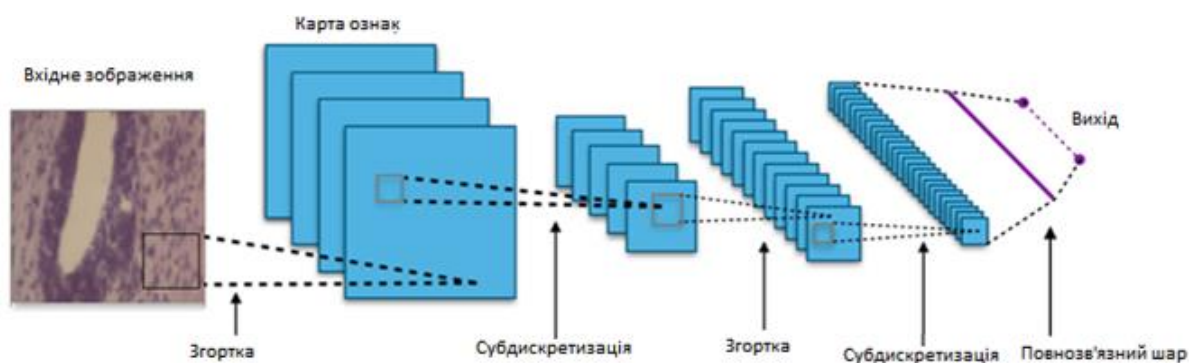


Рисунок 3.1 – Принцип роботи згорткової нейронної мережі

Наприклад, у повнозв'язаному шарі для (невеликого) зображення розміром 100×100 кількість ваг становить 10 000. Операція згортки вирішує цю проблему, оскільки вона скорочує кількість параметрів, що дозволяє мережі стати глибшою за меншу кількість параметрів.

Наприклад, незалежно від розміру зображення, області згортки розміром 5×5 з однаковими вагами вимагають лише 25 вільних параметрів. Таким чином, це вирішує проблему вибуху або зникнення градієнтів під час тренування традиційних багатошарових нейронних мереж з численними шарами за допомогою зворотного поширення.

Згорткові мережі можуть містити шари локального або глобального об'єднання, які комбінують виходи кластерів нейронів одного рівня передачі до наступного. Наприклад, у максимізаційному агрегуванні використовується максимальне значення з кожного кластеру нейронів попереднього шару, тоді як у випадку усередненого агрегування використовується середнє значення. Повнозв'язані шари з'єднують кожен

нейрон одного шару з кожним нейроном наступного, що, в сутності, відповідає структурі традиційної багатошарової перцептронної нейронної мережі. Згорткові шари використовують спільні ваги, тобто для кожного рецептивного поля шару використовується той самий фільтр (банк ваг), що сприяє зменшенню потреби у пам'яті та покращує продуктивність.

3.1.2 Глибока згорткова мережа з прямим зв'язком

За основу була взята згорткова нейронна мережа [17] з архітектурою, що наведена на рисунку 3.2.

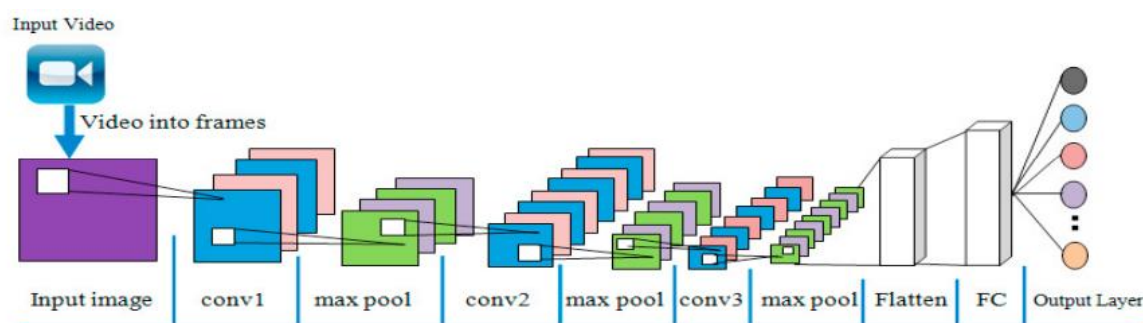


Рисунок 3.2 – Архітектура глибокої згорткової мережі з прямим зв'язком

Згортковий шар даної мережі функціонує на послідовних етапах:

а) згортковий шар: процес згортки включає наступні кроки:

- вхідне зображення та детектор функцій беруться в розгляд;
- кожний піксель зображення множиться на відповідний піксель функції;
- суму значень обчислюється;
- отримана сума розділяється на загальну кількість пікселів у зображенні;

б) об'єднання шарів: шари об'єднання або підвибірки зменшують розмір карти до меншого розміру. Процес об'єднання включає наступні кроки:

