МАГІСТЕРСЬКА РОБОТА

МР. ІПм - 42.00.00.000 ПЗ

Група ІПм-24-2

Федоренко Роман

2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Федоренко Роман Михайлович

(прізвище, ім'я, по батькові)

УДК 004.9

(індекс)

МАГІСТЕРСЬКА РОБОТА

Аналіз та моделювання емоцій в контексті релевантних та похідних даних

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Федоренко Р.М.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник Яцишин Микола Миколайович, к.т.н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту

Завідувач кафедри

доц. Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Нормоконтроль

доц. Вовк Р.Б.

(посада) (підпис) (дата) (ініціали та прізвище)

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

Івано-Франківськ – 2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:
Зав. кафедрою **ІПЗ**
доц. В.В. Бандура
“ 04 ” вересня 2025 р.

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Федоренку Роману Михайловичу

(прізвище, ім’я, по-батькові)

1. Тема магістерської роботи “ Аналіз та моделювання емоцій в контексті релевантних та похідних даних ”

керівник проекту (роботи) Яцишин М.М., к.т.н., доцент

затверджені наказом закладу вищої освіти від “ 05 ” листопада 2025 р. № 695/7

2. Строк подання студентом проекту (роботи) 15 грудня 2025 р.

3. Вихідні дані до проекту (роботи) Формальні моделі і методи побудови інформаційних та програмних технологій моделювання емоцій

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

1. Аналіз предметної області моделювання емоцій з використанням мультимодальних даних
2. Методологія автоматичного розпізнавання спонтанних лицьових одиниць дії
3. Методологія розпізнавання самооцінних емоцій за виразами обличчя
4. Імплементація методів та моделювання емоцій в контексті релевантних та похідних даних

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

1. Загальний огляд гібридного фреймворку глибокого навчання (рис. 1.1)
2. Структура гібридної мережі розпізнавання (рис. 1.2)
3. Огляд методу FACS3D-Net для мультиміткової детекції AU (рис. 1.3)
4. Оцінка пози голови (рис. 1.4)
5. Блок-схема системи розпізнавання (рис. 1.5)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата
Перевірка на плагіат	доц., к.т.н. Вовк Р.Б.	

7. Дата видачі завдання 04 вересня 2025 р.

Керівник _____
(підпис)

Завдання прийняв до виконання_____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назви етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Підбір і вивчення літератури по темі магістерської роботи	15.09.2025	виконано
2	Аналіз предметної області моделювання емоцій з використанням мультимодальних даних	01.10.2025	виконано
3	Методологія автоматичного розпізнавання спонтанних лицьових одиниць дії	17.10.2025	виконано
4	Методологія розпізнавання самооцінних емоцій за виразами обличчя	02.11.2025	виконано
5	Імплементація методів та моделювання емоцій в контексті релевантних та похідних даних	19.11.2025	виконано
6	Аналіз ефективності злиття фізіологічних сигналів та одиниць дії у задачах детекції болю	02.12.2025	виконано
7	Затвердження пояснювальної записки роботи завідувачем кафедри	15.12.2025	виконано

Студент – магістр _____
(підпис)

Керівник роботи _____
(підпис)

АНОТАЦІЯ

Магістерська робота: 75 с., 19 рис., 9 табл., 43 джерел.

Тема: Аналіз та моделювання емоцій в контексті релевантних та похідних даних

Метою роботи є розроблення та обґрунтування мультимодальних методів аналізу й моделювання емоцій із використанням релевантних та похідних даних.

Об'єктом дослідження є процеси формування, вираження та автоматичного розпізнавання емоційних станів людини.

Предметом дослідження є методи, моделі та алгоритми мультимодального аналізу емоцій із використанням фізіологічних сигналів, одиниць дії, виразів обличчя та похідних контекстуальних даних.

Результати дослідження

В роботі запропонована архітектура нейронної мережі забезпечує ефективне моделювання самооцінних емоцій за рахунок комбінування статичних та динамічних ознак.

Висновок

Досліджено методологію злиття ознак, що забезпує можливість врахування як фізіологічних реакцій, так і структурних поведінкових змін у лицьових експресіях, що значно підвищує точність розпізнавання емоційних станів

МОДЕЛЮВАННЯ ЕМОЦІЙ, МУЛЬТИМОДАЛЬНІ ДАНІ, ОДИНИЦІ ДІЇ, ФІЗІОЛОГІЧНІ СИГНАЛИ, РОЗПІЗНАВАННЯ ЕМОЦІЙ, ГЛИБИННЕ НАВЧАННЯ, ЧАСОВІ РЯДИ, НЕЙРОННІ МЕРЕЖІ, КОНТЕКСТУАЛЬНИЙ АНАЛІЗ.

ABSTRACT

Master Thesis: 75 pp., 19 fig., 9 tab., 43 sources.

Topic: Analysis and modeling of emotions in the context of relevant and derived data

The method of the work is the development and justification of multimodal methods for analyzing and modeling emotions using relevant and derived data.

The object of the study is the processes of formation, probability and automatic recognition of human emotional states.

The subject of the study is methods, models and algorithms for multimodal analysis of emotions using physiological signals, single actions, facial expressions and source contextual data.

Research results

The proposed neural network structure in the work provides effective modeling of self-evaluated emotions for combining static and dynamic features.

Conclusion

The feature fusion methodology has been studied, which provides the ability to take into account both physiological reactions and structural changes in behavior in personal expressions, which significantly increases the accuracy of recognizing emotional states.

EMOTIONS MODELING, MULTIMODAL DATA, ACTION UNITS, PHYSIOLOGICAL SIGNALS, EMOTIONS RECOGNITION, DEEP LEARNING, TIME SERIES, NEURAL NETWORKS, CONTEXTUAL ANALYSIS.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

AAM – Active Appearance Model

AU – Action Unit

AUC – Area Under the Curve

cGAN – Conditional Generative Adversarial Network

CPM – Confidence Preserving Machine

D-PAttNet – Dynamic Patch-Attentive Network

EDA – Electrodermal Activity

FACS – Facial Action Coding System

FC – Fully Connected

FPS – Frames Per Second

RNN – Recurrent Neural Network

SAANet – Siamese Action Unit Attention Network

VAS – Visual Analog Scale

ВСТУП

Актуальність теми.

Сучасні дослідження у сфері штучного інтелекту та обробки мультимодальних даних дедалі більше фокусуються на моделюванні емоційних процесів як одного з ключових аспектів людсько-комп'ютерної взаємодії. Емоції, будучи фундаментальними механізмами регуляції поведінки, сприйняття та прийняття рішень, визначають якість комунікації між людиною та інтелектуальними системами, зумовлюючи ефективність адаптивних інтерфейсів, розумних агентів, медичних систем підтримки рішень та технологій добробуту. У зв'язку з цим аналіз, моделювання та автоматичне розпізнавання емоцій набувають особливої актуальності. Проте сучасні підходи досі стикаються з низкою обмежень, пов'язаних із варіативністю людської експресії, складністю інтерпретації спонтанних реакцій та різнорідністю джерел даних, що потребує розвитку нових методів інтеграції та інтерпретації релевантних і похідних ознак.

Мультимодальний підхід до вивчення емоцій дозволяє об'єднати інформацію з різних каналів — виразів обличчя, одиниць дії (Action Units), фізіологічних сигналів, поведінкових патернів і контексту взаємодії — що забезпечує суттєве підвищення точності емоційних моделей у порівнянні з мономодальними системами. Особливої уваги потребують спонтанні експресії та самооцінні емоції, які проявляються більш тонко і не завжди піддаються прямому візуальному або поведінковому аналізу, а також специфічні стани, такі як емоційно-зумовлений біль. Такі завдання потребують ґрунтовного аналізу структури експресивних одиниць, вивчення контекстуальних факторів, побудови надійних методів екстракції ознак та розробки архітектур нейронних мереж, здатних до моделювання складної мультимодальної динаміки.

Крім того, сучасні набори даних, такі як BP4D+ чи EMOPAIN, містять різнорідні типи інформації — відеодані, маркери AU, сенсорні фізіологічні

показники та самооцінні шкали — що робить можливим проведення комплексного аналізу взаємозв'язків між фізичними, когнітивними та суб'єктивними аспектами емоцій. Це, у свою чергу, створює передумови для розробки нових методологій моделювання емоцій як багатовимірних психологічних конструкцій, а також для формування узагальнених моделей, здатних працювати у реальних умовах взаємодії.

У даній роботі реалізовано системний підхід до аналізу, моделювання та інтеграції емоційних ознак на основі релевантних та похідних даних, що включає дослідження одиниць дії, контекстуальних факторів, фізіологічних параметрів, а також суб'єктивних самооцінок. Особлива увага приділяється проблемам синхронізації модальностей, формуванню комбінованих просторово-часових представлень та створенню ефективних архітектур глибинного навчання, здатних до узгодженого злиття різних інформаційних потоків. Результати дослідження спрямовані на вирішення актуальних завдань у галузях автоматичного розпізнавання емоцій, медичної діагностики, адаптивних систем взаємодії та інтелектуальних інтерфейсів.

Таким чином, дана робота спрямована на розробку методологічних та практичних основ, необхідних для побудови адаптивних і надійних систем аналізу емоцій, здатних функціонувати з урахуванням мультимодальної природи емоційної інформації.

Актуальність дослідження зумовлена стрімким розвитком технологій штучного інтелекту і зростанням попиту на системи, що можуть інтерпретувати емоційний стан людини з високою точністю. У сферах медицини, психології, телеоператорних систем, освітніх платформ та безконтактної діагностики розпізнавання емоцій є ключовим компонентом для покращення якості взаємодії та прийняття рішень. Проте сучасні системи здебільшого базуються на одномодальних підходах, які не враховують складну природу емоцій, їхній контекст, фізіологічні прояви та суб'єктивні самооцінки. Це значно обмежує їхню ефективність та здатність до узагальнення.

Мультимодальний аналіз емоцій, що поєднує вирази обличчя, одиниці дії, фізіологічні сигнали та контекстуальні дані, дозволяє отримати більш повне уявлення про емоційний стан людини. При цьому особливої актуальності набувають методи глибинного навчання, здатні моделювати складні патерни та взаємозв’язки між різними видами даних. Додатковим аргументом на користь актуальності є зростання інтересу до систем, що здатні ідентифікувати не лише базові, але й самооцінні емоції та такі стани, як біль, що мають важливе значення у медичних застосуваннях.

Недостатність існуючих моделей у контексті обробки спонтанних експресій, проблеми дисбалансу наборів даних і потреба у часовій синхронізації сигналів визначають необхідність створення нових методологічних рішень. У цьому контексті дослідження, орієнтоване на інтеграцію релевантних та похідних даних, є важливим кроком до формування більш точних, інтерпретованих і стійких моделей емоцій.

Розвиток сучасних систем штучного інтелекту та людино-машинної взаємодії (HCI) вимагає переходу від простої детекції облич до глибинного розуміння емоційних станів людини. Афективні обчислення мають значний потенціал для застосування у медицині (моніторинг болю, психологічна реабілітація), освіті, системах безпеки та персоналізованих сервісах. Проте, незважаючи на значний прогрес у цій галузі, існуючі методи мають низку фундаментальних обмежень, вирішення яких визначає актуальність даного дослідження.

По-перше, проблема розбіжності між експресивними та відчутими емоціями. Більшість сучасних алгоритмів (state-of-the-art) фокусуються на розпізнаванні експресивних емоцій (те, що показує обличчя), ігноруючи самооцінні емоції (те, що суб'єкт відчуває насправді). Психологічні дослідження свідчать, що одна й та сама експресія (наприклад, усмішка) може супроводжувати діаметрально протилежні стани: радість, збентеження, страх або навіть фізичний біль. Розробка методів, здатних розрізняти ці стани

шляхом аналізу самооцінних звітів та їх кореляції з мікрорухами м'язів, є критично важливою для створення емпатичних систем AI.

По-друге, нехтування часовою динамікою та контекстом. Емоція є динамічним процесом, а не статичною подією, зафіксованою в одному кадрі. Традиційні методи часто аналізують пікові моменти експресії, втрачаючи інформацію про фази початку, розвитку та загасання емоції. Крім того, згідно з принципом контекстуальності (наприклад, у філософії Фреге та психології Барретт), емоцію неможливо коректно інтерпретувати ізольовано від ситуації. Актуальність полягає у створенні нових архітектур (зокрема, на базі 3D CNN та аналізу патернів AU), які інтегрують часову інформацію та контекстуальні ознаки для підвищення точності розпізнавання.

По-третє, необхідність об'єктивізації оцінки болю. Оцінка больового синдрому в клінічних умовах часто базується на суб'єктивних самозвітах, що може бути ненадійним або неможливим для певних категорій пацієнтів. Мультимодальні підходи, що поєднують аналіз лицьових одиниць дії (FACS) із фізіологічними сигналами (частота серцевих скорочень, електродермальна активність тощо), демонструють значний потенціал. Дослідження методів злиття (fusion) цих модальностей та їх синхронізації є актуальним завданням для створення об'єктивних діагностичних інструментів.

Таким чином, актуальність даної роботи обумовлена необхідністю розробки комплексних, контекстно-обізнаних та мультимодальних методів аналізу, які здатні враховувати часову природу емоцій та мінімізувати розрив між автоматичною детекцією експресії та реальним досвідом людини.

Метою роботи є розроблення та обґрунтування мультимодальних методів аналізу й моделювання емоцій із використанням релевантних та похідних даних.

Об’єктом дослідження є процеси формування, вираження та автоматичного розпізнавання емоційних станів людини.

Предметом дослідження є методи, моделі та алгоритми мультимодального аналізу емоцій із використанням фізіологічних сигналів, одиниць дії, виразів обличчя та похідних контекстуальних даних.

Завдання дослідження:

1. Проаналізувати теоретичні засади моделювання емоцій та структуру емоційної експресії.

2. Дослідити роль одиниць дії, часової динаміки та контексту у формуванні емоційних реакцій.

3. Виконати аналіз існуючих наборів даних, орієнтованих на розпізнавання складних і самооцінних емоцій.

4. Розробити методологію мультимодального моделювання з урахуванням релевантних та похідних даних.

5. Запропонувати архітектуру нейронної мережі для розпізнавання самооцінних емоцій.

Методи дослідження

У роботі застосовано методи машинного навчання та глибинного навчання, мультимодальної обробки сигналів, аналізу часових рядів, методи екстракції ознак з лицьових експресій, статистичні методи оцінювання достовірності результатів, а також експериментальні методи порівняння ефективності алгоритмів.

Наукова новизна отриманих результатів

Наукова новизна роботи полягає у розробці комплексної мультимодальної методології моделювання емоцій, яка інтегрує релевантні та похідні дані та дозволяє підвищити точність розпізнавання складних емоційних станів. Рапропоновано підхід, що поєднує часову динаміку одиниць дії, самооцінні емоції та фізіологічні сигнали у єдиному аналітичному процесі. Отримано нові результати щодо ефективності мультимодальних моделей у задачах розпізнавання болю, що дозволило встановити їхню перевагу над одномодальними підходами.

Практичне застосування результатів

Результати дослідження можуть бути застосовані у медичних діагностичних системах для виявлення болю та емоційного дискомфорту, у психологічних та соціальних дослідженнях для автоматичного аналізу емоційної поведінки, у системах людино-комп'ютерної взаємодії для адаптації інтерфейсу під емоційний стан користувача. Запропоновані мультимодальні методики можуть бути інтегровані у системи моніторингу пацієнтів, робототехнічні платформи, віртуальних агентів та освітні середовища, де важливо забезпечити емоційно-чутливу взаємодію.

Структура магістерської роботи. Представлена робота складається зі вступу, трьох розділів та висновків. Загальний обсяг роботи становить 75 сторінок, і містить 19 рисунків, 9 таблиць, перелік використаних джерел із 43 позицій.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ МОДЕЛЮВАННЯ ЕМОЦІЙ З ВИКОРИСТАННЯМ МУЛЬТИМОДАЛЬНИХ ДАНИХ

1.1. Аналіз та моделювання емоційних реакцій

Людські емоції є фундаментальними складовими комунікації та передачі внутрішніх станів. Характер та інтенсивність їхнього вираження демонструють значну міжіндивідуальну варіативність. Це дисертаційне дослідження присвячене емпіричному та теоретичному аналізу двох ключових категорій емоційних проявів: експресивних (виражених) та самооцінних (відчутих) емоцій.

Експресивні емоції визначаються як емоційні стани, які об'єктивуються через лицьові експресії. На противагу цьому, самооцінні емоції є суб'єктивними внутрішніми станами, які вимірюються шляхом прямого, ручного звітування суб'єкта про його/її актуальні емоційні переживання. Обидва типи емоцій тісно пов'язані з контекстом та характером стимульного впливу на суб'єкта.

Сучасні передові методики (state-of-the-art) розпізнавання емоцій переважно орієнтовані на детекцію експресивних лицьових емоцій, часто ігноруючи самооцінні емоції та контекстуальні чинники. Однак, нещодавні публікації переконливо продемонстрували, що контекст відіграє критичну роль в емоційному аналізі, і є необхідним для повного розуміння емоційних процесів.

Крім того, значна частина поточних досліджень сфокусована на моментальних, статичних виразах. Натомість, емоційна реакція на стимули має природну часову динаміку (temporal nature) і є процесом, а не миттєвою подією.

