МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

ТАРАСЮК (БРОЩАК) Софія Іванівна

Метод розпізнавання емоцій мовлення на основі глибокого навчання /Speech Emotion Recognition Method Based on Deep Learning

спеціальність: 122 - Комп'ютерні науки
освітньо-професійна програма - Комп'ютерні науки

Кваліфікаційна робота

Виконав студент групи КНм-21
Тарасюк (Брощак) С.І.
_____

Науковий керівник:
к.т.н., доцент, Ліп'яніна-Гончаренко Х.В.
_____

Кваліфікаційну роботу
допущено до захисту:
«___» _____ 20___ р.
В.о. завідувача кафедри
_____ Н.М. Васильків

ТЕРНОПІЛЬ - 2024

Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління
Освітній ступінь «магістр»
спеціальність: 122 – Комп'ютерні науки
освітньо-професійна програма – Комп'ютерні науки

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ М.П. Комар
«____» _____20__р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
ТАРАСЮК (БРОЩАК) Софія Іванівна

1. Тема кваліфікаційної роботи
Метод розпізнавання емоцій мовлення на основі глибокого навчання /Speech Emotion Recognition Method Based on Deep Learning
керівник роботи к.т.н., доцент, Ліп'яніна-Гончаренко Х.В.
затверджені наказом по університету від 12 грудня 2023 року № 753.

2. Строк подання студентом закінченої кваліфікаційної роботи 4 грудня 2024 р.

3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.

4. Основні питання, які потрібно розробити
- Проаналізувати проблему класифікації емоцій через мовлення.
- Оглянути існуючі підходи до класифікації емоцій, їхні переваги та недоліки.
- Розробити алгоритми та інформаційну базу для IG-CNN.
- Реалізувати обмеження інтерпретованості для набору GeMAPS.
- Забезпечити некорельованість репрезентацій у груповій CNN.
- Розробити метод розпізнавання емоцій на основі мовлення.
- Підготувати набори даних для експериментів та їх розподілу.
- Провести експерименти для оцінки ефективності IG-CNN.
- Дослідити інтерпретованість моделі IG-CNN та її взаємодії.

5. Перелік графічного матеріалу у роботі
- Фреймворк інтерпретованої групової згорткової нейронної мережі.

6. Консультанти розділів кваліфікаційної роботи

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

7. Дата видачі завдання 12 грудня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	Затвердження теми кваліфікаційної роботи, ознайомлення з літературними джерелами та складання плану роботи.	до 01.01. 2024 р.	
2	Написання 1 розділу кваліфікаційної роботи	до 01.03. 2024 р.	
3	Написання 2 розділу кваліфікаційної роботи	до 20.05.2024 р.	
4	Написання 3 розділу кваліфікаційної роботи	до 28.10. 2024 р.	
5	Представлення попереднього варіанту кваліфікаційної роботи, перевірка та внесення змін керівником	до 11.11.2024 р.	
6	Опрацювання зауважень та представлення завершеного варіанту кваліфікаційної роботи. Підготовка супроводжуючих документів.	до 25.11.2024 р.	
7	Перевірка кваліфікаційної роботи на оригінальність тексту.	до 30.11.2024 р.	
8	Оформлення кваліфікаційної роботи та отримання допуску до захисту	до 04.12.2024 р.	
9	Подання кваліфікаційної роботи до захисту на засіданні атестаційної комісії.	до 14.12. 2024 р.	

Студент _____ Тарасюк (Брощак) С.І.
підпис

Керівник роботи _____ к.т.н., доцент, Ліп’яніна-Гончаренко Х.В.
підпис

РЕЗЮМЕ

Кваліфікаційна робота на тему «Метод розпізнавання емоцій мовлення на основі глибокого навчання» на здобуття освітнього ступеня «Магістр» зі спеціальності 122 «Комп'ютерні науки» освітньої програми «Комп'ютерні науки» написана обсягом в 82 сторінок і містить 11 ілюстрацій, 5 таблиць, 6 додатків та 38 використаних джерел.