- вибирається розмір вікна;
- вибирається крок для переміщення вікна;
- вікно переміщується по фільтрованому зображенню;
- з кожного вікна обирається максимальне значення;
- випрямлена лінійна одиниця (ReLU): це функція активації, яка повертає нуль, коли вхід менше нуля, та сам вхід, коли він більше або рівний нулю. Вона обчислюється за формулою $R(x) = \max(0, x)$;
- повністю пов'язаний шар: це шар, в якому кожен нейрон пов'язаний з кожним нейроном попереднього шару. Цю операцію можна обчислити за допомогою формули $F(x) = \sigma(W * x)$;
- шар Softmax: цей шар дозволяє виконати зворотне поширення та покращує продуктивність мережі. Якщо N – розмір вхідного вектора, то цей шар обчислюється відповідно;
- вихідний шар: розмір вихідного шару відповідає кількості класів та представляє клас вхідного зображення.

3.1.3 Модифікація та навчання згорткової нейронної мережі

Цей підрозділ становить кульмінацію даної роботи, пропонуючи модифікацію, яка значно покращила якість результатів у вирішенні проблеми розпізнавання емоцій людини. Згідно з модифікацією, між кроками 3 і 4 згорткового шару, описаного у попередньому розділі, було додано щільний шар (dense layer).

Після цього було додано відповідну кількість епох та налаштування гіперпараметрів.

На рисунках 3.3 та 3.4 продемонстровано порівняння процесу навчання створеної нейронної мережі з використанням різних наборів гіперпараметрів.

Оскільки графік зменшення помилок продемонстрував монотонний спад, кількість епох була експериментально встановлена на рівні 600.

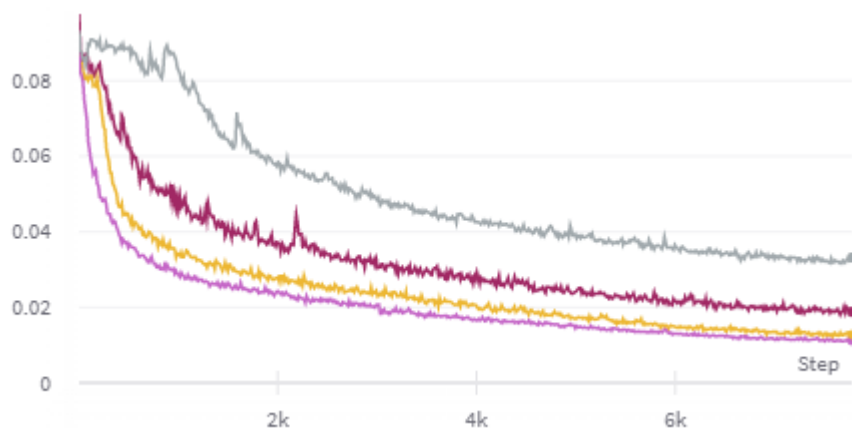


Рисунок 3.3 – Зменшення лосу (помилки моделі при навчанні)

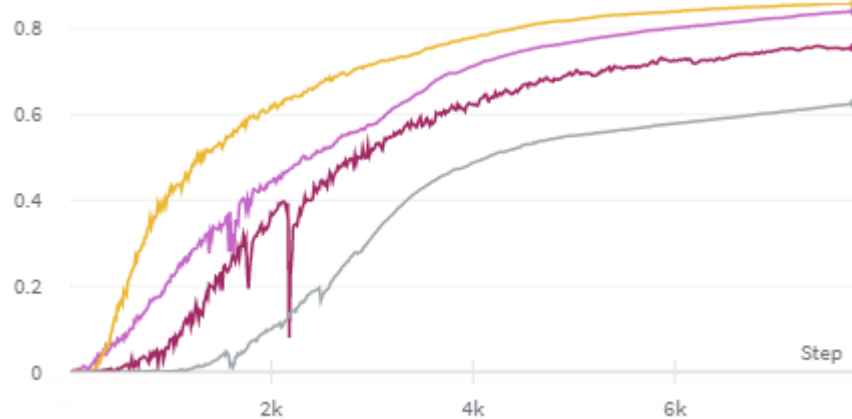


Рисунок 3.4 – Приріст точності моделі на навчальному датасеті

3.2 Обрана метрика та якісне порівняння методів

3.2.1 Обрана метрика

При оцінці ефективності методу використовувалися такі метрики, як точність (precision), повнота (recall) та F-міра (F-score). Давайте розглянемо їх більш детально.

Точність (Precision): Це відношення кількості правильно визначених релевантних елементів до загальної кількості елементів, які були визначені як релевантні.

Повнота (Recall). Це відношення кількості правильно визначених релевантних елементів до загальної кількості релевантних елементів, які могли бути визначені.

F-міра (F-score). Це гармонічне середнє між точністю та повнотою і визначається за допомогою формули, яка враховує обидві ці метрики, формула F-міри.

Використання цих метрик допомагає отримати більш об'єктивну оцінку ефективності тесту.