Дана робота дисертація пропонує методологічні підходи, що використовують тимчасову інформацію для підвищення ефективності розпізнавання самооцінних емоцій та контексту.

Паралельно з детекцією загальних лицьових виразів, дослідження також охоплює аналіз одиниць дії (Action Units, AUs) лиця [1], які відповідають специфічним рухам лицьових м'язів. Більшість робіт зосереджена на детекції виникнення AU, тоді як аналіз їхньої динаміки та використання після детекції залишається обмеженою областю. Оскільки динаміка лицьових експресій є тимчасовою, відповідні м'язові рухи (AUs) також демонструють тимчасовий профіль, який включає фази початку, пікової інтенсивності та завершення. Ця тимчасова природа AU, у поєднанні з їхньою спільним виникненням (co-occurrence), відкриває нові перспективи та наукові виклики для досліджень одиниць дії.

Загальна мета полягає у систематичному дослідженні тимчасової та контекстуальної природи емоцій на основі сучасних емпіричних наборів даних.

Основні результати магістерської роботи:

- Розроблено методологію, яка встановлює взаємозв'язок між лицьовими одиницями дії та досвідом/контекстом суб'єкта.

- Створено підхід до детекції самооцінних емоцій на основі аналізу лицьової експресії.

Ця робота пропонує інтегровані підходи для вирішення обох цих складних та критично важливих проблем в афективних обчисленнях.

Це дослідження зосереджено на вирішенні низки ключових викликів у сфері розпізнавання емоцій:

Питання 1. Розпізнавання самооцінних емоцій

Більшість сучасних робіт орієнтується на розпізнавання або експресивних лицьових емоцій, або індукованих емоцій. Люди переживають множинні та диференційовані емоції у відповідь на різні стимули, і дослідження з детекції самооцінних емоцій є вкрай недостатніми. Самооцінні емоції є найбільш безпосереднім виміром суб'єктивного емоційного досвіду.

Відкрите питання: чи можливо розпізнати множинні самооцінні емоції за допомогою тимчасової інформації лицьової експресії?

Питання 2. Дисбаланс даних у одиницях дії (AU)

У машинному навчанні розподіл даних є критичним для прогностичної точності. Частотний розподіл виникнення AU є незбалансованим, що негативно впливає на оціночні метрики (наприклад, F1-показник). Цей дисбаланс призводить до того, що більшість сучасних методів демонструють низьке стандартне відхилення (схожу точність) та сильно корелюють із розподілом виникнення AU. Пропонується методологія для вирішення складної проблеми дисбалансу даних у детекції одиниць дії.

Питання 3. Мультимодальне розпізнавання болю

Розпізнавання болю є важливим клінічним завданням. Переважна більшість обнадійливих робіт використовує єдину модальність (наприклад, лицьову експресію). Було показано, що використання кількох модальностей підвищує точність систем розпізнавання.

Відкрите питання: хи можуть мультимодальні системи підвищити точність розпізнавання болю для всіх вікових груп, і чи існує кореляція між модальностями? Аналіз кореляцій між модальностями може надати нові інсайти для розробки більш надійних систем.

Питання 4. Моделювання тимчасової динаміки AU для контекстуалізації

Незважаючи на значні успіхи в детекції AU, використання виявлених AU залишається недостатньо дослідженим відкритим питанням. Ця робота досліджує інноваційний підхід до моделювання тимчасової динаміки одиниць дії з метою автоматичного розпізнавання контексту, тобто досвіду суб'єкта.

1.2. Дослідження факторів впливу на розпізнавання емоцій

Лицьові експресії є ключовим каналом передачі емоційних станів у людини. Однак, індивідуальні біосоціальні характеристики, такі як етнічна

приналежність, стать та вік, можуть суттєво впливати на точність розпізнавання емоцій.

Ця проблема ускладнюється нерівномірним географічним та демографічним розподілом етнічних груп населення планети, що створює методологічні труднощі для збору збалансованих наборів даних для міжкультурного аналізу емоцій. Обмеження у формуванні репрезентативних вибірок може призвести до недостатньої генералізації методів детекції емоцій, що унеможливлює створення єдиної моделі, ефективної для всіх етнічних груп.

Однією з ключових методологій для об'єктивного представлення лицьових експресій є система кодування рухів обличчя (Facial Action Coding System, FACS) [34], яка кодує рухи м'язів обличчя через одиниці дії (Action Units, AUs). Емоційний вираз, як правило, активує одну або комбінацію AU, що дозволяє FACS репрезентувати широкий спектр лицьових експресій. Успішне дослідження цієї проблематики вимагає використання великих та різноманітних наборів даних. На щастя, було зібрано багато передових наборів даних (state-of-the-art) для детекції емоцій, що дозволило досягти значних результатів.

1.2.1. Аналіз одиниць дії (AU)

У контексті AU, в [11] провели аналіз міжкультурних та культурно-специфічних аспектів сприйняття та вираження лицьових емоцій. Вони виявили, що, хоча експресія та сприйняття деяких емоцій є подібними, існують статистично значущі відмінності для інших емоцій між культурами. Ці результати свідчать про потенційну варіативність активації AU для одних і тих же емоцій у різних культурних середовищах.

Інші дослідники також активно працюють над вдосконаленням детекції AU. В [12] застосували архітектуру, що поєднує згорткові нейронні мережі (CNN) та мережі довгої короткочасної пам'яті (LSTM) для трикласової класифікації AU, використовуючи як просторові, так і часові ознаки.

Запропонована мережа є привабливою завдяки природному моделюванню трьох комплементарних аспектів. Загальний огляд запропонованого фреймворку подано на рис. 1.1.

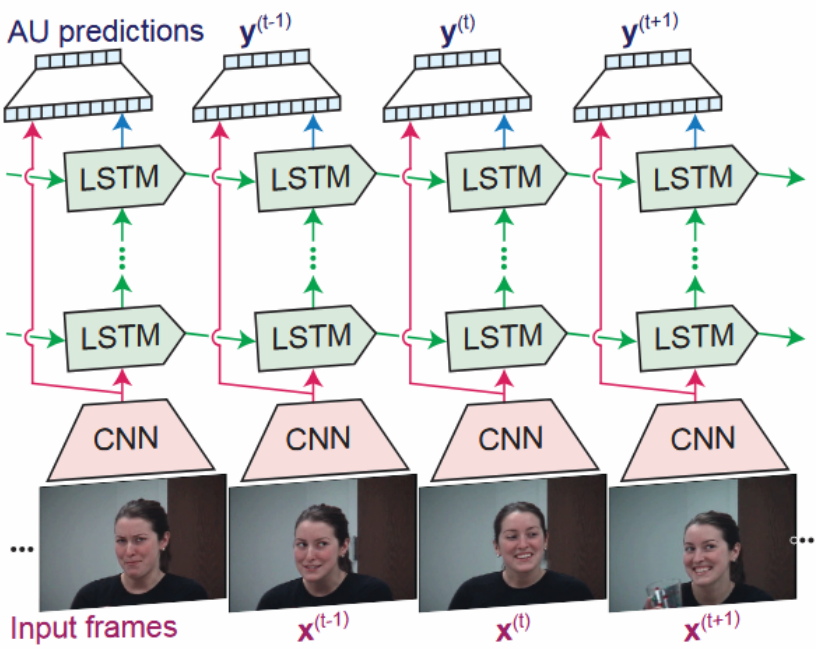


Рис. 1.1. Загальний огляд гібридного фреймворку глибокого навчання

Дана мережа поєднує переваги CNN та LSTMs для моделювання та використання як просторових, так і часових ознак. Далі використовується мережа злиття для комбінування обох ознак з метою покадрового прогнозування.

Для навчання генералізованого представлення залучається згорткова нейронна мережа (CNN), натренована для екстракції просторових ознак. Такі ознаки мінімізують поширені персоніфіковані упередження, властиві ознакам, отриманим вручну (hand-crafted features) і таким чином, надають можливість знизити навантаження на розробку складних класифікаторів.

Для захоплення часових залежностей (temporal dependencies) на просторові ознаки накладаються стекові мережі довгої короткочасної пам'яті (LSTMs). Нарешті, тут гаруються вихідні показники як від CNN, так і від

LSTMs у мережу fusion (злиття) для прогнозування 12 AUs для кожного кадру.

Навчені просторові ознаки, які додатково поєднуються з часовою інформацією, перевершують стандартну CNN та state-of-the-art методи, засновані на ознаках. Крім того, тут виконується візуалізація понять кожної AU, засвоєних моделлю, що, наскільки розкриває, як машини сприймають лицьові AU.

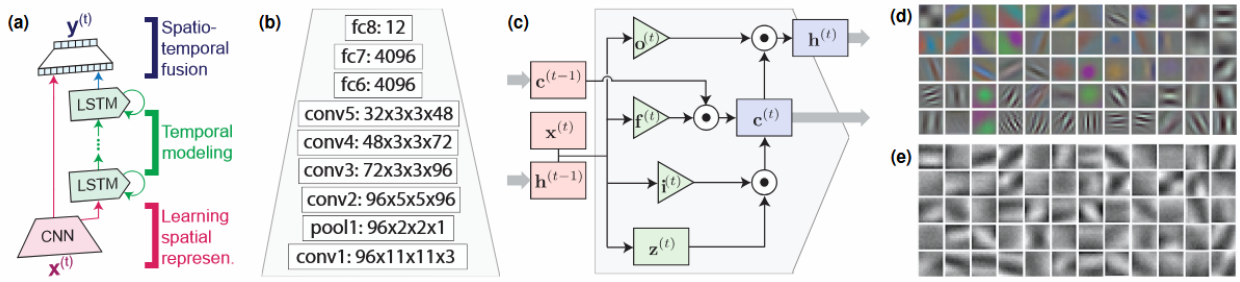


Рис. 1.2. Структура гібридної мережі розпізнавання

На рисунку 1.2 подано a) згорнута ілюстрація рисунку 1.1, що демонструє три компоненти навчання: просторове представлення (spatial representation), часове моделювання (temporal modeling) та просторово-часове злиття (spatiotemporal fusion). b) Архітектура 8-шарової згорткової нейронної мережі (CNN). c) Схематичне представлення блоку мережі довгої короткочасної пам'яті (LSTM). d) – e) Візуалізація шарів conv1 моделей, натренованих на наборах даних ImageNet та GFT відповідно. Як можна побачити, фільтри, навчені на нашому наборі даних облич, містять меншу кількість детекторів колірних плям (color blob detectors), що свідчить про меншу інформативність кольору для детекції одиниць дії (AU).

В дослідженні [13] розробили машину збереження впевненості (Confidence Preserving Machine, CPM). Їхня методологія передбачає навчання двох незалежних класифікаторів (позитивного та негативного) з подальшим використанням персоніфікованого класифікатора для детекції AU. В [14] запропонували глибоку архітектуру навчання з тимчасовою fusion-стратегією

для детекції AU, де незалежно навчаються регіони інтересу з локальними CNN, і застосовується мультиміткове навчання.

Незважаючи на ці результати, дисбаланс даних може впливати на процес навчання моделей, ускладнюючи генералізацію детекції емоцій для різних вікових, гендерних та етнічних груп. В [14] також підтвердили складність міждоменного розпізнавання AU (тобто між різними наборами даних). А в [15] представили FACS3D-Net, метод, що інтегрує 2D та 3D згорткові мережі для детекції AU. Вони показали, що комбінування просторової та часової інформації призводить до підвищення ефективності розпізнавання AU.

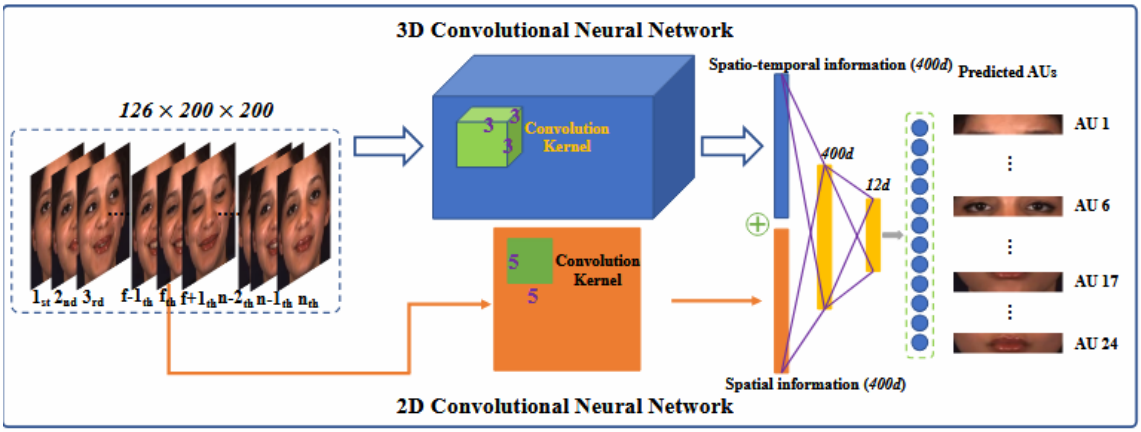


Рис. 1.3. Огляд методу FACS3D-Net для мультиміткової детекції AU

Відеокліп подається на вхід 3D згорткової мережі (3D convolutional network) для вивчення просторово-часової інформації. Отримана інформація потім конкатенується (об'єднується) з просторовою інформацією f-го кадру, отриманою з 2D згорткової мережі (2D convolutional network). Після проходження двох пов'язних шарів (dense layers) ми отримуємо результати детекції 12 одиниць дії (AUs) у f-му кадрі.

Останні роки також відзначені інтенсивними дослідженнями висновування емоцій з лицьових експресій з використанням 2D і 3D інформації. В [14] запропонували деформовану синтетичну модель (Deformable Synthesis Model) для створення синтетичних 3D лицьових даних,

які використовувалися для тренування глибоких нейронних мереж, демонструючи покращену генералізацію експресій на різних передових наборах даних. В [15] розробили підхід, що базується на навчанні залишкових компонентів де-експресії, де виразовий компонент вилучається за допомогою генеративної моделі та використовується для розпізнавання експресій.

1.2.2. Критична роль контексту

Хоча згадані вище роботи досягли значних успіхів, вони не враховували контекстуальні фактори, які, згідно з психологічними дослідженнями, є визначальними для розпізнавання емоцій.

Принцип контексту філософії Фреге [17] – "ніколи не запитувати про значення слова ізольовано, а лише в контексті речення" – слугує мотиваційним підґрунтям для цієї дисертації: емоції нерозривно пов'язані з контекстом.

Автори [18] емпірично продемонстрували, як контекст впливає на сприйняття емоцій. Вони показали, що візуальні сцени, інтонація голосу, культура та присутність інших осіб модифікують емоційне сприйняття суб'єкта. Вони підкреслили значимість емоційних факторів для ментальних процесів, досліджуючи, як психіка індивіда формується через взаємодію з фізичним, соціальним та культурним середовищем. Також показали, що психологічні риси є нейтральними за своєю суттю, а їхні наслідки залежать від контексту.

Крім того, в [2] виділили чотири компоненти, що характеризують емоцію: суб'єкт, стимул, стратегії вибору/імплементації та тип результату і запропонували соціодинамічну модель емоцій, акцентуючи увагу на ролі соціального контексту.

Мотивовані цими психологічними роботами, в [20] представили EmotiCon – метод контекстно-обізнаного прогнозування емоцій. EmotiCon використовує комбінацію лицьових та тілесних ознак, а також семантику

фону та соціальні взаємодії, щоб імітувати, як людина розпізнає емоції, спираючись на всеосяжний контекст ситуації. Вони використовували контекст суб'єкта та стратегію імплементації (оточення та інші особи) для прогнозування емоції.

Основні властивості EmotiCon:

- Контекстно-обізнаний (Context-Aware) - це головна відмінність від традиційних методів, які фокусуються лише на обличчі.

- Мультимодальний (Multimodal) - Інтегрує дані з обличчя та тіла (наприклад, ходи).

- Розпізнавання сприйнятої емоції (Perceived Emotion Recognition) - метод фокусується на розпізнаванні того, як емоція людини сприймається зовнішнім спостерігачем, а не на її фактичному, суб'єктивному емоційному стані.

- Мультиміткова класифікація (Multi-label Classification) - вихід моделі є класифікацією емоцій з можливістю присвоєння декількох міток (емоцій) для одного вхідного кадру/зображення [1.1].

Автор продемонстрував, що розпізнавання емоції іншою особою є ускладненим лише за лицьовою експресією. Наприклад, вираз обличчя футболіста одразу після забитого гола може бути інтерпретований як гнів. Проте, надання повного контексту (тобто факту забитого гола) дозволяє впевнено класифікувати емоцію як радість. Це підтверджує критичну необхідність інтеграції контексту та всієї сцени в системи машинного навчання для прогнозування емоцій. Наскільки нам відомо, відсутні роботи, які б враховували контекст для прогнозування емоцій. Ці висновки обґрунтовують дане дослідження контексту.