Метою роботи є розробка методу розпізнавання емоцій на основі мовлення з використанням глибоких нейронних мереж для досягнення високої точності класифікації емоційних станів.

Методи дослідження: використання згорткових нейронних мереж (CNN), рекурентних мереж (RNN), методів обробки аудіосигналів, аугментація даних та статистичні методи оцінки ефективності моделей.

Результати дослідження: запропоновано модель IG-CNN, яка поєднує інтерпретовані та автономні репрезентації, що забезпечує високу точність розпізнавання емоцій навіть у складних умовах. Модель перевірена на наборах даних IEMOCAP, RAVDESS, eNTERFACE'05 та CREMA-D і показала перевагу над базовими методами.

Практичне значення роботи: метод може бути застосований у системах голосового управління, психологічного аналізу, автоматизованого обслуговування клієнтів та в освітніх і медичних системах.

Ключові слова: РОЗПІЗНАВАННЯ ЕМОЦІЙ, ГЛИБОКЕ НАВЧАННЯ, ЗГОРТКОВІ МЕРЕЖІ, АУДІОАНАЛІЗ, АУГМЕНТАЦІЯ.

ABSTRACT

The qualification thesis titled "Speech Emotion Recognition Method Based on Deep Learning" for obtaining the Master's degree in the specialty 122 "Computer Science" of the educational program "Computer Science" comprises 82 pages and includes 11 illustrations, 5 tables, 6 appendices, and 38 references.

The aim of the thesis is to develop a speech emotion recognition method based on deep neural networks to achieve high accuracy in classifying emotional states.

Research methods: the use of convolutional neural networks (CNN), recurrent neural networks (RNN), audio signal processing methods, data augmentation, and statistical methods for evaluating model efficiency.

Research results: The IG-CNN model was proposed, combining interpretable and autonomous representations to ensure high accuracy in emotion recognition, even under challenging conditions. The model was tested on datasets such as IEMOCAP, RAVDESS, eNTERFACE'05, and CREMA-D, outperforming baseline methods.

Practical significance: The method can be applied in voice control systems, psychological analysis, automated customer service, and educational and medical systems.

Keywords: EMOTION RECOGNITION, DEEP LEARNING, CONVOLUTIONAL NETWORKS, AUDIO ANALYSIS, AUGMENTATION.

ЗМІСТ

Вступ 8
1 Аналіз предметної області і постановка задачі дослідження 12
1.1 Огляд та визначення проблеми класифікації емоцій 12
1.2 Огляд існуючих підходів для класифікації емоцій 14
1.3 Вибір перспективного шляху і постановка задачі дослідження 16
Висновки до розділу 1 18
2 Алгоритмічне та інформаційне забезпечення розпізнавання емоцій за допомогою мовлення 20
2.1 Розділення інтерпретованих та автономних репрезентацій у рамках IG-CNN для розпізнавання емоцій 20
2.2 Обмеження інтерпретованості для навчання моделі 21
2.3 Групова CNN та некорельованість репрезентацій 24
2.4 Метод розпізнавання емоцій на основі мовлення 26
Висновки до розділу 2 29
3 Програмно-технологічне забезпечення 31
3.1 Набори даних та умови експерименту 31
3.2 Загальна ефективність 34
3.3 Дослідження впливу компонентів моделі 34
3.4 Аналіз чутливості 36
3.5 Тестування роботи моделі в різних ситуаціях 41
3.6 Аналіз інтерпретованості 44
Висновки до розділу 3 47
Висновки 49
Список використаних джерел 52
Додаток А Порівняння ефективності з базовими методами на тестовому наборі 57
Додаток Б Порівняння продуктивності з варіантами IG-CNN 60

Додаток В Продуктивність IG-CNN при комбінаціях навчання швидкості та розміру партії 62
Додаток Г Т-критерій вікової специфічності CEP на CREMA-D 64
Додаток Д Продуктивність моделі при різних шуму рівнях 67
Додаток Е Апробація отриманих результатів 70