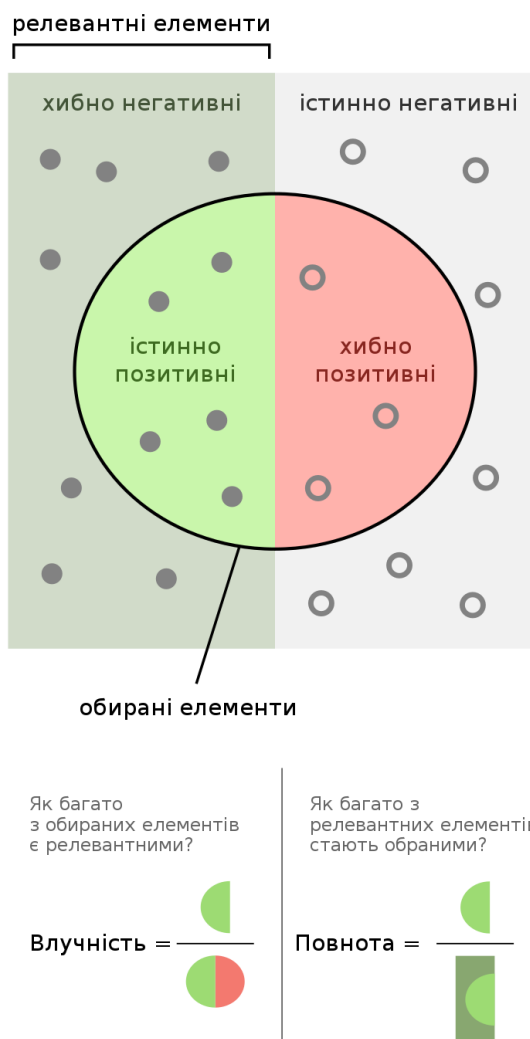


Рисунок 3.5 – Влучність та повнота

3.2.2 Результати попередніх методів

Розглянемо останні досягнення науки в області розпізнавання емоцій людини по обличчю. Останні досягнення в області розпізнавання емоцій людини по обличчю включають в себе використання глибоких нейронних мереж, особливо згорткових нейронних мереж (CNN), для автоматичного виявлення та класифікації емоцій на обличчях. Дані мережі можуть аналізувати вхідні зображення облич та визначати емоційний стан особи з високою точністю. На рисунку 3.6 наведені найкращі моделі, оцінені на датасеті AffectNet.

Facial Expression Recognition on AffectNet

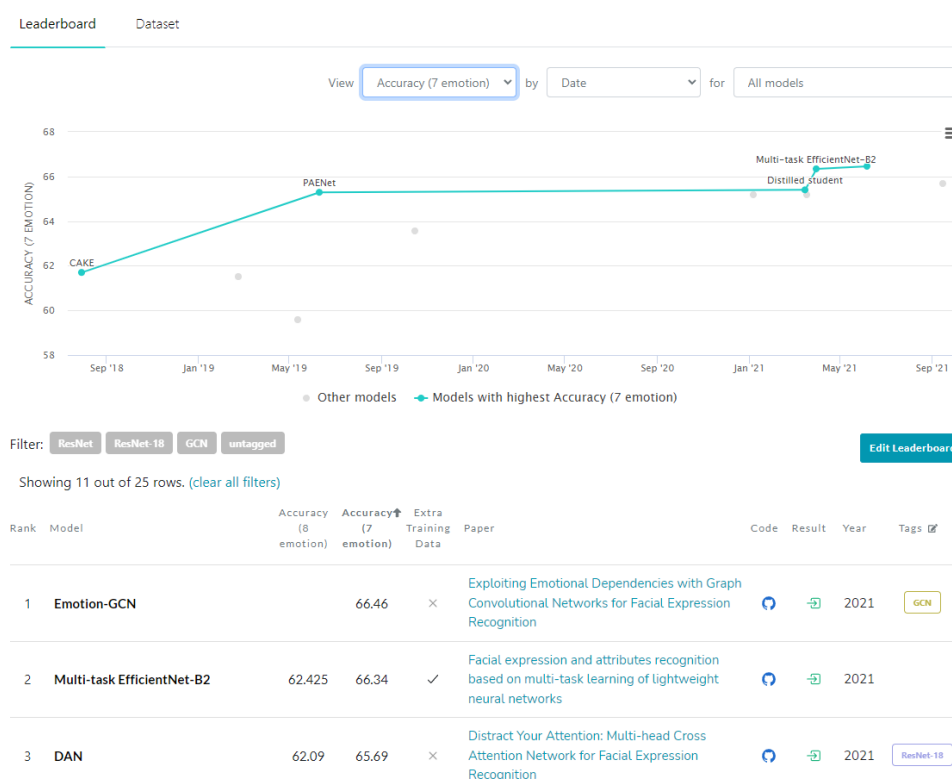


Рисунок 3.6 – Найкращі моделі, оцінені на датасеті AffectNet

Перша модель, відома як Emotion-GCN, забезпечила точність на рівні 66,46% на датасеті AffectNet.