Поряд із контекстом, упередженість (bias) та вплив гендеру є важливими факторами в емоційному аналізі. В [21] закликають до підвищення справедливості штучного інтелекту (AI). Вони підкреслюють, що проблема значною мірою спричинена тим, що більшість тренувальних даних для AI збирається в Сполучених Штатах. AI засвоює не лише ідентифікаційні

ознаки з даних, але й упередження та розподіл даних. Цей ефект був підтверджений в [15], які проаналізували точність гендерних класифікаторів для осіб з різними демографічними ознаками. Вони виявили, що найкраща продуктивність гендерних класифікаторів спостерігалася для світлошкірих чоловіків, а найгірша – для темношкірих жінок.

1.3. Методологія автоматичного розпізнавання спонтанних лицьових одиниць дії

1.3.1. Набір даних

У цьому дослідженні використано набір даних спонтанних лицьових експресій, отриманих від суб'єктів у природних умовах. Набір даних мав обсяг 300 ГБ і складався з кольорових зображень із роздільною здатністю 640 × 480 пікселів, глибиною кольору 8 біт на піксель та частотою кадрів 60 полів на секунду (черезрядкова розгортка 2:1).

Відеопослідовності включали обертання голови поза площиною до 75 градусів. До вибірки увійшли 17 суб'єктів: 3 азіати, 3 афроамериканці та 11 європеоїдів. Троє суб'єктів носили окуляри. Лицьова поведінка (протягом однієї хвилини відео на суб'єкта) була оцінена кадр за кадром двома командами експертів за системою FACS (система кодування рухів обличчя).

Незважаючи на значний розмір бази даних за сучасними стандартами цифрового відеозберігання, кількість спонтанних прикладів для кожної одиниці дії (AU) була відносно невеликою. Тому прототипування системи було виконано на трьох одиницях дії з найбільшою кількістю прикладів:

- Моргання (AU 45): 168 прикладів від 10 суб'єктів.
- Підняття брів (AU 1 + 2): 48 прикладів від 12 суб'єктів.
- Опускання брів (AU 4): 14 прикладів від 12 суб'єктів.

Негативні приклади для кожної категорії були сформовані шляхом випадкового відбору послідовностей, узгоджених за суб'єктом та довжиною

послідовності. Ці три лицьові дії є релевантними для таких застосувань, як моніторинг пильності, тривожності та сплутаності свідомості.

Представлена система використовує загальноцільові механізми навчання, які можуть бути застосовані для розпізнавання будь-якої лицьової дії, за умови наявності достатнього обсягу тренувальних даних. Це усуває необхідність розробки спеціалізованих метрик ознак для розпізнавання додаткових лицьових дій.

1.3.2. Система розпізнавання

Загальний огляд архітектури системи розпізнавання проілюстровано на рисунках 1.4 - 1.5.

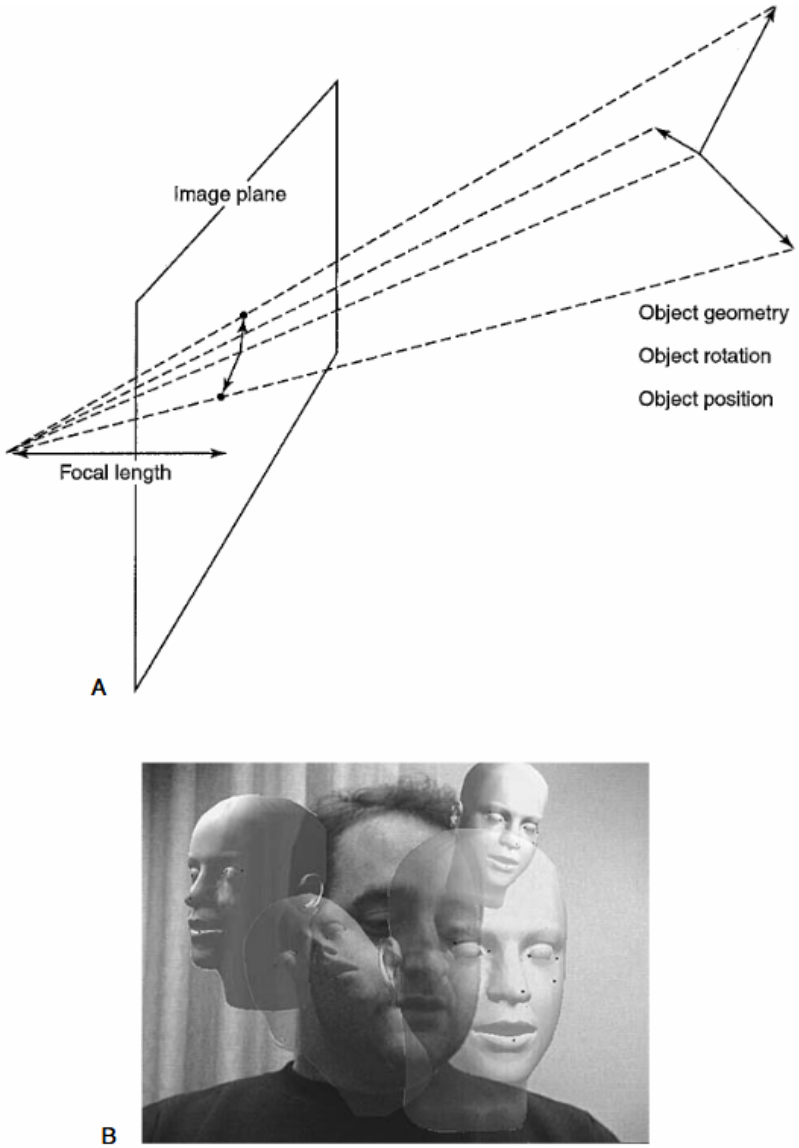


Рис. 1.4. Оцінка пози голови

А. Спершу, параметри камери та геометрія обличчя спільно оцінюються за допомогою ітераційного методу найменших квадратів (iterative least squares technique).

В. Далі, поза голови оцінюється в кожному кадрі з використанням стохастичної частинкової фільтрації (stochastic particle filtering). Кожна частинка (particle) являє собою модель голови з певною орієнтацією та масштабом.

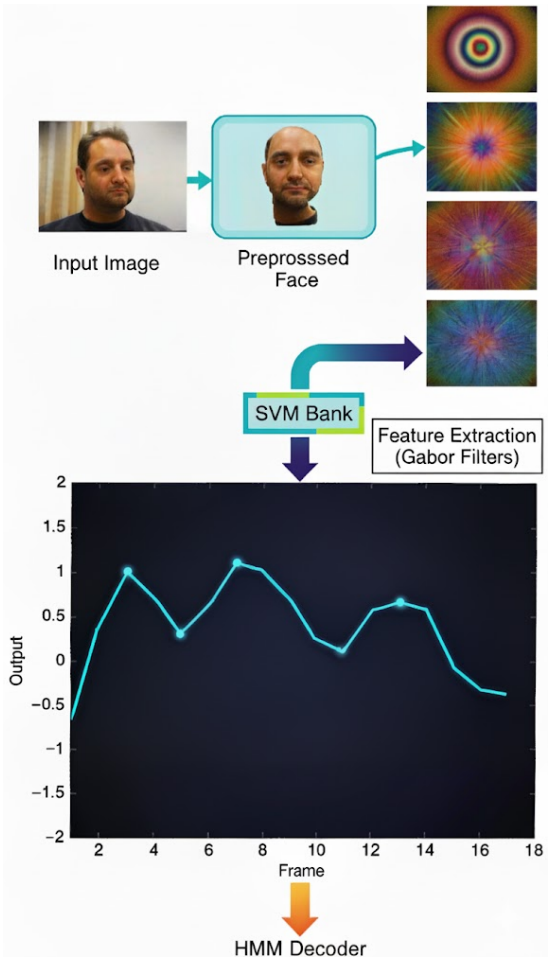


Рис. 1.5. Блок-схема системи розпізнавання

Згідно блок-схеми на рис. 1.5:

1. Спочатку виконується оцінка пози голови (Head Pose Estimation).

2. Зображення трансформуються (warped) до фронтального вигляду та канонічної геометрії обличчя.

3. Трансформовані зображення потім пропускаються через банк фільтрів Габора (Gabor filters).

4. Далі машини опорних векторів (SVMs) тренуються для класифікації лицьових дій на основі Габорового представлення в окремих відеокадрах.

5. Вихідні траєкторії SVM для повних відеопослідовностей спрямовуються до прихованих марковських моделей (hidden Markov models, HMMs)

Спочатку виконувалася оцінка пози голови (Head Pose Estimation) у відеопослідовностях із використанням частинкового фільтра (particle filter) зі 100 частинками.

Далі зображення обличчя вирівнювалися (warped) на канонічну модель обличчя з фіксованою геометрією, поверталися до фронтального вигляду, а потім проєктувалися назад у площину зображення. Це вирівнювання використовувалося для визначення та кадрування підрегіону обличчя, що містить очі та брови. Вертикальна позиція очей становила 0,67 висоти вікна; відстань між очима — 105 пікселів; відстань від очей до рота — 120 пікселів. Яскравість пікселів була лінійно перемасштабована до діапазону [0,255]. Потім була застосована м'яка гістограмна еквілізація для рівнів сірого шляхом застосування логістичного фільтра з параметрами, підібраними для узгодження середнього значення та дисперсії рівнів сірого в нейтральному кадрі.

1.3.3. Екстракція ознак та класифікація

Отримані зображення були згорнуті банком фільтрів Габора (Gabor kernels) з п'ятьма просторовими частотами та вісьмома орієнтаціями. Вихідні величини (магнітуди) були нормалізовані до одиничної довжини, а потім дискретизовані зі зменшенням розміру в чотири рази.

Отримані габорові представлення були передані до банку машин опорних векторів (Support Vector Machines, SVMs). Нелінійні SVM були навчені розпізнавати лицьові дії в окремих відеокадрах. Тренувальні зразки для SVM відповідали пікам активності дії, ідентифікованим експертами

FACS, а негативні приклади були випадково відібраними кадрами, узгодженими за суб'єктом.

Для тестування генералізації на нових суб'єктах використовувалася перехресна валідація з виключенням одного (leave-one-out cross-validation). Вихід SVM (який є відстанню вздовж нормалі до розділової гіперплощини, тобто маржею) формував траєкторії виходів для повних відеопослідовностей тестових суб'єктів.

Ці траєкторії потім передавалися прихованим марковським моделям (Hidden Markov Models, HMMs). HMM — це імовірнісні динамічні моделі, які навчаються розподілам ймовірностей послідовностей. Вони є домінантним підходом у сучасних системах розпізнавання мовлення.

HMM були навчені моделювати послідовності виходів SVM, типові для кожної AU. Для кожної окремої AU тренувалася одна HMM, отже, кожна HMM виступала як експерт для відповідної AU (подібно до того, як кожна HMM у розпізнаванні мовлення стає експертом для певної фонеми).

На етапі тестування нова послідовність подавалася до кожної HMM для оцінки правдоподібності (likelihood) цієї послідовності за умови кожної потенційної AU. Обиралася та AU, яка відповідала HMM з максимальною правдоподібністю. Важливо, що цей підхід класифікує лицьові дії без використання інформації про те, який саме кадр містив пік активності дії. Генералізація на нових суб'єктів знову тестувалася з використанням перехресної валідації з виключенням одного.

Висновки до розділу

В даному розділі проведений аналіз предметної області дозволив встановити, що моделювання емоцій є багатокомпонентним завданням, яке потребує урахування фізіологічних, поведінкових та контекстуальних чинників. Розгляд теоретичних основ формування емоцій показав, що емоційна експресія не може бути адекватно інтерпретована лише на основі

статичних лицьових ознак, оскільки вона залежить від інтенсивності стимулу та індивідуальних особливостей суб’єкта. Дослідження одиниць дії засвідчило їхню значущість як структурованих маркерів мікрокомпонентів експресій, здатних відображати як базові, так і складні емоційні реакції. Разом із тим було встановлено, що AU-патерни потребують аналізу у часовій динаміці, оскільки тривалість та послідовність активацій можуть суттєво впливати на точність класифікації. Особливу увагу приділено ролі контексту, який виявився критично важливим для коректного розпізнавання емоцій, адже ситуаційні, соціальні та середовищні умови формують різні траєкторії експресивної поведінки.

РОЗДІЛ 2. МЕТОДОЛОГІЯ РОЗПІЗНАВАННЯ САМООЦІННИХ ЕМОЦІЙ ЗА ВИРАЗАМИ ОБЛИЧЧЯ

2.1. Аналіз самооцінних емоцій та їх зв'язок із лицьовою експресією

Останніми роками спостерігається інтенсивний розвиток досліджень, спрямованих на висновування емоційних станів людини за допомогою лицьових експресій, з використанням як 2D, так і 3D інформації.

Як було описано в попередньому розділі в [15] запропонували використання архітектури FACS3D-Net для детекції одиниць дії (Action Units, AU). Їхній метод інтегрує 2D та 3D згорткові нейронні мережі (CNN) для виконання цього завдання та продемонстрував, що поєднання просторової та часової інформації призводить до підвищення ефективності детекції AU. В [2] створили синтетичні 3D лицьові дані, які використовувалися для тренування глибоких нейронних мереж. Вони показали, що застосування цих синтетичних тренувальних даних сприяє генералізації детекції лицьових експресій на різних передових наборах даних (state-of-the-art datasets). Крім того, вони представили метод навчання залишкових компонентів де-експресії, який розпізнає лицьові експресії шляхом вилучення виразового компонента за допомогою генеративної моделі.

Незважаючи на обнадійливі результати в розпізнаванні емоцій на основі лицьових експресій, численні дослідження показали, що експресії суттєво варіюються між культурами та окремими індивідами. Виявили, що, хоча існують докази, які підтверджують загальноприйняту відповідність між емоцією та експресією (наприклад, усмішка при радості та нахмурення при смутку), способи їхньої комунікації значно різняться навіть у межах однієї й тієї ж ситуації. Також було продемонстровано, що в лицьових експресіях може виникати множинність емоцій (multiple emotions).

З огляду на це, критично важливим є дослідження суб'єктивно пережитих емоцій (тобто самооцінних емоцій, self-report emotion), щоб глибше зрозуміти зв'язок між експресіями та внутрішнім емоційним станом суб'єкта. Більшість робіт із розпізнавання емоцій зосереджені на детекції емоції, яка мала бути викликана стимулом. Однак, в [22] розпочали дослідження зв'язку між самооцінними емоціями суб'єктів та лицьовими експресіями, зосередившись на емоціях, які відчуваються під час виникнення усмішки. Вони встановили, що усмішка виникає при таких емоціях, як веселощі, збентеження, страх та біль, причому ці усмішки візуально відрізнялися, про що свідчили вимірювання одиниць дії. Ця робота стала мотивацією для нашого поточного дослідження самооцінних емоцій суб'єктів.

Оскільки було доведено, що експресії варіюються між людьми, в лицьових експресіях може виникати множинність емоцій, і пряме висновування емоції з експресії є складним (рис. 2.1), це спонукало дослідити самооцінені емоції суб'єктів у їхньому зв'язку з лицьовими експресіями.

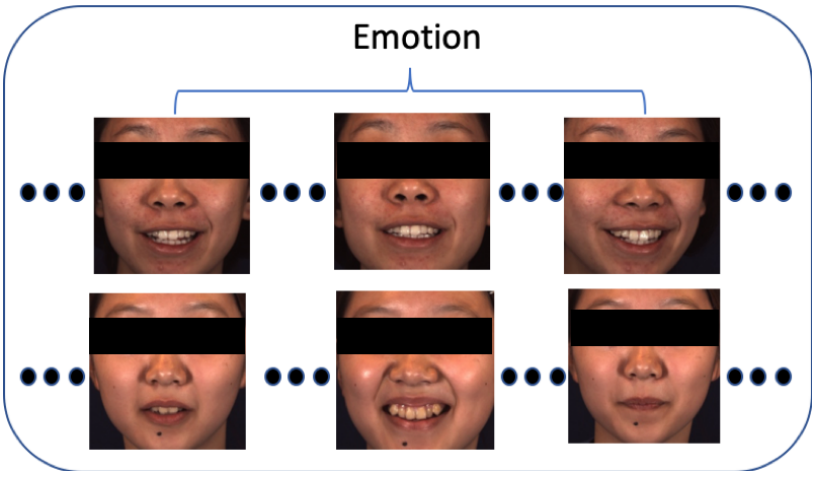


Рис. 2.1. Варіативність експресій та самооцінних емоцій

На рис. 2.1 різні суб'єкти виконують одне й те саме завдання, призначене для викликання радості/веселощів. Самооцінні емоції: (вгорі) –

розслабленість: 5; веселощі: 4; співчуття: 2; зляканість: 0; здивування: 1. (внизу) – розслабленість: 1; веселощі: 1; співчуття: 0; зляканість: 2; здивування: 1. Усі інші самооцінні емоції дорівнюють 0 для обох суб'єктів.

В даній роботі ми проводимо аналіз самооцінних звітів суб'єктів на основі мультимодального корпусу спонтанних емоцій BP4D+ , досліджуючи їхній розподіл за гендером у співвідношенні до емоції, яка мала бути викликана завданням. Ми пропонуємо архітектуру 3D CNN для розпізнавання множинних самооцінних емоцій на основі лицьових експресій.

2.2. Аналіз самооцінних емоцій у корпусі даних BP4D+

2.2.1. Характеристика набору даних

Для емпіричного дослідження було використано мультимодальний корпус спонтанних емоцій BP4D+.