ВСТУП

Актуальність дослідження полягає в зростаючому інтересі до автоматичного розпізнавання емоцій, особливо в контексті людського мовлення. У сучасному світі емоції відіграють ключову роль у спілкуванні, впливаючи на якість та ефективність взаємодії між людьми. З розвитком технологій і зростанням застосування штучного інтелекту, особливо в галузях, пов'язаних із розпізнаванням мови та емоцій, виникла потреба у створенні систем, здатних аналізувати емоційний стан людини на основі аудіосигналів. Це має широке застосування в різних сферах, таких як освіта, медицина, розробка персональних асистентів, системи безпеки, розважальні системи, а також у психологічній та соціальній адаптації.

Однією з головних причин актуальності цього дослідження є обмеженість існуючих підходів до розпізнавання емоцій у мовленні. Багато традиційних методів обмежені у здатності точно інтерпретувати складні та динамічні емоційні стани через різноманітність мовних ознак, таких як інтонація, темп, тон, паузи та інші характеристики голосу. Нинішні методи часто покладаються на ручну інженерію ознак, що робить їх менш гнучкими та придатними для застосування в реальних сценаріях. Ці обмеження створюють потребу в розробці нових підходів, які зможуть автоматично виділяти релевантні ознаки з мовних сигналів та враховувати їхню складну природу.

З появою та розвитком методів глибокого навчання з'явилася можливість значно підвищити точність і надійність систем розпізнавання емоцій. Глибокі нейронні мережі, зокрема згорткові (CNN) та рекурентні нейронні мережі (RNN), продемонстрували високу ефективність в обробці та аналізі аудіосигналів. Вони дозволяють автоматично навчатися інформативним ознакам з необроблених даних, усуваючи необхідність ручної інженерії ознак. Дослідження в цій області спрямовані на

вдосконалення таких моделей, підвищення їх точності та розширення можливостей застосування у різних реальних умовах, включаючи ситуації з шумом, різними акцентами та індивідуальними особливостями мовців.

Актуальність цього дослідження також підкріплюється зростаючими вимогами до покращення якості взаємодії між людиною та машиною. У міру того, як системи штучного інтелекту стають все більш інтегрованими в наше повсякденне життя, виникає необхідність у розробці більш чуйних та "емоційно інтелектуальних" систем, які здатні розпізнавати та враховувати емоційний стан користувачів. Це підвищує ефективність та природність взаємодії, дозволяючи створювати більш персоналізовані та адаптивні системи. Таким чином, розробка методу розпізнавання емоцій у мовленні на основі глибокого навчання є актуальним завданням, яке має значний потенціал для покращення різних аспектів технологій, орієнтованих на людину.

Мета дослідження полягає в розробці методу розпізнавання емоцій на основі мовлення з використанням глибоких нейронних мереж, який забезпечить високу точність класифікації емоційних станів на основі аудіосигналів.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких завдань:

1. Провести аналіз проблеми класифікації емоцій та визначити її ключові аспекти в контексті розпізнавання емоцій за допомогою мовлення.

2. Оглянути та проаналізувати існуючі підходи до класифікації емоцій, визначити їхні переваги та недоліки, а також обґрунтувати вибір перспективного підходу для розпізнавання емоцій.

3. Розробити алгоритмічне та інформаційне забезпечення для розділення інтерпретованих та автономних репрезентацій у рамках IG-CNN для розпізнавання емоцій.

4. Визначити та реалізувати обмеження інтерпретованості для навчання моделі на основі набору GeMAPS, включаючи обмеження на евклідову відстань між вивченими та створеними людиною репрезентаціями.

5. Реалізувати групову CNN та забезпечити некорельованість репрезентацій для підвищення ефективності розпізнавання емоцій.

6. Розробити метод розпізнавання емоцій на основі мовлення, включаючи етапи аугментації даних, екстракції ознак, побудови та навчання моделі.