Друга модель, Multi-task EfficientNet-B2, була випробувана на цьому ж датасеті та демонструє точність 66,34% у розпізнаванні семи емоційних станів.

Третя модель у рейтингу DAN показала результат 65,69% за тих самих умов.

Ці моделі є останніми досягненнями в цій галузі і були розроблені у 2021 році.

3.2.3 Результати розробленого методу

На датасеті IEMOCAP була розроблена нейронна мережа, яка досягла точності (accuracy) на рівні 71,5%. Цей результат перевищує точність найближчих конкурентів на понад 5%.

3.3 Опис компонентів побудованої системи

Основним внеском даної роботи є проведення тренувальних експериментів та розробка нейронної мережі. Розроблена система, хоч і має просту структуру, виконує важливу демонстраційну функцію. Центральним елементом системи є файл з розширенням `.ipynb`, який можна виконати за допомогою пакету Jupyter Notebook.

Цей документ містить пайплайн обробки зображень та налаштування демонстраційного стенду. Сам стенд використовує безкоштовну та зручну бібліотеку Gradio, що забезпечує привабливий інтерфейс для демонстрації.

Після виконання цього файлу автоматично генерується унікальне посилання, згідно з якими можна налаштувати доступ до стартового інтерфейсу без прив'язки до місця чи країни. Приклад використання даного інтерфейсу наведено на рисунку 3.7.

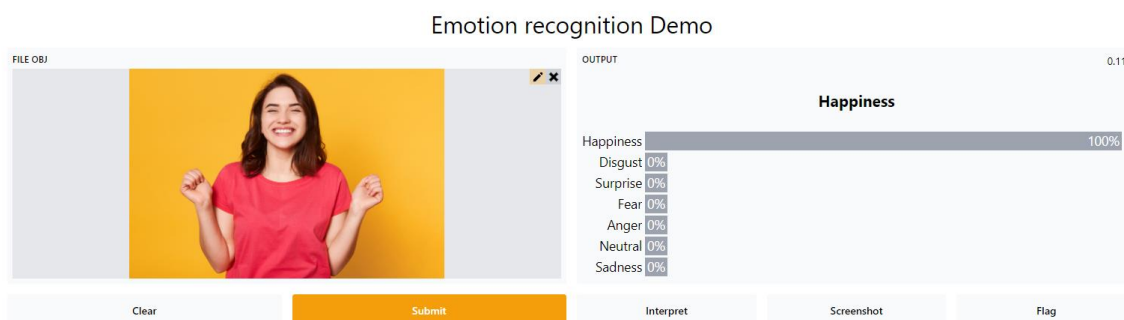


Рисунок 3.7 – Приклад використання створеного інтерфейсу

3.4 Висновок до розділу

У цьому розділі було детально описано структуру та архітектуру нейронної мережі, яка була використана для побудови моделі. Описано кожен шар мережі, його функції та параметри, що використовувалися. Крім того, були розглянуті зміни, які були внесені до початкової архітектури моделі з метою покращення її результатів. Особлива увага приділялася змінам, які стосувалися різних шарів мережі, таких як згорткові, пулінгові та повністю пов'язані шари, а також функцій активації.

Детально була розглянута обрана метрика оцінки якості моделі – точність (ассигасу). Пояснено, як вона обчислюється та які переваги та недоліки вона має у вимірюванні ефективності моделі.

Надалі, у розділі наведено результати останніх досліджень у сфері розпізнавання емоцій за обличчям. Зазначено, що найкращі з них досягли точності на рівні 66,34%. Проте, надано важливий контекст, що на обраному датасеті IEMOCAP розроблена модель досягла значно кращого результату – точність становила 71,5%. Це свідчить про ефективність та перспективність використання даної нейронної мережі для задач розпізнавання емоцій на вибраному датасеті.

4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту

Таблиця 4.1 надає огляд основної концепції проекту, потенційних сфер практичного застосування продукту та основних переваг, доступних користувачам при використанні розробленого додатку.

Таблиця 4.1 – Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Створення системи для виявлення емоцій у людей під час перегляду реклами.	Маркетинг	Здатність до більш точного та ефективного розміщення реклами відповідно до реакції користувача.

Наступним кроком буде проведено аналіз техніко-економічного потенціалу, параметрів розробки та порівняння з конкуруючими додатками (таблиця 4.2). Основними з них є EmoDetect, Хеома та Face Reader (Noldus).