Набір даних включає 140 суб'єктів (58 чоловіків та 82 жінки) віком від 18 до 66 років, що представляють різні етнічні групи, зокрема європеоїдну, афроамериканську, азіатську та латиноамериканську.

Корпус містить різноманітні модальності даних:

- Відеоінформація: 2D та теплові зображення.
- Геометрична інформація: 3D моделі, 2D та 3D орієнтири обличчя.
- Поведінкові дані: Одиниці дії (AU) за системою FACS.
- Фізіологічні дані.

Для індукції конкретних емоційних станів було розроблено десять стандартизованих завдань (таблиця 2.1). Для 138 суб'єктів було зібрано самооцінні звіти про емоції, які вони фактично відчували під час кожного завдання. Суб'єктам було дозволено обирати множинні емоції (наприклад, розслабленість, здивування, сум, щастя) для кожного завдання. Інтенсивність самооцінних емоцій вимірювалася за 5-бальною шкалою Лайкерта. Наш подальший аналіз ґрунтується на даних від усіх 138 суб'єктів, які надали самооцінки.

Таблиця 2.1.

Опис емоційно-індукційних завдань корпусу BP4D+

Завдання	**Діяльність**	**Цільова Емоція**
T1	Інтерв'ю: Прослуховування гумористичного жарту	Щастя
T2	Графічне шоу: Перегляд 3D-аватара учасника	Здивування
T3	Відеокліп: Дзвінок на екстрену лінію 911	Сум
T4	Відчуття раптового акустичного спалаху	Переляк або здивування
T5	Інтерв'ю: Питання типу "правда чи брехня"	Скептицизм
T6	Імпровізація гумористичної пісні	Сором
T7	Переживання фізичної загрози в грі з дротиками	Страх або нервозність
T8	Холодний пресор: Занурення руки в крижану воду	Фізичний біль
T9	Інтерв'ю: Отримання скарги на низьку продуктивність	Гнів або розчарування
T10	Відчуття неприємного запаху	Огида

2.2.2 Аналіз самооцінок суб'єктів

Натхненні дослідженнями, які вказують на можливість переживання множинних емоцій під час лицьової експресії, ми дослідили розподіл самооцінних емоцій суб'єктів у всіх завданнях BP4D+.

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Relaxed	0.63	0.25	0.01	0.00	0.09	0.09	0.04	0.07	0.06	0.03
Amused	0.96	0.50	0.00	0.12	0.14	0.43	0.25	0.02	0.14	0.01
Disgusted	0.00	0.04	0.04	0.00	0.00	0.02	0.01	0.00	0.01	0.97
Afraid	0.01	0.00	0.53	0.35	0.01	0.02	0.71	0.12	0.15	0.13
Angry	0.00	0.00	0.42	0.12	0.04	0.01	0.04	0.09	0.47	0.09
Frustrated	0.01	0.00	0.06	0.04	0.06	0.12	0.02	0.08	0.41	0.04
Sad	0.00	0.01	0.65	0.00	0.00	0.01	0.00	0.00	0.09	0.00
Sympathetic	0.05	0.00	0.36	0.01	0.01	0.01	0.01	0.00	0.07	0.01
Nervous	0.22	0.05	0.27	0.10	0.09	0.40	0.55	0.12	0.33	0.16
Pained	0.00	0.01	0.02	0.04	0.00	0.00	0.02	0.99	0.00	0.04
Embarrassed	0.05	0.32	0.00	0.08	0.03	0.94	0.01	0.00	0.27	0.02
Startled	0.03	0.17	0.37	0.99	0.09	0.02	0.33	0.17	0.20	0.17
Surprised	0.23	0.81	0.09	0.64	0.46	0.17	0.31	0.12	0.19	0.12
Skeptical	0.04	0.03	0.01	0.00	0.98	0.04	0.13	0.03	0.28	0.04

Рис. 2.2. Відсоток суб'єктів, які відчували емоцію в кожному завданні

Як видно з рис. 2.2, значна кількість суб'єктів повідомила про множинні емоції для одного завдання. Наприклад, у завданні Т1 (Щастя), незважаючи на цільову емоцію, значний відсоток суб'єктів відчував не лише забавленість (0.96), але й розслабленість (0.63), здивування (0.23) та нервозність (0.22). Хоча цільова або подібна емоція фіксується найбільшим відсотком суб'єктів (наприклад, Огида в Т10 – 0.97, Біль у Т8 – 0.99), спостерігається високе виникнення доповнювальних емоцій.

Особливо виділяється завдання Т3 (сум): тоді як сум відчували 65% суб'єктів, страх (0.53), гнів (0.42), співчуття (0.36), нервозність (0.27) та зляканість (0.37) також фіксувалися з відносно високою частотою. У равданні Т9 (шнів) розчарування (0.41) відчувалося з майже такою ж частотою, як і гнів (0.47). Цей аналіз підтверджує висновки літератури [9] про індивідуальну варіативність емоційних реакцій у стандартизованих ситуаціях. Виникнення множинних емоцій у межах одного завдання може частково пояснювати складність висновування емоцій виключно з лицьових експресій.

Ми також провели незалежний аналіз самооцінок для чоловіків та жінок. Хоча більшість самооцінних емоцій є схожими в обох гендерних групах, виявлено деякі відмінності. Наприклад, у Т3 (сум) сум повідомили 72% жінок проти 55% чоловіків.

Для подальшого дослідження цих відмінностей було оцінено статистичну значущість між гендерними групами.

Статистична значущість різниці у частоті виникнення самооцінних емоцій була оцінена за допомогою парних t-тестів. Статистична значущість різниці у інтенсивності самооцінних емоцій (за 5-бальною шкалою Лайкерта) також була розрахована (таблиця 2.2).

Аналіз частоти виникнення (таблиця 2.3). Для більшості емоцій статистично значущої різниці не виявлено (позначено як 'n.s'). Винятки спостерігаються переважно у Т3 (Сум), Т4 (Переляк) та Т8 (Біль). Наприклад, у Т3 є значна різниця у повідомленні про Сум ($p<0.005$).

Таблиця 2.2.

Значущість різниць між інтенсивністю самооцінних емоцій чоловіків та жінок

Task	Relaxed	Amused	Disgusted	Afraid	Angry	Frustrated	Sad	Sympathetic	Nervous	Pained	Embarrassed	Startled	Surprised	Skeptical
T1	*n.s*	*n.s*	-	*n.s*	-	*n.s*	-	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T2	*n.s*	*n.s*	*n.s*	-	-	-	*n.s*	-	*n.s*	*n.s*	*	*n.s*	*n.s*	*n.s*
T3	*n.s*	-	**	*n.s*	*n.s*	*n.s*	***	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*
T4	-	***	-	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-
T5	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T6	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T7	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*
T8	***	*n.s*	-	*n.s*	*n.s*	*n.s*	-	-	*n.s*	*	-	*	*	*n.s*
T9	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T10	*n.s*	*n.s*	*n.s*	*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*	*n.s*	*n.s*

Таблиця 2.3.

Значущість різниць між виникненням самооцінних емоцій чоловіків та жінок

Task	Relaxed	Amused	Disgusted	Afraid	Angry	Frustrated	Sad	Sympathetic	Nervous	Pained	Embarrassed	Startled	Surprised	Skeptical
T1	*n.s*	*n.s*	-	-	-	-	-	-	*n.s*	-	*n.s*	-	*n.s*	-
T2	*n.s*	*	*n.s*	-	-	-	-	-	*n.s*	-	**	*n.s*	*n.s*	*n.s*
T3	-	-	*n.s*	*n.s*	*n.s*	*n.s*	***	*n.s*	*n.s*	*n.s*	-	*n.s*	*	-
T4	-	*	-	*n.s*	*n.s*	*n.s*	-	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-
T5	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T6	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T7	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	-	*n.s*	-	*n.s*	*	*n.s*	*n.s*
T8	***	*n.s*	-	*n.s*	*n.s*	*n.s*	-	-	*n.s*	*n.s*	-	*	*n.s*	*n.s*
T9	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T10	*n.s*	-	*n.s*	*	*n.s*	*n.s*	-	-	*n.s*	*n.s*	*n.s*	**	*n.s*	*n.s*

Аналіз інтенсивності (таблиця 2.2). Врахування інтенсивності змінює картину значущих відмінностей. Наприклад, у T2 (Здивування), хоча частота повідомлень про Забавленість не має значущої різниці між статями, інтенсивність цієї емоції, коли вона відчувається, суттєво відрізняється ($p<0.005$).

Для кількісної оцінки цього ефекту було розраховано середню інтенсивність емоцій у двох випадках:

- Діапазон [0,5] (з урахуванням 0, коли емоція не відчувалася): Середня інтенсивність для всіх завдань та емоцій становить 0.49 для жінок та 0.42 для чоловіків.

- Діапазон [1,5] (лише коли емоція відчувалася): Середня інтенсивність становить 3.67 для жінок та 3.43 для чоловіків.

2.3. Методологія та експериментальна оцінка

2.3.1. Попередня обробка даних

На початковому етапі було виконано детекцію та нормалізацію облич (вирівнювання щодо обертання, масштабування та центрування) з використанням бібліотеки Dlib. Вихідні зображення розміром 256×256 пікселів були субдискретизовані (downsampled) до роздільної здатності 175×175 пікселів.

Архітектура 3D CNN накладає обмеження на вхідні дані, вимагаючи фіксованої довжини часової послідовності. Хоча вибірка N послідовних кадрів є можливим технічним рішенням, такий підхід не забезпечує репрезентативності всієї послідовності. Для вирішення цієї проблеми було застосовано метод вибірки з повної послідовності. Враховуючи високу автокореляцію сусідніх кадрів, було відібрано 200 рівновіддалених кадрів з кожної послідовності, що дозволило максимізувати збереження темпоральної інформації.

2.3.2 Запропонована архітектура нейронної мережі

Базуючись на архітектурі FACS3D-Net, в даному дослідженні пропонується розширена багатоканальна (multi-branch) архітектура для мультиміткового розпізнавання емоцій, яка використовує спільні шари 3D згортки.

3D згортка забезпечує екстракцію просторово-часових ознак протягом усього емоційного епізоду, тоді як незалежні вихідні гілки виконують регресійний аналіз для кожної окремої емоції. Вихідні канали мережі використовують глибокі ознаки (deep features), отримані зі спільної 3D CNN, для прогнозування ймовірності наявності самооцінної емоції. Такий підхід дозволяє розглядати кожну емоцію незалежно, водночас враховуючи явище їхнього співвиникнення (co-occurrence).

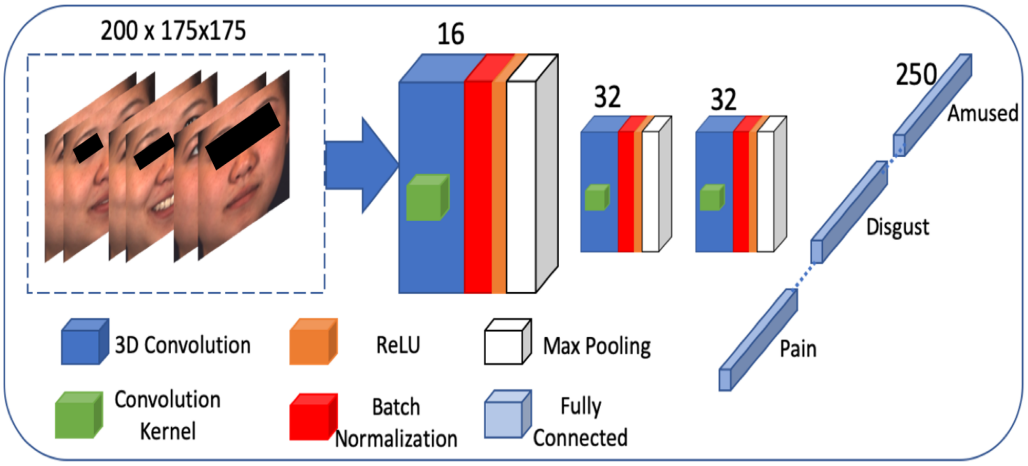


Рис. 2.3. Архітектура пропонованої 3D згорткової нейронної мережі для мультиміткового розпізнавання емоцій

Ця архітектура призначена для розпізнавання емоцій або станів людини (наприклад, "Amused", "Disgust", "Pain") на основі відеопослідовності зображень обличчя.

Наведемо опис кожного етапу обробки даних у цій мережі:

1. Вхідні дані (Input)

Формат - на вхід подається тензор розміром 200×175×175.

Кількість кадрів (часова розмірність, глибина відеопослідовності): 200

Прострова роздільна здатність зображення (висота та ширина в пікселях) - 175×175

Зміст - послідовність зображень обличчя людини.

2. Згорткові шари

Мережа складається з трьох послідовних блоків обробки. Кожен блок виконує вилучення просторово-часових ознак (spatial-temporal features).

Перший блок використовує 3D Convolution (тривимірну згортку) з 16 фільтрами (каналами). Включає ядро згортки (Convolution Kernel). Після згортки застосовується Batch Normalization (пакетна нормалізація) для стабілізації навчання. Далі йде функція активації ReLU (помаранчевий шар) для внесення нелінійності. Завершується шаром Max Pooling (білий блок) для зменшення розмірності даних.

Другий блок має аналогічну структуру, але кількість фільтрів збільшується до 32. Складається з 3D Conv → Batch Norm → ReLU → Max Pooling.

Третій блок також використовує 32 фільтри. Структура також повторюється.

3. Класифікація (Classification)

Після проходження через згорткові шари, отримані ознаки передаються на класифікатор. Дані "вирівнюються" (flattened) у вектор розміром 250. Цей шар об'єднує всі виділені ознаки для прийняття фінального рішення.

Вихід (Output):

Мережа видає передбачення класу емоції. На схемі показані приклади класів: amused (задоволений/розважений), disgust (огида), pain (біль).

Головна особливість цієї мережі — використання 3D-згорток, що дозволяє аналізувати не просто статичні картинки, а динаміку зміни виразу обличчя у часі (рух м'язів), що є критичним для точного розпізнавання емоцій.

2.3.3. Валідація даних

Враховуючи значну міжіндивідуальну варіативність емоційних реакцій навіть за ідентичних ситуативних умов, експериментальна стратегія базувалася на перехресній валідації за завданнями для конкретних суб'єктів (subject-specific task cross-validation). Зокрема, модель тренувалася на 9 завданнях та тестувалася на 1 завданні для одного суб'єкта. Ця процедура передбачала 10 ітерацій для кожного з 70 суб'єктів.

Обсяг вибірки для кожного суб'єкта становив 1800 кадрів для тренування та 200 кадрів для тестування. Результати представлено як усереднені показники ефективності за всіма суб'єктами. У процесі навчання множинні самооцінні емоції слугували еталонними мітками (ground truth) для кожної послідовності завдань.

У наборі даних BP4D+ спостерігається нерівномірний розподіл повідомлених емоцій за завданнями. Для задачі мультиміткового розпізнавання було відібрано шість категорій емоцій, частота виникнення яких перевищувала поріг у 100 випадків (таблиця 2.4).

Таблиця 2.4.

Метрики оцінки розпізнавання самооцінних емоцій

Емоція	**Усі**			**Жінки**			**Чоловіки**		
	Acc	**F1 Bin**	**AUC**	**Acc**	**F1 Bin**	**AUC**	**Acc**	**F1 Bin**	**AUC**
Розслаблений	0.89	0.45	0.70	0.89	0.48	0.72	0.89	0.42	0.68
Забавлений	0.80	0.42	0.65	0.81	0.45	0.68	0.78	0.38	0.62
Нервовий	0.81	0.42	0.68	0.79	0.44	0.68	0.84	0.38	0.66
Болючий	0.86	0.35	0.66	0.84	0.31	0.64	0.89	0.40	0.68
Збентежений	0.86	0.44	0.70	0.84	0.39	0.66	0.87	0.49	0.75
Здивований	0.75	0.39	0.62	0.75	0.39	0.61	0.74	0.39	0.63
Середнє	**0.83**	**0.41**	**0.67**	**0.82**	**0.41**	**0.67**	**0.84**	**0.41**	**0.67**

Таблиця 2.5.

Дисперсія та стандартне відхилення точності розпізнавання

Метрика	**Стандартне відхилення**	**Дисперсія**
Завдання	0.101	0.01
Суб'єкти	0.16	0.026

Таблиця 2.4 демонструє усереднені показники точності (Accuracy), F1-бали та площу під кривою (AUC) для чоловічої та жіночої підгруп, а також для загальної вибірки. Запропонована мережа досягла середньої точності 83%, значення F1-балу 0.41 та AUC 0.67.

Згідно з даними таблиці 2.5, низькі значення стандартного відхилення та дисперсії як за завданнями, так і за суб'єктами свідчать про стабільність (робастність) запропонованого методу незалежно від індивідуальних особливостей суб'єктів. Як було зазначено, патерни виникнення самооцінних

емоцій демонструють значну схожість між гендерними групами. Це спостереження корелює з результатами, які вказують на подібність оціночних метрик для досліджуваних емоцій між статями (таблиця 2.4).