7. Підібрати та підготувати набори даних для експериментів, встановити умови експерименту, включаючи процедури розподілу даних на навчальні, валідаційні та тестові набори.

8. Провести експерименти для оцінки загальної ефективності моделі IG-CNN порівняно з базовими моделями, а також дослідити вплив компонентів моделі та гіперпараметрів на її продуктивність.

9. Здійснити аналіз інтерпретованості моделі IG-CNN, досліджуючи взаємодію між інтерпретованими та автономними репрезентаціями та їх внесок у точність розпізнавання емоцій.

Об’єктом дослідження є процес розпізнавання емоцій за допомогою мовлення, включаючи етапи обробки аудіосигналів, екстракції ознак та класифікації емоційних станів з використанням методів глибокого навчання.

Предметом дослідження є моделі та методи глибокого навчання, такі як згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN), які застосовуються для аналізу та класифікації емоційних ознак мовлення з метою підвищення точності розпізнавання емоцій.

Методи дослідження включають застосування глибокого навчання, зокрема згорткових та рекурентних нейронних мереж, для розробки моделі розпізнавання емоцій на основі мовлення. Було використано методи попередньої обробки аудіоданих, такі як нормалізація, аугментація та виділення ознак (Mel-frequency cepstral coefficients, Chroma_stft, zero crossing

rate), з метою покращення якості вхідних сигналів. Ефективність моделі оцінювалася експериментальними методами на різних наборах даних, а також під час тестування з різними параметрами навчання. Для аналізу результатів застосовувалися статистичні методи, які дозволяли порівняти продуктивність моделі з існуючими підходами та провести t-тести для оцінки статистичної значущості отриманих даних.

Наукова новизна отриманих результатів полягає в розробці методу розпізнавання емоцій з мовлення на основі глибокого навчання, який поєднує інтерпретовані та автономні репрезентації для підвищення точності та пояснюваності моделі, вирішуючи проблему кореляції між ознаками, що відрізняє його від існуючих підходів.

Практичне значення отриманих результатів полягає в можливості застосування розробленого методу розпізнавання емоцій з мовлення в різних сферах, таких як системи голосового управління, психологічний аналіз, автоматичний супровід клієнтів, а також в освітніх і медичних застосуваннях, де необхідний аналіз емоційного стану людини для покращення взаємодії та надання персоналізованих послуг.

Структура та обсяг роботи. Кваліфікаційна робота складається зі вступу, трьох основних розділів, висновків, списку літератури та додатків.

Апробація результатів дослідження. Основні теоретичні положення роботи й практичні результати дослідження доповідалися й обговорювалися: V Міжнародної наукової конференції «Стратегічні напрямки розвитку науки: фактори впливу та взаємодії», яка відбулася 11 жовтня 2024 року у місті Луцьк, Україна; III Міжнародної наукової конференції «Інтелектуальний ресурс сьогодення: наукові задачі, розвиток та запитання», яка відбулася 20 вересня 2024 року у місті Одеса, Україна.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

1.1 Огляд та визначення проблеми класифікації емоцій

Класифікація емоцій за аудіоданими є складною задачею в галузі обробки природної мови (NLP) та аналізу сигналів, яка вимагає точного визначення та аналізу людських емоцій. Основна мета таких систем полягає у визначенні емоційного стану особи на основі її голосових сигналів. Цей процес включає етапи збору аудіоданих, їх попередньої обробки, виділення характеристик і подальшої класифікації. У таких дослідженнях зазвичай розглядаються базові емоції, як-от радість, сум, гнів, страх, здивування та відраза, хоча можуть враховуватися також складніші та більш нюансовані емоційні стани.