Таблиця. 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№	Техніко-економічні характеристики проекту	Мій проект	Конкуренти			W	N	S
			1	2	3			
1	Висока точність розпізнавання емоцій	+	-	-	-	+		
2	Простота використання	+	+	-	+		+	
3	Здатність до аналізу у реальному часі	+	+	+	+		+	
4	Можливість побудови графіків	-	+	+	+			+

4.2 Технологічний аудит ідеї проекту

Далі ми проведемо аудит технологічних засобів, які будуть використовуватися для реалізації проекту. Результати аудиту подано в таблиці 4.3.

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

Ідея проекту	Реалізація	Наявність	Доступність
Аналіз емоцій людини/людей під час перегляду рекламних роликів	Pytorch	наявна	доступна
	HDFS	наявна	доступна
	Mongo DB	наявна	доступна
	Rest Api	наявна	доступна
	Gradio	наявна	доступна
	AWS EC2	наявна	доступна
	AWS S3	наявна	доступна

4.3 Аналіз ринкового потенціалу для старту стартап-проекту

Аналіз ринку (таблиця 4.4.)

Таблиця 4.4 – Потенційний ринок стартап-проекту

№	Стану ринку	Опис
1	Кількість гравців	5-7
2	Обсяг загальних продажів, вимірюваний в гривнях на одиницю виміру (наприклад, на один місяць або одну одиницю товару)	Середня вартість ліцензії на використання, яка становить від 10 до 20 тисяч доларів.
3	Ринок демонструє	Стабільна тенденція постійного зростання
4	Обмеження для входу	У середньому, конкуренція на ринку в цьому сегменті є помірною, проте при прийнятті рішень щодо безпеки споживачі в основному довіряють компаніям з визнаною репутацією
5	Вимоги щодо стандартизації та сертифікації в даній області	Поточне законодавство не містить специфічних вимог

У таблиці 4.5 наведено потенційно-зацікавлені групи, їх вимоги та потреби.

Таблиця 4.5 – Ознаки можливих користувачів стартапу

№	Створення попиту на ринку	Сегментація цільової аудиторії проекту	Аналіз унікальних особливостей поведінки різних груп споживачів	Очікування та вимоги користувачів до функціональності проекту
1	Бажання ефективно використовувати ресурси, у тому числі рекламний бюджет, під час реалізації маркетингової кампанії	Взаємодія з маркетинговими агентствами для ефективної стратегії сприйняття	Розуміння та врахування емоційного аспекту	Оптимізація зручності використання та продуктивності в роботі

Після ідентифікації потенційних цільових аудиторій було проведено аналіз ринку на предмет наявності загроз (таблиця 4.6) та можливостей (таблиця 4.7).

Таблиця 4.6 – Фактори загроз

№	Фактор	Зміст загрози	Можлива реакція компанії
1	Конкуренція	На ринку спостерігається активна присутність стартапів, що впроваджують передові технології	Для запобігання можливій втраті співпраці через цей фактор, ми плануємо розширення функціоналу нашого продукту
2	Недостатній функціонал продукту	Це може призвести до відмови маркетингових агентств у співпраці з нашою компанією	Якщо високий відсоток відмов від співпраці стане проблемою, ми розглянемо можливість додаткового розширення функціоналу в майбутньому

Таблиця 4.7 – Фактори можливостей

№	Фактор	Зміст можливості	Можлива реакція компанії
1	Можливості залучення інвестицій	Існує потенціал залучення інвестицій, де основними фінансовими суб'єктами будуть фонди та інвестиційні компанії	відкриває шлях до розширення нашого розробницького складу та функціоналу додатку

Ознаки конкуренції приводяться у таблиці 4.8.

Таблиця 4.8 – Ступеневий аналіз конкуренції ринку

Специфіка конкурентного становища	Проява характеристики	Вплив на вектор дій компанії (потенційні кроки для підтримки конкурентоспроможності)
Характер конкуренції: Чиста конкуренція	Присутність незалежної конкуренції на ринку	Інтенсивна робота у сфері технологій та постійний розвиток
Географічний масштаб конкуренції: Міжнаціональний	Конкурентні битви з міжнародними компаніями	Присутність на міжнародному ринку
Внутрішньогалузева ознака конкуренції	Наявність розгалуженого програмного забезпечення	Зосередженість на впровадженні маркетингових ініціатив
Специфіка конкуренції за товарним видом: Товарно-видова	Достовірна участь відомих конкурентів у сфері	Привертання та заслужений авторитет серед висококваліфікованих фахівців у відповідній галузі
Інтенсивність конкуренції: Марочна.	Розпізнавання конкурентних брендів	Створення власного та унікального бренду продукту

Аналіз конкурентних умов за допомогою моделі М. Портера (таблиця 4.9).

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти на ринку	Можливі конкуренти на ринку	Постачальники	Клієнти	Товари-замінники
	EmoDetect / Face Reader				Середні та малі компанії
Результат	Жвава конкурентна боротьба, що характеризується активними зусиллями учасників ринку	Конкурентність ринку утримується завдяки створенню сприятливих умов для конкурентів	Постачальники встановлюють високі вартості ліцензій, що впливає на загальні витрати	Високі вимоги клієнтів до якості продукції та послуг від виробників	Відсутність адекватних альтернативних варіантів для заміщення продукту

Існує можливість успішного вступу на ринок, але це передбачатиме необхідність забезпечення конкурентоздатної цінової політики та розширеного функціоналу нашого продукту.