2.3.4. Висновки щодо переваг пропонованої методології

У ході дослідження було проведено аналіз самооцінних емоцій. Результати продемонстрували, що, незважаючи на загальну тенденцію до схожості патернів реакцій на стимули у чоловіків та жінок, існують окремі випадки дивергенції між групами, які потенційно можуть мати стохастичну природу. Було ідентифіковано статистично значущі відмінності у виникненні та інтенсивності емоцій для різних гендерних груп у розрізі експериментальних завдань.

Емпіричні дані свідчать про те, що у вибірці BP4D+ суб'єкти жіночої статі в середньому демонструють вищі показники інтенсивності індукованих емоцій. Цей аналіз корелює з наявними літературними джерелами, підтверджуючи тезу про міжсуб'єктну варіативність емоційного реагування в межах ідентичних ситуативних контекстів.

Було розроблено та запропоновано архітектуру 3D згорткової нейронної мережі (3D CNN) для автоматизованого розпізнавання самооцінних емоцій на основі часових послідовностей виконання завдань. У роботі представлено комплексні метрики оцінки ефективності моделі.

Незважаючи на перспективність отриманих результатів, дане дослідження має низку обмежень, які окреслюють напрямки для майбутньої роботи:

- експериментальна база обмежувалася підмножиною суб'єктів з корпусу BP4D+. Майбутні дослідження вимагають масштабування аналізу на повну когорту суб'єктів, а також валідації на альтернативних наборах даних, що містять самооцінки емоційного стану.

- враховуючи високу міжсуб'єктну варіативність самооцінних емоцій, для верифікації узагальнюючої здатності (generalization) методу доцільно

застосувати протокол перехресної валідації з виключенням одного суб'єкта (leave-one-subject-out cross-validation).

Висновки до розділу

У другому розділі було здійснено поглиблене дослідження самооцінних емоцій, що відзначаються великою варіативністю та складністю інтерпретації через їхній внутрішній, рефлексивний характер. Аналіз теоретичних особливостей цього класу емоцій показав, що вони формуються на перетині соціальних, когнітивних та емоційних процесів, що зумовлює нетривіальність їх зовнішнього вираження. Використання корпусу BP4D+ дозволило емпірично виявити закономірності у співвідношенні між внутрішніми самооцінками суб’єктів та їхніми мікроекспресіями. Дослідження показало, що навіть невеликі зміни у динаміці AU можуть сигналізувати про наявність складних емоційних станів, пов’язаних із почуттями провини, сорому чи гордості. Запропонована методологія побудови архітектури нейронної мережі, яка інтегрує кілька форм представлення даних, продемонструвала здатність моделювати як статичні, так і динамічні компоненти експресій. Валідація моделей підтвердила, що включення часових характеристик значно підвищує точність прогнозування, особливо у випадках слабко виражених або прихованих емоцій. Результати свідчать про те, що одночасне врахування AU-патернів та самооцінок формує більш стійку модель емоційної поведінки, ніж використання лише зовнішніх ознак.

РОЗДІЛ 3. ІМПЛЕМЕНТАЦІЯ МЕТОДІВ ТА МОДЕЛЮВАННЯ ЕМОЦІЙ В КОНТЕКСТІ РЕЛЕВАНТНИХ ТА ПОХІДНИХ ДАНИХ

3.1. Методологія мультимодальної оцінки емоцій на основі фізіологічних сигналів та лицьових одиниць дії

Клінічна оцінка больового синдрому є складним діагностичним завданням. На сьогодні самозвіт пацієнта (self-report) залишається найбільш поширеним методом оцінки. Проте, при застосуванні в клінічних умовах, такі міри є притаманно суб'єктивними і часто позбавлені критично важливої часової інформації (temporal information).

Автоматизовані методи розпізнавання болю мають потенціал для покращення якості життя пацієнтів, надаючи об'єктивний інструментарій для оцінки. З огляду на це, останніми роками спостерігається зростання кількості перспективних досліджень у цьому напрямку, які охоплюють аналіз лицьових експресій, фізіологічних сигналів, кінематики та м'язової активності.

3.1.1. Ключові описи та набори даних

Для сприяння прогресу в автоматичному розпізнаванні болю використано набір даних EmoPain, який включає:

- Багатовидові відеозаписи обличчя;
- Аудіодані;
- 3D захоплення руху (motion capture);
-Електроміографічні (EMG) сигнали м'язів спини.

Вибірка набору даних складається з 22 пацієнтів (7 чоловіків / 15 жінок) із хронічним болем у спині та контрольної групи з 28 здорових суб'єктів (14 чоловіків / 14 жінок).

На основі цього набору даних було проведено низку важливих досліджень:

В [21] запропонували архітектуру глибокого навчання BodyAttentionNet для розпізнавання захисної поведінки (protective behavior). Ця архітектура навчається на часовій інформації та визначає релевантні частини тіла за допомогою механізму уваги. Автори продемонстрували підвищення точності розпізнавання при зменшенні кількості параметрів моделі у 6–20 разів порівняно з існуючими аналогами (state-of-the-art).

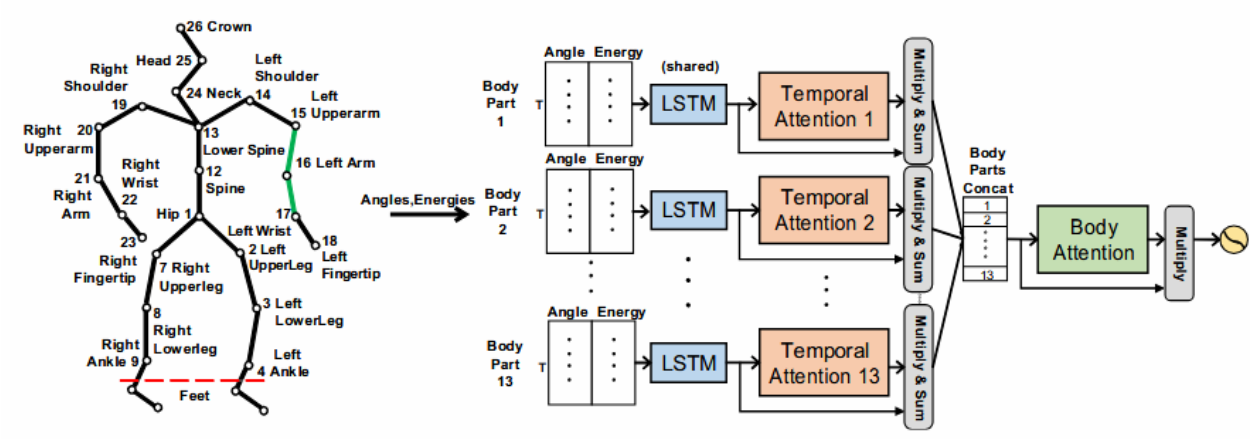


Рис. 3.1. Архітектура процесу глибокого навчання BodyAttentionNet для розпізнавання захисної поведінки

В дослідженні [22] розробили метод автоматичної класифікації рівнів болю на основі кінематики та м'язової активації. Застосування алгоритмів Random Forest та SVM на даних EmoPain дозволило досягти високої точності у розрізненні низького та високого рівнів болю, а також станів здорових суб'єктів. У подальшому дослідженні вивчали вплив афективних факторів на щоденне функціонування при хронічному болю, виявивши, що дані про рух є інформативними для розпізнавання рівнів стресу та болю.

В роботі [23] запропонували метод злиття фізіологічних сигналів (sensor fusion) для розпізнавання емоцій. Їхній підхід, заснований на зваженому злитті, продемонстрував високу ефективність у детекції болю на наборі даних BP4D+ [139], досягнувши точності 98.48%.

У контексті лицьового аналізу, застосували підхід на основі Active Appearance Model для детекції болю у відео з набору даних UNBC-McMaster

shoulder pain, використовуючи одиниці дії (FACS). Отримані результати підтвердили валідність використання AU як модальності для ідентифікації болю.

3.1.2. Пропонована методологія мультимодального злиття

Базуючись на проаналізованих роботах, у цьому дослідженні пропонується метод розпізнавання болю шляхом мультимодального злиття фізіологічних даних та лицьових одиниць дії.

Використовувані модальності:

1. Фізіологічні сигнали - частота серцевих скорочень (HR), частота дихання (respiration), кров'яний тиск (blood pressure) та електродермальна активність (EDA).

2. Одиниці дії (Action Units) - використовуються AU з найбільш експресивних зон обличчя.

Експериментальна валідація проводилася на наборі даних BP4D+. Ключовим етапом попередньої обробки була часова синхронізація фізіологічних сигналів із зображеннями обличчя, що містять активовані AU (one-to-one correspondence).

Відповідно основними результатами є:

- Розробка методу розпізнавання болю на основі злиття (fusion) фізіологічних сигналів та одиниць дії.

- Аналіз кореляцій між фізіологічними параметрами та активацією лицьових м'язів.

Проведення порівняльних експериментів у гендерному розрізі (чоловіки проти жінок) для виявлення специфіки проявів болю та кореляцій між модальностями.

Ефективність запропонованого методу злиття була перевірена експериментально. Результати свідчать про позитивний синергетичний ефект об'єднання модальностей на точність детекції.

Таблиця 3.1.

Порівняльний аналіз ефективності розпізнавання болю за модальностями (фізіологічні сигнали, одиниці дії, злиття / M = Чоловіки; F = Жінки)

Модальність	**Група 1 (Acc / F1)**	**Група 2 (Acc / F1)**	**Група 3 (Acc / F1)**	**Група 4 (Acc / F1)**	**Група 5 (Acc / F1)**
Фізіологічні	77.70 % / 0.300	75.25 % / 0.285	76.98 % / 0.269	69.14 % / 0.350	75.43 % / 0.219
Одиниці дії	89.02 % / 0.734	88.00 % / 0.668	90.73 % / 0.778	88.27 % / 0.725	87.50 % / 0.753
Злиті (fused)	89.20 % / 0.750	88.58 % / 0.689	—	—	—

3.2. Методологія мультимодального розпізнавання болю

Для вирішення задачі розпізнавання болю ми пропонуємо підхід мультимодального злиття (multimodal fusion), що інтегрує фізіологічні сигнали та одиниці дії обличчя (Action Units, AU). Аналіз фокусується на найбільш експресивних сегментах послідовностей завдань, призначених для індукції емоційних станів. Експресивний сегмент визначається як часовий інтервал, що містить кадри, для яких було виконано ручну анотацію одиниць дії (AU) згідно з системою FACS. Експериментальна валідація методу здійснюється на базі мультимодального корпусу емоцій BP4D+ [139].

3.2.1 Характеристика набору даних

BP4D+ являє собою комплексний мультимодальний набір даних, що включає:

- Візуальні дані: 2D (RGB) та теплові зображення, 3D моделі обличчя.
- Геометричні дані: 2D та 3D орієнтири (landmarks) обличчя.
- Поведінкові анотації: Вручну закодовані одиниці дії (загалом 33 AU).

Фізіологічні дані (8 каналів) це діастолічний артеріальний тиск, середній артеріальний тиск, електродермальна активність (EDA), систолічний артеріальний тиск, "сирий" (raw) сигнал артеріального тиску,

частота пульсу, частота дихання та амплітуда/напруга дихання (respiration valance).

Вибірка складається зі 140 суб'єктів (58 чоловіків та 82 жінки) віком від 18 до 66 років. Протокол збору даних передбачає виконання суб'єктами 10 завдань для індукції різних емоційних станів. Оскільки дане дослідження фокусується на детекції болю, основна увага приділяється завданню холодового пресора (Cold Pressor Task), під час якого суб'єкти занурюють руку у відро з крижаною водою на тривалий час.

У нашому експериментальному дизайні використовуються дані 139 суб'єктів. Суб'єкт F082 був виключений з вибірки через неповноту даних (відсутні ознаки). Анотація AU доступна для приблизно 20 секунд найбільш експресивних кадрів для чотирьох цільових емоцій: щастя, збентеження, страх та біль. Відповідно, ці чотири емоційні категорії формують основу для наших експериментів.

3.2.2. Часова синхронізація модальностей

Існує значна розбіжність у частоті дискретизації даних: відеозаписи у BP4D+ мають частоту 25 кадрів на секунду (fps), тоді як фізіологічні сигнали реєструються з частотою приблизно 1000 Гц. Для коректного злиття модальностей необхідна процедура синхронізації.

Процес синхронізації включає наступні етапи:

1. Розрахунок кількості відеокадрів у цільовій послідовності.

2. Субдискретизація (downsampling) фізіологічних сигналів до кількості відеокадрів із використанням техніки однокрокового бутстрепінгу (one-step bootstrapping). Цей метод дозволяє зменшити кількість зразків, зберігаючи при цьому інформативність сигналу.

3. Встановлення відповідності "один-до-одного" між кожним відеокадром та вектором фізіологічних показників.

Враховуючи варіативність тривалості відеопослідовностей між суб'єктами, було виконано фінальну нормалізацію (resampling) найбільш експресивних сегментів до фіксованої довжини у 5000 кадрів. Отримані 5000 синхронізованих кадрів (пари AU та фізіологічних сигналів) використовуються для подальшого злиття.

3.2.3. Злиття ознак

На основі 5000 синхронізованих кадрів формується об'єднаний простір ознак. Для кожного кадру t виконується конкатенація вектора з 33 AU та вектора з 8 фізіологічних сигналів, утворюючи локальний вектор ознак f_t:

$$f_t = [AU_1, \ldots, AU_{33}, Phys_1, \ldots, Phys_8]$$

де AU_i представляють інтенсивність або наявність 33 одиниць дії, а $_{Physj}$ — значення 8 фізіологічних сигналів. Розмірність вектора f_t становить 41.

Для врахування часової динаміки (temporal information) у процесі виявлення болю, ми об'єднуємо вектори f_t для всіх 5000 кадрів послідовності у єдиний глобальний вектор ознак F:

$$\mathbf{F} = [f_1, f_2, \ldots, f_{5000}]$$

Така агрегація призводить до формування вектора ознак високої розмірності (41×5000=205,000), який подається на вхід класифікатора. Загалом, для 139 суб'єктів та 4 завдань було сформовано 556 векторів ознак.

3.2.4. Особливості побудови методології

У рамках задачі бінарної класифікації "Біль проти Відсутності болю":

- Позитивний клас (Positive Class) - послідовності, що відповідають завданню індукції болю.

- Негативний клас (Negative Class) - послідовності, що відповідають завданням індукції щастя, збентеження та страху.

Цей дизайн узгоджується з попередніми дослідженнями на наборі даних BP4D+. Для класифікації було використано алгоритм випадкового лісу (Random Forest) з ансамблем із 275 дерев.

Валідація методу проводилася з використанням протоколу незалежної від суб'єкта 10-кратної перехресної перевірки (subject-independent 10-fold cross-validation) на всій вибірці (139 суб'єктів).

Окрім загальної оцінки, було проведено:

- Міжгендерні експерименти (Cross-gender experiments).

- Гендерно-специфічні експерименти (Gender-specific experiments).

У кожному з експериментальних сценаріїв оцінювалася ефективність як окремих модальностей (лише фізіологічні сигнали або лише AU), так і запропонованого підходу мультимодального злиття.

3.3. Аналіз ефективності злиття фізіологічних сигналів та одиниць дії у задачах детекції болю

3.3.1. Ефективність розпізнавання болю

Для оцінки ефективності системи розпізнавання болю використовувалися метрики загальної точності (Accuracy) та F1-міри (F1-score).

У рамках суб'єктно-незалежних експериментів (subject-independent experiments) на повній вибірці, застосування методу злиття (fusion) фізіологічних сигналів та одиниць дії (AU) дозволило досягти точності 89.2% та F1-міри 0.75.

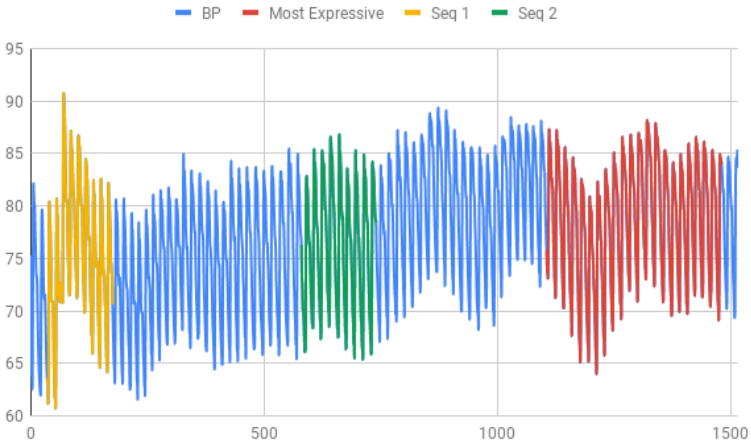
Аналіз окремих модальностей показав наступне:

- Фізіологічні сигнали продемонстрували точність 77.7%, проте F1-міра виявилася низькою — 0.3. Це свідчить про те, що класифікатор, навчений

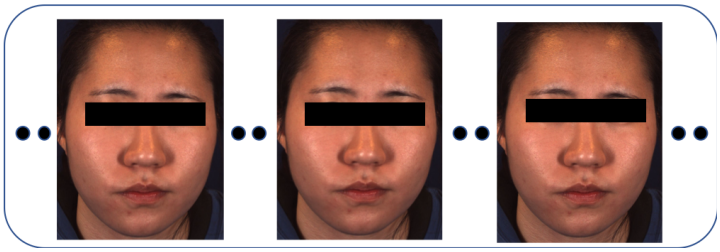
виключно на цій модальності, схильний відносити більшість зразків до класу "відсутність болю" (який є мажоритарним у вибірці).