Емоційний стан людини (рисунок 1.1) характеризується специфічними акустичними ознаками, такими як висота тону, інтенсивність, швидкість мовлення та частотні характеристики, які можна виділити для подальшого аналізу. Проте, варіативність цих ознак залежно від індивідуальних особливостей, контексту, а також присутність шуму в записах створюють серйозні виклики для розробників таких систем. Вибір та оптимізація методів виділення ознак і обробки сигналів, що здатні мінімізувати вплив цих факторів, є критично важливими.

Наукова спільнота активно застосовує методи машинного навчання, особливо глибинного навчання, для розв'язання завдань класифікації емоцій. Конволюційні нейронні мережі (CNN), рекурентні нейронні мережі (RNN) та їх комбінації (CRNN) демонструють високі показники ефективності в цьому напрямі. Однак створення надійних моделей вимагає великого обсягу анотованих даних, які нерідко є обмеженим ресурсом, особливо для рідкісних чи недостатньо представлених емоцій.

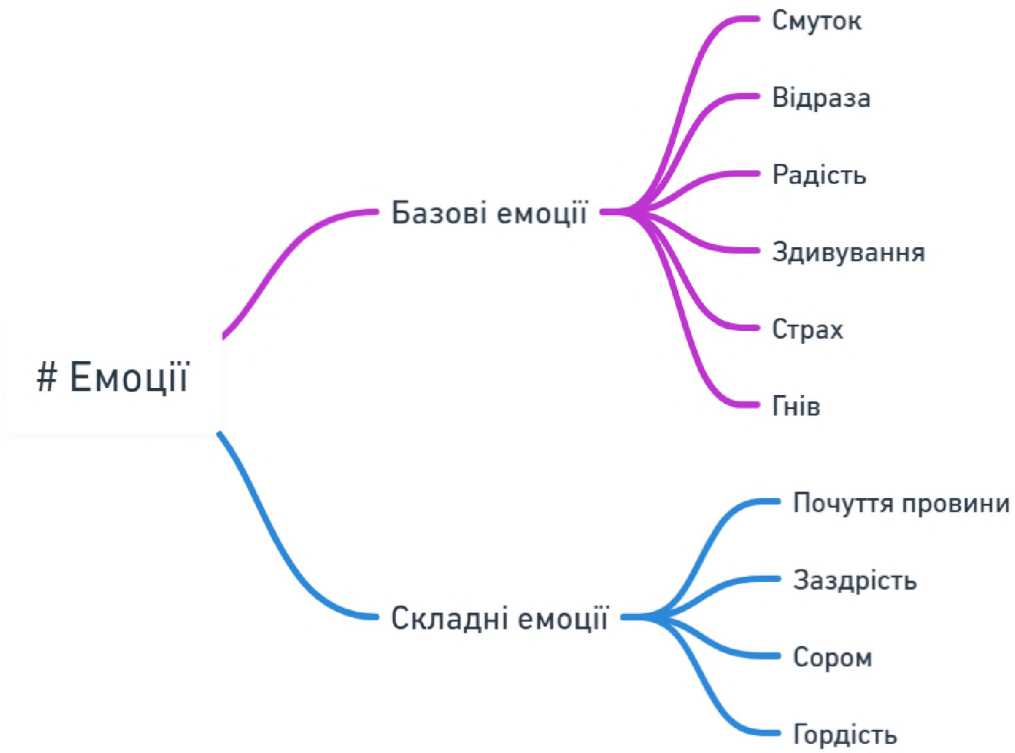


Рисунок 1.1 – Типи емоцій

Крім того, існує значний ризик перенавчання моделі на специфічних характеристиках окремого набору даних, що може знизити здатність моделі до узагальнення результатів на ширший спектр даних. Це потребує використання методів регуляризації та впровадження стратегій для збалансування класів, щоб забезпечити рівномірне навчання моделі для всіх категорій емоцій.

Класифікація емоцій за аудіоданими є складним завданням, яке вимагає глибокого розуміння як лінгвістичних, так і нелінгвістичних особливостей голосу, а також здатності ефективно аналізувати та обробляти аудіосигнали. Також важливо враховувати соціокультурні та індивідуальні фактори, що впливають на розпізнавання емоцій, для забезпечення точності та універсальності системи класифікації.