Загальний огляд конкурентних переваг представлено у наступній таблиці 4.10.

Таблиця 4.10 – Основні ознаки конкурентоспроможності

№	Ознаки конкурентоспроможності	Важливість факторів для порівняння конкурентних проєктів:
1	Найбільша упізнаваність	Основна функція – розпізнавання емоцій
2	Часткова упізнаваність	Розпізнання в реальному часі емоцій та на основі цього приймаються рішення про час показу реклами
3	Вартість ліцензії	Конкурентоспроможна ціна

Важливим кроком є зіставлення сильних та слабких сторін проекту у порівнянні з конкуруючими компаніями (таблиця 4.11).

Таблиця 4.11 – Порівняння сильних та слабких сторін продукту

№	Рівень конкурентоспроможності	Оцінка 1-20	Рейтинг конкурентів						
			-3	-2	-1	0	1	2	3
1	Максимальна точність	19						x	
2	Розпізнавання у реальному часі	18					x		
3	Вартість ліцензії	15			x				

Останнім кроком є SWOT-аналіз (таблиця 4.12).

Таблиця 4.12 – SWOT-аналіз стартап-проекту

Переваги: висока точність розпізнавання, легкість використання та доступність	Недоліки: обмежений інтерфейс та функціонал
Розпізнавання емоцій в реальному часі	Загрози: велика конкуренція серед компаній

Таблиця 4.13 – Альтернативи ринкового впровадження стартап-проекту

№	Альтернативна ринкова стратегія	Можливість отримання ресурсів.	Термін реалізації
1	Об'єднання компаній для виконання проєктів	Відчуження розробки зацікавленій компанії.	До двох років
2	Модернізація розробки у відкритий проєкт з вихідним кодом	Залучення розробників, експертів	До одного року
3	Монетизація	Використання платформ мінімальним вкладом	До 3-х місяців

4.4 Розробка ринкової стратегії стартап-проекту

В таблиці 4.14 наведено опис цільових груп потенційних споживачів продукту. Такими групами стали малий, середній та великий бізнес у різних галузях інформаційних технологій.

Таблиця 4.14 – Вибір цільових груп потенційних споживачів

№	Основні клієнти	Готовність до придбання продукту	Можливий попит	Конкуренція	Рівень складності
1	Малий бізнес	Максимальна	Максимальний попит	Достатня	Проста
2	Середній бізнес	Максимальна	Високий, при умові доступної вартості	Середня	Середня
3	Великий бізнес	Мінімальна готовність	Високий, при умові розвитку та попиту	Висока	Складна

Після аналізу отриманих результатів було вирішено, що для подальшого розвитку найбільш доцільним буде спрямувати зусилля на групу малих та середніх підприємств. Це обумовлено тим, що в масштабах великих компаній прийняття рішень зазвичай відбувається повільно і потребує багатоетапних процедур та документального оформлення, що може ускладнити процес масштабування та підтримки розробки.

Після оцінки отриманих результатів було вирішено вибрати альтернативний шлях, який передбачає продаж власної реалізації та технологій за менш високою, але прийнятною ціною, що є більш привабливим варіантом з комерційної точки зору.

Обрання основної ринкової стратегії наведено у таблиці 4.15.

Таблиця 4.15 – Обрання основної ринкової стратегії

№	Альтернативний розвиток	Стратегія розвитку	Конкурентні стратегії	Ключова стратегія розвитку
1	Реалізація технологій	Зменшення вартості та співпраця з менш відомими платформами	Зменшення собівартості та задоволення потреб більшої кількості споживачів	Вигідніші умови в порівнянні з конкурентами

Основна ідея розвитку проекту з урахуванням конкурентоспроможності відображена у таблиці. 4.16.

Таблиця 4.16 – Формування основної концепції конкурентного розвитку

Новизна проєкту на ринку:	Так
Залучення нової аудиторії користувачів або аудиторії конкурентів:	Залучення нових користувачів
Використання типових, напрацьованих розробок:	Ні, виключно новаторські та власні розробки
Конкурентна стратегія:	Першість на ринку

Враховуючи головні вимоги користувачів до продукту, обрані стратегії конкурентної поведінки на ринку та основні концепції конкурентного розвитку, ми визначимо стратегію позиціонування (таблиця 4.17).