- Одиниці дії (AU), як мономодальність досягли точності 89.02% та F1-міри 0.734. Ці результати узгоджуються з літературними даними, підтверджуючи, що AU є дискримінативним дескриптором для автоматичного розпізнавання болю.

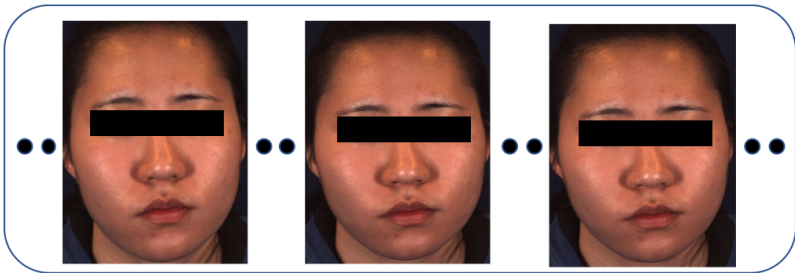
а) дихання



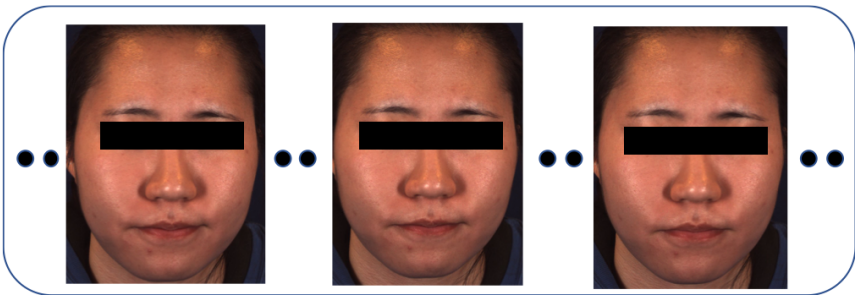
б) тиск



в) Послідовність 1: суб'єкт занурює руку в крижану воду



г) Послідовність 2: приблизно 20 с після занурення



д) Послідовність 3: відбувається через 40 с після занурення

Рис. 3.2. Візуальне зіставлення динаміки фізіологічних сигналів (дихання та артеріальний тиск) і лицьових експресій (одиниць дії)

Гендерно-специфічні експерименти (навчання та тестування в межах однієї статі) показали схожі тенденції:

- Для фізіологічних сигналів точність становила близько 76%, а F1-міра — 0.28, знову ж таки демонструючи зміщення в бік класу "відсутність болю".

- Одиниці дії продемонстрували високу ефективність: точність ~89%, F1-міра 0.668 (для чоловіків) та 0.778 (для жінок).

У перехресних (cross-gender) експериментах (наприклад, навчання на жінках, тестування на чоловіках) результати залишалися порівнянними, при цьому фізіологічні сигнали знову показали низьку F1-міру (~0.28). Важливим спостереженням стало те, що мультимодальне злиття не покращило результати в цьому сценарії.

При навчанні на чоловіках і тестуванні на жінках точність злиття (88.27%, F1 0.725) не перевищила результатів для AU.

При навчанні на жінках і тестуванні на чоловіках спостерігалося незначне зниження точності (на 1.19%) порівняно з використанням лише AU.

Це явище можна частково пояснити відмінностями в кореляційних патернах між статями при аналізі найбільш експресивних сегментів послідовностей (рис. 3.3 б та 3.3 в). Детальні результати наведено в таблиці 3.1.

Таблиця 3.1.

Результати розпізнавання болю з використанням фізіологічних сигналів, одиниць дії та злиття обох модальностей

Модальності	Усі суб'єкти		Чоловіки		Жінки		Навчання (Ч) / Тест (Ж)		Навчання (Ж) / Тест (Ч)	
Метрика	Точність	F1	Точність	F1	Точність	F1	Точність	F1	Точність	F1
Фізіологічні	77.70%	0.3	75.25%	0.285	76.98%	0.269	69.14%	0.35	75.43%	0.219
Одиниці дії	89.02%	0.734	88.00%	0.668	90.73%	0.778	88.27%	0.725	87.50%	0.753
Злиті	89.20%	0.75	88.58%	0.689	91.35%	0.787	88.27%	0.725	86.31%	0.75

3.3.2. Порівняння з сучасними методами

Наскільки відомо, прямі порівняльні дослідження з використанням ідентичних модальностей на наборі даних BP4D+ у літературі відсутні. Найбільш релевантною є робота [33], в якій досліджували злиття фізіологічних сигналів на BP4D+, проте їхній підхід не був суб'єктно-незалежним. Оскільки це найближчий аналог, ми наводимо детальне порівняння.

Існуючий метод. Використовуючи нейронну мережу прямого поширення (Feed-forward NN), досягли точності 98.48%. Однак при використанні класичних методів результати були нижчими: SVM (92.64%), Random Forest (90.27%), Naive Bayes (89.77%).

Пропонований метод. Використовуючи Random Forest лише на фізіологічних сигналах у суб'єктно-незалежному режимі, ми отримали точність 77.7%. Проте, запропонований метод мультимодального злиття

(Phys + AU) досягає точності 89.2%, що є співставним із існуючим результатом для алгоритму Random Forest, незважаючи на більш суворі умови валідації (суб'єктна незалежність).

3.3.3 Аналіз отриманих рехультатів

На основі аналізу взаємозв'язку фізіологічних сигналів та одиниць дії в контексті болю, можна виділити два ключові спостереження:

1. Часова асинхронність піків.

Пікові значення фізіологічних сигналів та інтенсивності лицьових експресій відбуваються в різні моменти часу (див. рис. 3.2).

2. Динаміка кореляції.

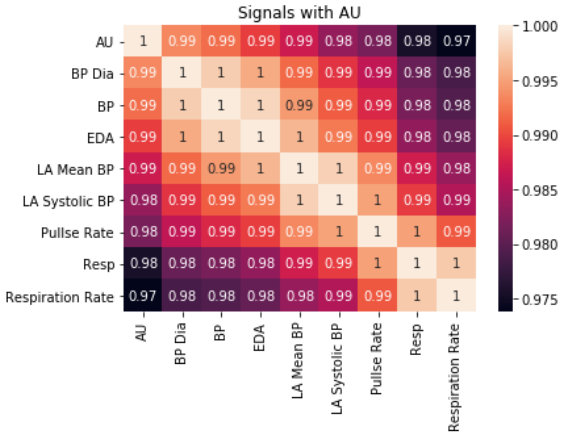
Фізіологічні сигнали демонструють сильну кореляцію в найбільш експресивних сегментах послідовності, проте кореляція відсутня при аналізі повного сигналу (див. рис. 3.3).

Як показано на рисунку 3.2, у момент найбільшої варіабельності фізіологічних сигналів (послідовність 1, занурення руки), м'язова реакція обличчя відсутня (рис. 3.2 в). Натомість, коли виникає лицьова експресія (рис. 3.2 д), фізіологічні сигнали стабілізуються (вирівнюються), і їхня варіабельність стає мінімальною.

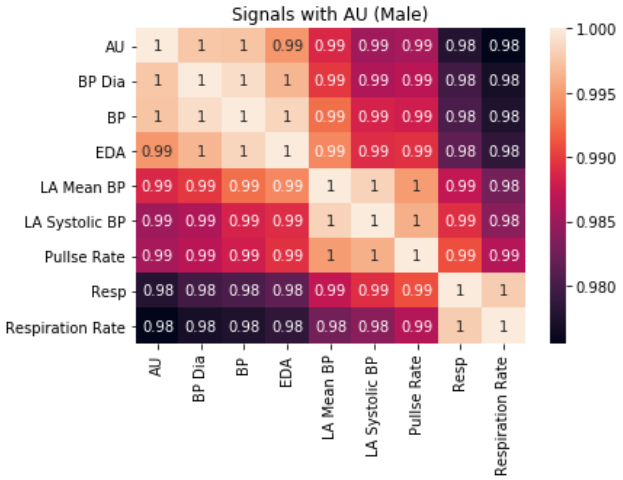
Для глибшого аналізу цього ефекту було розраховано коефіцієнт кореляції між сигналами як для найбільш експресивних сегментів, так і для повного запису.

На рис. 3.3 г, протягом усього запису фізіологічні сигнали корелюють слабо. На рис. 3.3 а - при вибірковому аналізі найбільш експресивних сегментів спостерігається суттєве зростання кореляції між фізіологічними сигналами. Цей ефект є стійким для обох статей.

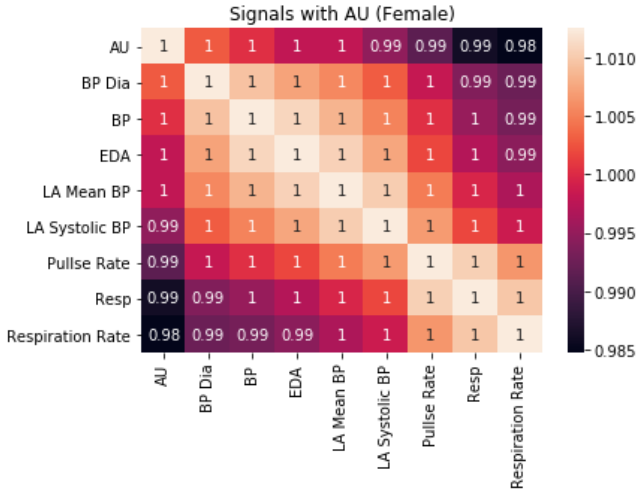
Зміна патернів варіабельності та кореляції під час експресивної фази може бути одним із пояснень низької точності розпізнавання болю виключно за фізіологічними даними, оскільки попередні дослідження вказують на те, що вища варіабельність сигналів зазвичай сприяє кращій детекції болю.



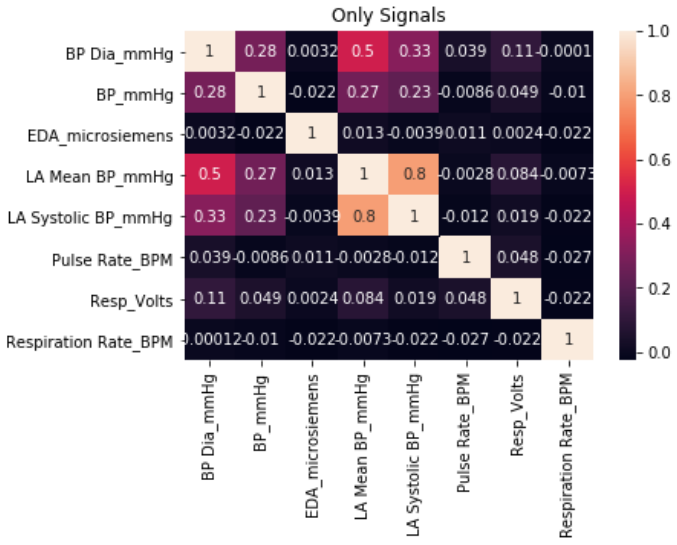
а) Фізіологічні дані з середньою частотою виникнення AU (найекспресивніша послідовність)



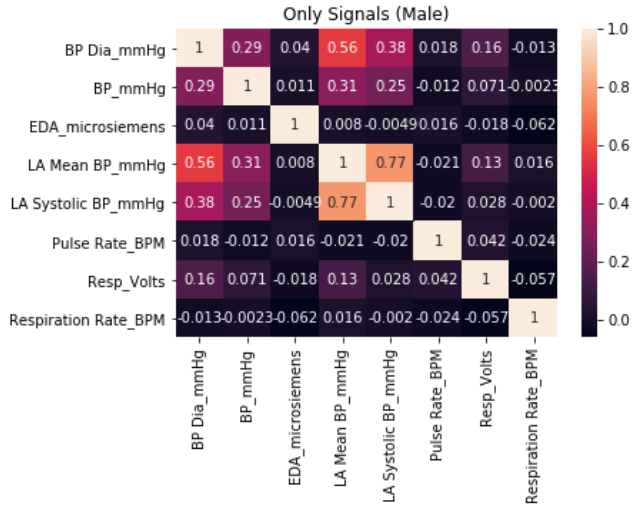
б) Фізіологічні дані з середньою частотою виникнення AU (найекспресивніша послідовність) - чоловіки



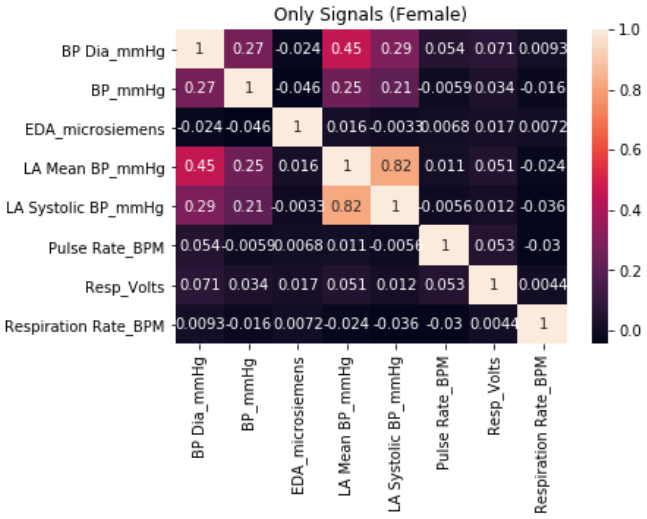
в) Фізіологічні дані з середньою частотою виникнення AU (найекспресивніша послідовність) — жінки



г) Повний фізіологічний сигнал



д) Повний фізіологічний сигнал — чоловіки



е) Повний фізіологічний сигнал — жінки

Рис. 3.3. Матриці кореляцій фізіологічних сигналів та одиниць дії

Рисунки 3.3 а – в - кореляції в найбільш експресивних сегментах послідовності, а г – е - кореляції для повного фізіологічного сигналу (без урахування AU).

3.4. Метод розпізнавання контексту з використанням динаміки лицьових експресій на основі патернів одиниць дії

Афективні обчислення (Affective Computing) є міждисциплінарною науковою галуззю, що об'єднує комп'ютерні науки, психологію та когнітивістику для дослідження та аналізу людських емоцій, які є предметом ґрунтовного вивчення в психології.

В роботі [36] автор досліджував диференціацію проявів сорому (embarrassment) у порівнянні з пов'язаною емоцією веселощів (amusement), які часто плутають. Він встановив, що як морфологія, так і динамічні патерни цих двох емоцій є відмінними та мають різний кореляційний зв'язок із самозвітами суб'єктів. В [37] аналізували часові характеристики та морфологію усмішок, які сприймаються як різні емоції (наприклад, веселощі та нервозність). Їхні результати свідчать про те, що усмішки, які сприймаються як схожі, насправді відрізняються за амплітудою, тривалістю, фазами початку (onset) та завершення (offset), а також залучають різні одиниці дії (AUs). Наприклад, усмішки, інтерпретовані як прояв веселощів, були тривалішими, мали більшу амплітуду та частіше включали активацію AU6. Ця робота слугує мотивацією для поточного дослідження, оскільки вона підкреслює важливість інтенсивності та часової динаміки для аналізу емоцій. Ми висуваємо гіпотезу, що використання цієї інформації призведе до покращення розпізнавання контексту на основі динаміки лицьових експресій.

Робота [30] також підкріплює нашу гіпотезу. Автори обговорюють, як емоції викликаються за певних обставин (тобто в контексті) та сприймаються через специфічні лицьові експресії. Вони аргументують, що для повного розуміння інформаційного наповнення лицьових експресій необхідно

відображати безперервні сигнали обличчя, тіла та голосу (тобто часову інформацію) на багатогранні переживання, використовуючи статистичні методи та методи машинного навчання.

Дослідження [27] ідентифікувало варіативність емоційних проявів у різних культурах. Аналізуючи 22 емоції у 5 культурах, вони виявили, що культура та контекст суттєво впливають на експресію, виходячи за межі шести базових емоцій, які широко вивчаються в літературі. Вони зосередилися на шести найбільш популярних емоціях: гнів, страх, огида, сум, щастя та здивування. Їхні висновки вказують на те, що хоча експресії іноді відповідають конкретним емоціям (наприклад, нахмурення при сумі), культура та контекст є визначальними факторами.

Крім того, такі лицьові рухи, як гримаса, не завжди відповідають певному емоційному стану. Однією з рекомендацій їхньої роботи є необхідність дослідження контекстно-чутливих способів, якими люди використовують лицьові м'язи для вираження емоцій. Наша робота усуває це обмеження шляхом дослідження розпізнавання контексту на основі часової динаміки лицьових експресій.