Таблиця 4.17 – Визначення стратегії позиціонування

№	Потреби аудиторії	Стратегія	Конкурентоспроможність	Комплексна позиція розробки
1	Економія бюджету для реклами через використання соціальних мереж	Залучення фахових ІТ спеціалістів	Доступна ціна ліцензії для використання	Доступність та легкість у використанні

4.5. Розроблення маркетингової програми стартап-проекту

На етапі початкового запуску необхідно розробити маркетингову концепцію продукту, орієнтовану на потреби користувачів. Результати аналізу представлені в таблиці 4.18.

Таблиця 4.18 – Узагальнення переваг концепції розроблюваного продукту

№	Потреба	Цінність товару	Переваги перед конкурентами
1	Економія бюджету проєкту	Мінімальне капіталовкладення	Унікальність технологій

Маркетингова модель товару (таблиця 4.19).

Таблиця 4.19 – Опис трьох рівнів моделі товару

Рівні	Сутності та складові
1. Ціль	Впровадження стабільної економічної системи
2. Реалізація	Характеристики: розпізнавальна система
	Система розпізнавання з API
	Якість: використанням інтеграційних та юніт тестів під час розробки продукту
	Марка: лого, впізнаваний мерч та назва
3. Підкріплення товару	До продажу: код продукту
	Після продажу: права на володіння

Запобігання копіюванню розробленого товару забезпечується завчасною компіляцією, яка ускладнює доступ до вихідного коду, оскільки він вже перетворений у виконуваний формат. Додатково, існують користувацькі ліцензійні угоди та прямі договори, що чітко забороняють копіювання розробленого продукту та вчинення плагіату.

У таблиці 4.20 наведено цінові рамки, що стануть ключовими аспектами при ціноутворенні продукту.

Таблиця 4.20 – Визначення цінових рамок

№	Вартість товарів-замінників	Вартість товарів-аналогів	Прибуток	Межа ціноутворення на товар/послугу
1	\$15500-35000	\$6000-50000	>\$500000	\$0-\$9000

Збут продукції наведено в таблиці 4.21. Було прийнято рішення проводити збут власними силами.

Таблиця 4.21 – Формування системи збуту

№	Поведінка потенційних клієнтів	Функції збуту	Рівень каналу збуту	Система збуту
1	Купівля ліцензійного продукту-програми	Укладення договорів-збуту	Однорівневий	Право інтелектуальної власності належить розробнику програмного забезпечення. Вертикальна система

Маркетингові стратегії, як фінальний етап реалізації, наведено у таблиці 4.22.

Таблиця 4.22 – Концепція маркетингових комунікацій на ринку

№	Цільова аудиторія	Канали зв'язку з цільовою аудиторією	Ключові стратегії	Ключова ціль рекламних компаній	Концепція
1	Одноразова підписка під час придбання	Реклама на різних майданчиках, рекомендації	Зручність та доступність	Наглядна демонстрація плюсів продукту	«Краще конкурентів»

4.6 Висновок до розділу

У цьому розділі описується та проводиться аналіз стартап-проекту, у тому числі його потенційних переваг та недоліків. Визначаються та демонструються можливі сильні та слабкі сторони ідеї. Також проведено оцінку технічної можливості реалізації цього проекту.

В результаті дослідження стало відомо, що ринок, на якому буде діяти проект, відзначається високим рівнем конкуренції через присутність значної кількості компаній. Однак, ціна ліцензій для використання їхніх розробок також є високою.

Після аналізу потенційних стратегій на ринку було визначено, що кращим рішенням буде спрямування на малі та середні підприємства, оскільки великі компанії вже мають встановлені зв'язки з іншими гравцями на ринку.

Згідно стратегії позиціонування, було визначено, що необхідно сконцентруватись на розробці зручного та популярного додатку. Також, важливо регулярно розширювати можливості функціоналу проекту. Встановлення низької рентабельної ціни на продукт виявилось ключовим для успішного та швидкого входу на ринок і забезпечення здорової чесної конкуренції.

ВИСНОВКИ

У даній роботі була детально розглянута проблема розпізнавання емоцій людини на відео. Очевидно, що в сучасному світі значення штучного інтелекту в житті людини важко переоцінити. Вивчені основні труднощі, з якими зіштовхуються вчені, які працюють над розпізнаванням емоцій людини на відео. Серед них варто відзначити наступні:

Проблему датасетів, яка полягає у складності збору та підготовки великої кількості даних для тренування моделей.

Похибки в розпізнаванні неповністю виражених емоцій, що може виникати через складність інтерпретації нечітких виразів обличчя або жестів.

Відмінність між західною та східною культурами, яка може впливати на сприйняття та вираження емоцій, що робить процес розпізнавання більш складним.

Наступним кроком було дослідження існуючих готових реалізацій, таких як EmoDetect, Xeoma, FaceReader (Noldus), Microsoft Oxford Project Emotion Recognition та eMotion Software, для визначення їхньої ефективності та можливостей у вирішенні цих проблем.

Серед основних недоліків зазначених систем можна відзначити низьку точність розпізнавання або обмежену можливість протестування через високу вартість користування повним функціоналом. Деякі системи також характеризуються застарілим інтерфейсом.