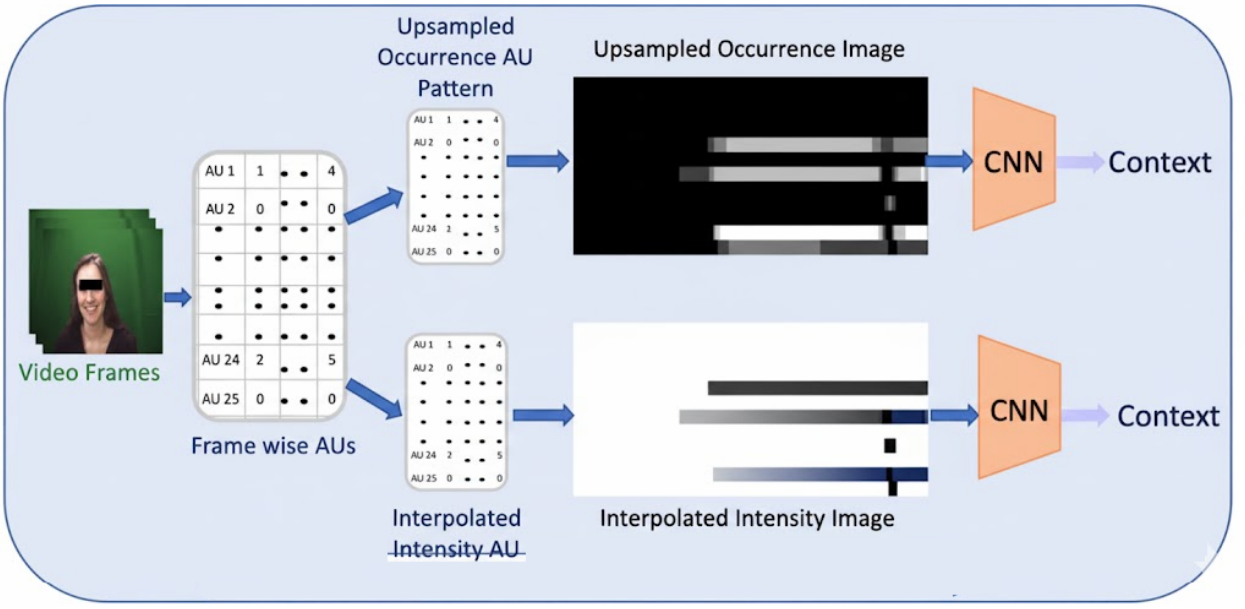


Рис. 3.4. Огляд підходу до моделювання динаміки обличчя на основі патернів AU

З вхідного відео вилучаються зображення, після чого покадрові одиниці дії (AUs) піддаються передискретизації (up-sampled) для визначення факту виникнення (occurrence) та інтерполяції для визначення інтенсивності. Патерни AU потім трансформуються у простір зображень (image-space), які використовуються для тренування окремих згорткових нейронних мереж (CNN) для розпізнавання контексту. Іноді окремі мережі використовуються для порівняння ефективності патернів виникнення AU проти патернів інтенсивності AU.

На рисунку 3.4 зображено схему підходу до моделювання динаміки обличчя на основі патернів одиниць дії (AU) для розпізнавання контексту. Цей підхід трансформує часові ряди активності м'язів обличчя у візуальні "зображення", які потім аналізуються за допомогою згорткових нейронних мереж (CNN).

Процес складається з наступних етапів.

1. Вхідні дані та екстракція AU (Input & Extraction):

- На вхід системи подаються відеокадри (Video Frames) із зображенням обличчя суб'єкта.

- Для кожного кадру виконується детекція одиниць дії, що формує матрицю Frame-wise AUs. У цій матриці рядки відповідають окремим AU (наприклад, AU1, AU2... AU25), а стовпці — послідовності кадрів. Значення в комірках відображають інтенсивність активації (від 0 до 5).

2. Формування патернів (Pattern Generation) - система розділяється на дві паралельні гілки для обробки різних аспектів даних:

- Верхня гілка (Occurrence) обробляє дані про факт активації м'язів. Дані піддаються передискретизації (Upsampled Occurrence AU Pattern) для створення Upsampled Occurrence Image. Це чорно-біле зображення, де візуалізується бінарна активність або дискретна присутність AU у часі (світлі смуги на чорному тлі).

- Нижня гілка (Intensity) обробляє безперервні значення сили активації м'язів. Дані інтерполюються (Interpolated Intensity AU) для створення

Interpolated Intensity Image. Це зображення відображає плавну динаміку зміни інтенсивності емоцій.

3. Класифікація (Classification via CNN):

- Отримані "зображення" (які фактично є візуалізацією часових рядів активності AU) подаються на вхід окремим згортковим нейронним мережам (CNN).

- CNN навчаються розпізнавати специфічні текстури та патерни на цих картах, що відповідають певним емоційним станам у часі.

4. Вихідні дані. Кінцевим результатом роботи мереж є визначення контексту (Context), тобто ситуації або стимулу, що викликав дану емоційну реакцію.

Суть методу полягає в тому, що пропонується розглядати часову динаміку емоцій не як простий вектор чисел, а як двовимірне зображення, що дозволяє використовувати потужний апарат комп'ютерного зору (CNN) для аналізу часових патернів поведінки людини.

На основі психологічних досліджень, сфера афективних обчислень зазнала значного розвитку протягом останніх двох десятиліть. Було показано, що емоції є важливою складовою людського інтелекту [24]. Вони є фундаментальною особливістю когніції, а емоційні результати є центральними для мислення, дій та сприйняття. З огляду на це, дана галузь має широкий глобальний вплив із застосуванням у таких сферах, як медицина, освіта, безпека, розваги та покращення клієнтського досвіду.

В роботі [27] був розроблений D-PattNet для вирішення трьох основних проблем у детекції AU на основі патчів:

1) обертання голови;

2) спільне виникнення (co-occurrence) AU;

3) просторово-часова динаміка.

Для подолання цих викликів D-PattNet кодує локальні патчі на ранніх етапах мережі, потім об'єднує інформацію на основі патчів за допомогою

механізму уваги, і, нарешті, використовує 3D-CNN для моделювання просторово-часової динаміки.

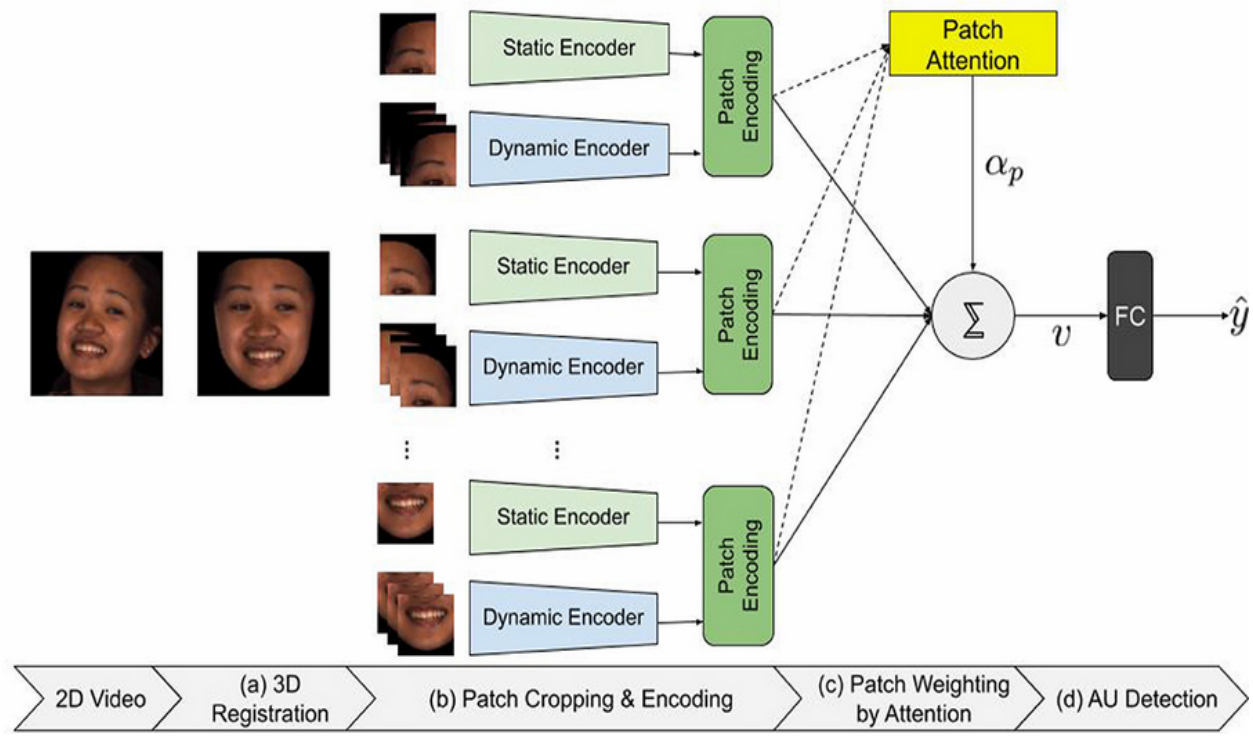


Рис. 3.5. Підхід D-PAttNet

a) Оцінюється щільний набір орієнтирів обличчя (facial landmarks) та реконструюється щільна 3D-сітка (dense 3D mesh) обличчя.

b) Патчі, що містять регіони обличчя, пов'язані з конкретними AU, кадруються та подаються на вхід різним CNN для кодування. Для кожного патча 2D-CNN використовується для кодування статичної інформації на рівні кадру (frame-level), а 3D-CNN — для кодування динамічної інформації на рівні сегмента (segment-level). Кодування патча отримується шляхом конкатенації статичного та динамічного представлень.

c) Патчі зважуються за допомогою сигмоїдального механізму уваги для детекції конкретних AU.

d) Кодування обличчя подаються на повнозв'язний шар (FC) для детекції AU.

Рисунок 3.5 демонструє структурні компоненти архітектури динамічної мережі з механізмом уваги до патчів (Dynamic Patch-Attentive Network, D-PAttNet).

1. На першому етапі виконується щільна 3D-реєстрація (dense 3D registration) на основі 2D-відеопослідовностей (рис. 3.5a).

2. Далі здійснюється кадрування патчів (patches), що містять локальні ділянки обличчя.

3. Для кожної локалізації патча застосовуються окремі нейронні мережі: 2D-CNN для кодування локальної статичної інформації та 3D-CNN для кодування локальної динамічної інформації.

4. Статичні та динамічні кодування конкатенуються (об'єднуються) для отримання підсумкового кодування патча (рис. 3.5b).

5. Застосовується сигмоїдальний механізм уваги (sigmoidal attention mechanism) для зважування внеску кожного патча у процес детекції специфічних одиниць дії (AUs) (рис. 3.5c).

На завершальному етапі, використовуючи фінальне представлення (кодування) обличчя, здійснюється детекція 12 AU (рис. 3.5 d).

Для забезпечення функціонування цих застосувань використовується широкий спектр модальностей, включаючи, але не обмежуючись: 2D та 3D зображення обличчя, тепловізійні зображення, фізіологічні сигнали, жести та аудіо. При використанні 2D зображень часто детектуються одиниці дії (AUs) за системою FACS, які є стандартом для кількісної оцінки рухів обличчя. У Відомо, що патерни AU впливають на детекцію, та запропонували підхід до тренування нейронних мереж на патернах AU для покращення їхнього розпізнавання. Базуючись на цьому, ми висуваємо гіпотезу, що патерни AU також можуть бути використані для розпізнавання контексту, який викликав емоційну реакцію.

Отже, запропоновано підхід до моделювання часової динаміки лицьової експресії на основі патернів AU. Ми демонструємо, що запропонований метод здатний точно розпізнавати контекст, використаний

для індукції емоційної реакції, на трьох передових наборах даних (state-of-the-art datasets).

Проведено експерименти з використанням як патернів виникнення (occurrence), так і патернів інтенсивності (intensity) AU для розпізнавання контексту. Ці експерименти свідчать, що патерни інтенсивності краще підходять для цього завдання. Крім того, результати вказують на необхідність використання більшої кількості AU для підвищення точності розпізнавання контексту.

3.5. Методологія експериментального дослідження методу розпізнавання контексту з використанням нейронної мережі

Для вирішення задачі розпізнавання контексту на основі аналізу часових патернів одиниць дії (AU) було використано набір даних BP4D+. Основний фокус дослідження зосереджено на вивченні темпоральної динаміки AU, що активуються під час виконання завдань емоційної індукції. Експериментальний дизайн дослідження включає наступні етапи:

1. Попередня обробка даних

1.1. Нормалізація даних. З метою уніфікації вхідних векторів було проведено нормалізацію даних, що забезпечує фіксовану довжину послідовностей AU. Цей процес включає часове вирівнювання (temporal alignment) послідовностей до встановленої кількості кадрів.

1.2. Селекція експресивних сегментів. Вибірка даних обмежувалася найбільш експресивними сегментами відеопослідовностей, для яких наявна експертна ручна анотація одиниць дії. Такий підхід дозволяє зосередити аналіз на найбільш релевантних та інформативних даних для задачі детекції контексту.

У рамках експериментального дизайну було застосовано два типи архітектур нейронних мереж:

1) згорткова нейронна мережа (Convolutional Neural Network, CNN)

2) повнозв'язна нейронна мережа прямого поширення (Feed-Forward Fully-Connected Neural Network).

Для вирішення задачі розпізнавання контексту було розроблено та навчено 7-шарову згорткову нейронну мережу, на вхід якої подаються зображення патернів AU (рис. 3.6). Конфігурація шарів мережі є наступною.

Перший та другий шари (згорткові). Перший шар містить 64 фільтри, другий — 128 фільтрів. Обидва шари мають розмір ядра згортки (kernel size) (5,5), крок (stride) 2 та використовують функцію активації ReLU.

Третій шар – це шар пакетної нормалізації (Batch Normalization).

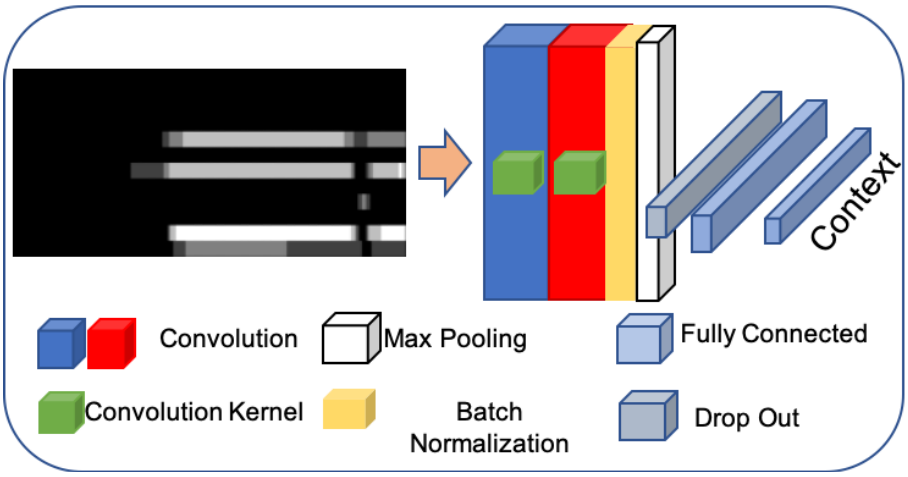


Рис. 3.6. Архітектура згорткової нейронної мережі

Зображення патернів одиниць дії (як інтенсивності, так і факту виникнення) використовуються для навчання мережі розпізнаванню контексту.

Четвертий шар - шар макс-пулінгу (Max Pooling) з розміром ядра (2,2).

П'ятий шар - шар регуляризації (Dropout) з коефіцієнтом виключення 0.3.

Шостий шар - повнозв'язний (Dense) шар, що складається з 1000 нейронів з функцією активації ReLU.

Сьомий шар (вихідний) - кількість нейронів відповідає кількості цільових класів контексту; застосовано функцію активації SoftMax.

Гіперпараметри навчання наступні:

- Функція втрат: категоріальна перехресна ентропія (Categorical Crossentropy).

- Оптимізатор: Adadelta.

- Швидкість навчання (Learning rate): 0.001.

- Розмір пакету (Batch size): 5.

- Тривалість навчання: 20 епох.

2. Екстракція ознак (Feature Extraction)

Архітектура системи базується на гібридному підході, що поєднує згорткові нейронні мережі (CNN) для вилучення просторових ознак із зображень обличчя та рекурентні нейронні мережі (RNN) для моделювання часових залежностей у послідовностях AU:

- CNN (просторові ознаки) - застосовується для екстракції високорівневих просторових репрезентацій із зображень. Архітектура включає каскад шарів згортки (convolutional layers) та субдискретизації (pooling layers).

- RNN (часові ознаки) - використовується для виявлення темпоральних залежностей та моделювання динаміки послідовностей AU. Зокрема, застосовано архітектуру довгої короткочасної пам'яті (Long Short-Term Memory, LSTM), яка ефективно вирішує проблему зникаючого градієнта при роботі з часовими рядами.

3. Класифікація

Етап класифікації контексту реалізовано за допомогою повнозв'язних шарів (fully connected layers), які інтегрують виходи підмереж CNN та RNN. Навчання моделі проводилося на тренувальній вибірці з подальшою оцінкою ефективності на тестовому наборі даних.

4. Валідація моделі

Для оцінки узагальнюючої здатності (generalization capability) моделі та запобігання перенавчанню (overfitting) було застосовано метод k-кратної перехресної перевірки (k-fold cross-validation).

Потім представлено емпіричні результати експериментів із розпізнавання контексту на основі часових патернів AU. Ефективність запропонованої моделі оцінювалася за допомогою стандартних метрик:

- точності (Accuracy),
- F1-міри (F1-score),
- площі під ROC-кривою (AUC).

Було проведено порівняльний аналіз ефективності різних архітектурних підходів: використання виключно просторових ознак (CNN), виключно часових ознак (RNN) та їхньої комбінації. Результати, наведені у таблиці 3.2, демонструють перевагу гібридного підходу.

Таблиця 3.2.

Ефективність розпізнавання контексту різними методами

Метод	**Точність (%)**	**F1**	**AUC**
CNN (лише просторові)	85.3	0.82	0.88
RNN (лише часові)	87.1	0.84	0.89
CNN + RNN (гібридний)	90.5	0.88	0.92

Валідація запропонованого методу проводилася шляхом порівняння з існуючими підходами до розпізнавання контексту (таблиця 3.3). Результати свідчать про те, що інтеграція часових патернів AU забезпечує суттєвий приріст точності порівняно з методами, що ігнорують темпоральну динаміку.

Таблиця 3.3.

Порівняння точності розпізнавання контексту з альтернативними методами

Метод	**Точність (%)**	**F1 оцінка**
Метод 1 (лише статичні вирази обличчя)	78.2	0.75
Метод 2 (лише фізіологічні сигнали)	80.5	0.78
Метод (часові патерни AU)	90.5	0.88

Детальний аналіз випадків некоректної класифікації показав, що зниження точності переважно спостерігається у сценаріях з високою контекстуальною складністю або у випадках, коли часові патерни AU характеризуються низькою виразністю або неоднозначністю.

Отримані результати емпірично підтверджують гіпотезу про те, що часові патерни AU є високоефективними дескрипторами для автоматизованого розпізнавання контексту. Це відкриває перспективи для вдосконалення систем афективних обчислень, особливо в умовах, де контекстуальна інформація є критичною для інтерпретації емоційного стану.

Перспективним вбачається дослідження більш складних архітектур глибокого навчання (наприклад, трансформерів) та інтеграція додаткових модальностей, таких як акустичні характеристики (аудіо) або семантичний аналіз тексту, що потенційно дозволить підвищити робастність та точність мультимодальних систем розпізнавання емоцій та контексту.

Висновки до розділу

У третьому розділі було реалізовано практичну інтеграцію теоретичних положень та методів мультимодального аналізу емоцій, що дозволило оцінити їхню ефективність у реальних сценаріях. Запропонований підхід до поєднання фізіологічних сигналів і патернів одиниць дії продемонстрував, що такі комбінації забезпечують суттєве підвищення точності розпізнавання складних емоційних станів. Дослідження наборів даних показало, що фізіологічні параметри є чутливими до внутрішніх афективних змін, тоді як лицьові експресії зберігають інформативність щодо поведінкових аспектів емоцій. У межах методології розпізнавання болю було встановлено, що мультимодальні моделі дозволяють коректно ідентифікувати як явні, так і слабко виражені ознаки цього стану. Часова синхронізація модальностей виявилася критично важливою, оскільки несинхронні сигнали можуть спотворювати результати і знижувати узагальнювальні властивості моделей.

Порівняння результатів із сучасними методами підтвердило переваги застосування інтегрованих технік злиття ознак, які здатні працювати із зашумленими та неоднорідними даними.

Аналіз ефективності показав, що запропонований підхід є універсальним і може масштабуватися на інші емоційні категорії, що відкриває перспективи його практичного використання. Результати експериментів довели, що мультимодальні моделі забезпечують більш стабільну інтерпретацію динаміки емоцій, ніж одномодальні рішення, особливо в критично важливих сферах, таких як медицина чи людиноорієнтовані системи.

Таким чином, третій розділ підтвердив, що мультимодальне злиття інформації є ключовою умовою точного моделювання емоцій у складних сценаріях взаємодії.

ВИСНОВКИ

У ході дослідження, спрямованого на аналіз та моделювання емоцій у контексті релевантних та похідних даних, була розроблена системна методологія, що інтегрує сучасні підходи комп'ютерного зору, аналізу поведінкових патернів, нейронних мереж та мультимодального злиття інформації. Отримані результати дозволили сформувати комплексне бачення природи емоційної експресії, її залежності від контекстуальних чинників та можливостей автоматизованої інтерпретації емоційних станів за допомогою комбінованих джерел даних.

Перший розділ роботи продемонстрував фундаментальну складність процесів розпізнавання та моделювання емоцій. Проведений аналіз предметної області включав дослідження концептуальних основ емоційних реакцій, механізмів формування експресивних проявів та особливостей людської поведінки, що унеможливлюють їх інтерпретацію виключно на основі статичних ознак. Поглиблений огляд одиниць дії (Action Units), як базових структурних компонентів лицьових експресій, засвідчив їхню важливість у побудові формальних ознак для моделей машинного навчання. Разом з тим встановлено, що AU-патерни мають бути розглядатися не ізольовано, а в контексті ситуаційних факторів, зокрема: соціального середовища, індивідуальних особливостей особи та спонтанності емоційної реакції. Аналіз методологій автоматичного розпізнавання спонтанних AU підкреслив значну роль якісних наборів даних, точних методів екстракції ознак та адаптивних алгоритмів класифікації, що дозволяють підвищувати достовірність прогнозів.

У другому розділі було реалізовано поглиблене дослідження самооцінних емоцій — складного класу афективних станів, тісно пов'язаних зі самоідентифікацією суб'єкта, його особистісними уявленнями та соціально-поведінковими установками. Аналіз даних набору BP4D+ продемонстрував значну варіативність самооцінних емоцій, їхню чутливість

до контексту та відображення у мікродинаміці лицьових одиниць дії. Запропонована архітектура нейронної мережі, орієнтована на мультикомпонентне представлення експресивних ознак, забезпечила ефективне виявлення кореляцій між AU-патернами та самооцінними станами. Валідація моделей підтвердила доцільність використання комбінованого підходу, який інтегрує структурні ознаки, часову інформацію та внутрішні самооцінні дані суб'єктів. Порівняльний аналіз також довів переваги запропонованої методології, зокрема її здатність працювати з обмеженими та нерівномірними даними, характерними для дослідження інтимних та індивідуалізованих емоційних проявів.

Третій розділ роботи був присвячений практичній імплементації підходів до моделювання емоцій на основі релевантних та похідних даних. Було розроблено дві методології мультимодального аналізу — для загальної оцінки емоцій та для детекції болю. Інтеграція фізіологічних сигналів із патернами одиниць дії показала, що різні модальності доповнюють одна одну, зменшуючи похибки, пов'язані із зашумленістю та неоднозначністю кожного окремого каналу інформації. Методологія мультимодального злиття забезпечила можливість врахування як швидкоплинних фізіологічних реакцій, так і структурних поведінкових змін у лицьових експресіях, що значно підвищило точність розпізнавання емоційних станів.

Особливу увагу приділено аналізу ефективності злиття ознак у задачах розпізнавання болю — однієї з найбільш складних афективних категорій через її високий ступінь суб'єктивності та спонтанності. Дослідження продемонструвало, що комбінування даних дозволяє моделі не лише точніше ідентифікувати ознаки болю, але й розрізняти його інтенсивність, що підтверджено порівнянням із сучасними підходами в галузі. Крім того, встановлено, що AU-патерни, розглянуті у часовому вимірі, є критично важливими для контекстуалізації емоційних станів, формування структурованих поведінкових моделей та побудови алгоритмів розпізнавання, здатних адаптуватися до динамічних змін.

Комплексний аналіз отриманих результатів на всіх етапах роботи засвідчив, що моделювання емоцій на основі релевантних та похідних даних є перспективним напрямом, який вимагає інтеграції різнорідних інформаційних джерел, використання мультимодальних моделей та урахування контексту. У межах дослідження підтверджено, що поєднання лицьових експресій, фізіологічних реакцій та внутрішніх самооцінних даних забезпечує набагато вищу точність і стійкість прогнозів порівняно з одномодальними підходами. Запропоновані методології демонструють можливість побудови систем, здатних до глибинного інтерпретаційного аналізу емоційних станів, що відкриває перспективи для створення ефективних рішень у сфері медицини, соціальної робототехніки, адаптивних інтерфейсів та інтелектуальних рекомендаційних систем.

Загалом результати магістерської роботи формують підґрунтя для подальших досліджень у напрямі мультимодального моделювання емоцій, зокрема щодо автоматизації обробки спонтанних емоційних проявів, удосконалення структурних моделей AU-динаміки та інтеграції контекстуальних даних широкого спектра.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Learning Spatial and Temporal Cues for Multi-label Facial Action Unit Detection . - Wen-Sheng Chu // https://www.ri.cmu.edu/pub_files/2017/5/ant_low.pdf
2. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), Second Edition
3. Paul Ekman Erika L. Rosenberg. – https://www.aqualide.com/upload/texte/text98.pdf
4. FACS3D-Net: 3D Convolution based Spatiotemporal Representation for Action Unit Detection . - Le Yang // https://www.jeffcohn.net/wp-content/uploads/2019/07/ACII2019_3DCNN.pdf
5. Frontiers | D-PAttNet: Dynamic Patch-Attentive Deep Network for Action Unit Detection - https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2019.00011/full#F1
6. Dual Learning for Facial Action Unit Detection Under Nonfull Annotation / Shangfei Wan. - https://bigdata.ustc.edu.cn/paper_pdf/2020/Shangfei-Wan-ITC2020,.pdf
7. Al-Eidan, Rasha M., Hend Al-Khalifa, and AbdulMalik Al-Salman. Deep-Learning-Based Models for Pain Recognition: A Systematic Review. Applied Sciences, 10 (17), 5984, MDPI, 2020.
8. Bhatti, Anubhav, Behnam Behinaein, Dirk Rodenburg, Paul Hungler, and Ali Etemad. “Attentive Cross-modal Connections for Deep Multimodal Wearable-based Emotion Recognition.” ArXiv preprint, arXiv:2108.02241, 2021.
9. Egede, Joy O., Siyang Song, Temitayo A. Olugbade, Chongyang Wang, Amanda Williams, Hongying Meng, Min Aung, Nicholas D. Lane, Michel Valstar, and Nadia Bianchi-Berthouze. EMOPAIN Challenge 2020: Multimodal Pain Evaluation from Facial and Bodily Expressions. ArXiv preprint, arXiv:2001.07739, 2020.

10. Farmani, Javad, Maryam Fard, Forouzan Golkar, and Sina Khorasgani. “A CrossMod-Transformer Deep Learning Framework for Multimodal Pain Recognition Using ECG and EDA Signals.” Scientific Reports, 15, (2025), Article 14238, Nature Publishing Group, 2025.
11. Guo, Yunfei, Tao Zhang, and Wu Huang. “Emotion Recognition Based on Multimodal Electrophysiology: Multi-Head Attention Contrastive Learning.” ArXiv preprint, arXiv:2308.01919, 2023.
12. Kollias, Dimitrios, and Stefanos Zafeiriou. “Affect Analysis In-The-Wild: Valence-Arousal, Expressions, Action Units and a Unified Framework.” ArXiv preprint, arXiv:2103.15792, 2021.
13. Liu, Wei, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. “Multimodal Emotion Recognition Using Deep Canonical Correlation Analysis.” arXiv:1908.05349, 2019.
14. Nadeeshani, Madhuka, et al. “Facial Emotion Prediction through Action Units and Deep Learning.” Conference Paper, 2021.
15. Phan, Kim Ngan, Ngumimi Karen Iyor Tsuun, Sudarshan Pant, Hyung-Jeong Yang, and Soo-Hyung Kim. “Pain Recognition With Physiological Signals Using Multi-Level Context Information.” IEEE Access, (2017).)
16. Picard, Marie-Eve, Miriam Kunz, Jen-I Chen, Pierre Rainville, et al. “A Distributed Brain Response Predicting the Facial Expression of Acute Nociceptive Pain: The Facial Expression Pain Signature (FEPS).” eLife, 13, e87962, eLife Sciences Publications, 2024.
17. Pons, Gerard, and David Masip. “Multi-Task, Multi-Label and Multi-Domain Learning with Residual Convolutional Networks for Emotion Recognition, arXiv:1802.06664, 2018.
18. Thiam, Patrick, Heinke Hihn, Daniel A. Braun, and Friedhelm Schwenker. “Multi-Modal Pain Intensity Assessment Based on Physiological Signals: A Deep Learning Perspective.” Frontiers in Physiology, 12, 720464, Frontiers Media, 2021.

19. Wang, Zhifeng. “End-to-End Multimodal Emotion Recognition by Using Deep Neural Networks.” Conference Paper (ABCs 2022), Australian National University, 2022.
20. Yang, Xudong, Hongli Yan, Anguo Zhang, Pan Xu, Sio Hang Pan, Mang I. Vai, and Yueming Gao. “Emotion Recognition Based on Multimodal Physiological Signals Using Spiking Feed-Forward Neural Networks.” Biomedical Signal Processing and Control, 82, 105921, Elsevier, 2024.
21. Zhi, R., et al. “Action Unit Analysis Enhanced Facial Expression Recognition via Evolutional Deep Learning.” Neurocomputing, 452, 2021, pp. 389–402
22. Belhouchette, Karim, et al. “Facial Action Units Detection to Identify Interest Emotion.” Journal of Intelligent & Fuzzy Systems, 42 (2022), pp. 1–12. World Scientific.
23. Scanavan, L., et al. “Multimodal Physiological-Based Emotion Recognition.” Proceedings of ICPRW 2020 (Workshops), 2020.
24. Wang, Zhuozheng, and Yihan Wang. “Emotion Recognition Based on Multimodal Physiological Electrical Signals.” Frontiers in Neuroscience, 19, 1512799, Frontiers Media, 2025.
25. Picard, M.-E., Kunz, M., Rainville, P., et al. “Brain Signature for Pain Expression (FEPS): Predicting the Facial Action Coding System Composite Score from fMRI.” BioRxiv preprint, 2023.
26. Benavent-Lledó, M., et al. “PainFusion+: A Multimodal Transformer Architecture for Multimodal Pain Assessment.” Computer Methods and Programs in Biomedicine, 240, 2025.
27. Fang, R., Liao, Y., & others. “Survey on Pain Detection Using Machine Learning Models.” JMIR Artificial Intelligence, 1, e53026, JMIR Publications, 2025.
28. Science Review: Liao, Yilong, Yuan Gao, Fang Wang, Li Zhang, Zhenrong Xu, and Yifan Wu. “Emotion Recognition with Multiple Physiological

Parameters Based on Ensemble Learning.” Scientific Reports, 15, 19869, Nature Publishing Group, 2025.

29. Nadeeshani, M., et al. “Facial Emotion Prediction Through Action Units and Deep Learning.” International Conference on Affective Computing and Intelligent Interaction, 2021.
30. Egede, J. O., et al. “EMOPAIN Challenge 2020: Multimodal Pain Evaluation from Facial and Bodily Expressions.” Proceedings of International Conference on Multimedia and Affective Computing, 2020.
31. Liu, W., Qiu, J., Zheng, W., Lu, B.-L. “Multimodal Emotion Recognition Using Deep Canonical Correlation Analysis.” IEEE Transactions on Neural Networks & Learning Systems, 2024
32. Yang, X., Yan, H., Zhang, A., Xu, P., Pan, S. H., Vai, M. I., & Gao, Y. “Emotion Recognition Based on Multimodal Physiological Signals Using Spiking Feed-Forward Neural Networks.” Biomedical Signal Processing and Control, 2024.
33. Zhi, R., et al. “Action Unit Analysis Enhanced Facial Expression Recognition via Evolutionary Deep Learning.” IEEE Transactions on Affective Computing, 2021.
34. Belhouchette, K., et al. “Facial Action Units Detection to Identify Interest Emotion.” Journal of Intelligent & Fuzzy Systems, 2022.
35. Scanavan, L., et al. “Multimodal Physiological-Based Emotion Recognition.” International Conference on Pattern Recognition Workshops (ICPRW), 2020.
36. Wang, Z., & Wang, Y. “Multimodal Physiological Emotion Recognition using EEG and ECG via Att-1DCNN-GRU.” Frontiers in Neuroscience, 2025.
37. Ekman, P., & Friesen, W. V. (2018). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, CA: Consulting Psychologists Press.

38. Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., & Girard, J. M. (2016). "BP4D+: Multimodal Spontaneous Emotion Corpus." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1577–1585). Las Vegas, NV, USA: IEEE.
39. Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). "DISFA: A Spontaneous Facial Action Intensity Database." IEEE Transactions on Affective Computing, 4(2), 151–160.
40. Lucey, P., Cohn, J. F., Prkachin, K. M., Solomon, P. E., & Matthews, I. (2011). "Painful data: The UNBC-McMaster shoulder pain expression archive database." Image and Vision Computing, 29(6), 358–370.
41. Aung, M. S., Kaltwang, S., Romera-Paredes, B., Martinez, B., Singh, A., Caine, M., & Pantic, M. (2016). "The Automatic Detection of Chronic Pain-Related Expression: Requirements, Challenges and the EmoPain Corpus." IEEE Transactions on Affective Computing, 7(4), 435–451.
42. Yang, H., Ciftci, U., & Yin, L. (2019). "FACS3D-Net: 3D Convolutional Neural Networks for Facial Action Unit Detection." In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (pp. 2262–2270). Seoul, South Korea: IEEE.
43. Mittal, T., Guhan, P., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020). "EmotiCon: Context-Aware Multimodal Emotion Recognition using Frege's Principle." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 14234–14243). Seattle, WA, USA: IEEE.