У цьому дослідженні розглядалися різні методи для виявлення обличчя на фотографіях. Серед них варто відзначити наступні:

- локальні бінарні шаблони (LBP);
- гістограма направлених градієнтів (HOG);
- геометричний метод розпізнавання обличчя;
- метод Віоли-Джонса;
- метод гнучкого порівняння на графах.

Частина вищезазначених методів виявилися застарілими та майже не використовуються в сучасних дослідженнях, тоді як інші продемонстрували гарний результат і активно використовуються дослідниками в наш час.

Наступним кроком у цьому дослідженні був аналіз традиційних методів розпізнавання емоцій, серед яких зазначені:

- метод К-найближчих сусідів (KNN);
- розріджена логістична регресія зі штрафом $L1/2$;
- метод опорних векторів (SVM).

Описані методи є методами машинного навчання, застосованими до проблеми класифікації, і виявилися не найефективнішими у завданні розпізнавання емоцій людини.

Одним із найважливіших аспектів даної роботи було вибір датасету. Було проведено дослідження дев'яти традиційних датасетів для проблеми розпізнавання емоцій, а саме:

- Dreamer;
- AffectNet;
- Emotic;
- Ascertain;
- Google Facial Expression Comparison Dataset;
- Extended Cohn-Kanade Dataset (CK+);
- K-EmoCon;
- FER-2013;
- Interactive Emotional Dyadic Motion Capture (IEMOCAP).

Жоден із вивчених датасетів не відповідає всім критеріям «золотого стандарту» датасетів, який включає якісні зображення, значний обсяг даних, різноманітність, збалансованість та належний поділ на навчальну, валідаційну та тестову вибірки. У зв'язку з цим було вирішено обрати датасет IEMOCAP, який найбільш відповідає цим критеріям.

Після цього було обрано метрику та проведено аналіз існуючих результатів. Найвища точність серед існуючих моделей становила 66,34%.

Подальшим кроком було розглянуто архітектуру глибокої згорткової нейронної мережі з прямим зв'язком, а також запропоновано її вдосконалення. Ці заходи призвели до підвищення точності до 71,5%.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

- 1) Face detector (emotions) in xeoma – felenasoft. *Video surveillance, vsaas, outsourcing – Felenasoft.* URL: <https://felenasoft.com/xeoma/en/articles/emotions-recognition-in-xeoma/> (date of access: 07.04.2024).
- 2) Happy? Sad? Angry? This Microsoft tool recognizes emotions in pictures - The AI Blog. *The AI Blog.* URL: <https://blogs.microsoft.com/ai/happy-sad-angry-this-microsoft-tool-recognizes-emotions-in-pictures/> (date of access: 07.04.2024).
- 3) eMotionSoftware. URL:: <https://tecnitop.com/en/emotion-sensefly-en/> (date of access: 07.04.2024).
- 4) Shakhnarovich G., Darrell T., Indyk P. Nearest-Neighbor Methods in Learning and Vision. 2005.
- 5) Cristianini N., Shawe-Taylor J. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press. (2000).
- 6) AffectNet – mohammad H. mahoor, phd. *Mohammad H. Mahoor, PhD – Professor, ECE – Ritchie School of Engineering and Computer Science – University of Denver.* URL: <http://mohammadmahoor.com/affectnet/> (date of access: 27.04.2024).
- 7) ASCERTAIN dataset. *ASCERTAIN dataset.* URL: <https://ascertain-dataset.github.io/> (date of access: 24.05.2024).
- 8) Stamos K., Naeem R. DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *Zenodo.* URL: <https://zenodo.org/record/546113> (date of access: 24.05.2024).
- 9) Resources – jeffrey cohn. *Jeffrey Cohn.* URL: <http://www.jeffcohn.net/Resources/> (date of access: 27.04.2024).
- 10) Emotions in context. *AI well.* URL: <http://sunai.uoc.edu/emotic/> (date of access: 27.04.2024).

11) Fer2013. *Kaggle: Your Machine Learning and Data Science Community*. URL: <https://www.kaggle.com/deadskull7/fer2013> (date of access: 27.04.2024).

12) Datasets – google research. *Google Research - Explore Our Latest Research in Science and AI*. URL: <https://research.google/tools/datasets/google-facial-expression/> (date of access: 27.04.2024).

13) K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations – Scientific Data. *Nature*. URL: <https://www.nature.com/articles/s41597-020-00630-y> (date of access: 27.04.2024).

14) IEMOCAP- home. *Signal Analysis and Interpretation Laboratory (SAIL) Ming Hsieh Department of Electrical Engineering and Computer Engineering; Department of Computer Scienceâ USC Viterbi School of Engineering*. URL: <https://sail.usc.edu/iemocap/> (date of access: 27.04.2024).

15) Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*. 2012. No.1. P. 1097–1105.