1.2 Огляд існуючих підходів для класифікації емоцій

Визначення емоційних станів із голосових даних здійснюється за допомогою множини методів, які можна умовно розділити на дві головні категорії: класичні машинні алгоритми навчання та сучасні глибинні нейронні мережі. Класичні підходи [1 - 4] включають методи, такі як машини опорних векторів (SVM), k-найближчих сусідів (k-NN), наївний Байєсовий класифікатор, та дерева рішень. Ці методи зазвичай вимагають ретельної предобробки даних та вибору характеристик, що є критичними для їх ефективності. Наприклад, можуть використовуватися такі ознаки, як частота основного тону (F0), енергія, мел-частотні кепстральні коефіцієнти (MFCCs), і ритм мовлення.

З іншого боку [5 - 7], методи глибинного навчання, зокрема конволюційні нейронні мережі (CNN) та рекурентні нейронні мережі (RNN), зокрема, з довгою короткочасною пам'яттю (LSTM) та блоками затриманої пам'яті (GRU), дозволяють автоматизувати процес екстракції ознак, опираючись на великі обсяги даних для виявлення складних неявних зв'язків в даних. Глибинні нейронні мережі демонструють високу здатність до узагальнення і здатні опрацьовувати сирі аудіосигнали без жорсткої предобробки, тим самим підвищуючи точність класифікації.

Останніми роками глибинне навчання стало особливо популярним завдяки його здатності до самонавчання на основі неструктурованих даних. Зокрема, мережі з увагою (attention networks) та трансформери (transformers) відкривають нові перспективи для визначення емоцій, дозволяючи моделям зосереджуватися на найбільш інформативних частинах аудіосигналу.

Крім алгоритмічного аспекту, важливими є стратегії навчання та аугментації даних. Такі техніки, як шуміння, зміна швидкості мовлення, та пітч-шіфтинг, використовуються для збільшення обсягів навчального датасету та підвищення стійкості моделі до варіацій в даних. Також,

використовується метод ансамблювання, який об'єднує результати кількох моделей для поліпшення загальної точності.

Слід відзначити, що в дослідженнях емоційного розпізнавання враховуються й культурні та індивідуальні варіативності, оскільки емоційна експресія може суттєво варіюватися залежно від культурних та особистісних характеристик осіб. Тому ефективність класифікації вимагає високої адаптивності та здатності моделі до генералізації на різноманітні демографічні та культурні групи.

У даному дослідженні зосереджено на аналізі різних алгоритмів розпізнавання емоцій за допомогою мови, враховуючи їхні переваги та недоліки. Таблиця 1.1 нижче надає порівняльний огляд ключових характеристик кожного розглянутого алгоритму, що дозволяє глибше зрозуміти їхні особливості та обмеження в контексті розпізнавання емоцій.

Таблиця 1.1 - Порівнює переваги та недоліки різних алгоритмів розпізнавання емоцій за допомогою мови

Назва дослідження	**Переваги**	**Недоліки**
Розпізнавання емоцій за допомогою мови з використанням машини підтримки векторів [1]	Висока точність завдяки вдосконаленій селекції ознак.	Необхідність ретельного вибору ознак, які впливають на результати.
Вплив алгоритму вибору ознак на розпізнавання емоцій за допомогою глибоких згорткових нейронних мереж [9]	Застосування глибокого навчання дозволяє ефективно враховувати різноманітність емоційних станів.	Потреба у великих обсягах даних та значні вимоги до обчислювальних ресурсів.
Алгоритми вилучення ознак для підвищення точності розпізнавання емоцій [10]	Підвищення точності розпізнавання завдяки новим методам вилучення ознак.	Може знадобитися додаткове налаштування для оптимальної роботи в різних умовах.
Автоматичне розпізнавання емоцій за допомогою рекурентних нейронних мереж з локальною увагою	Використання механізмів уваги допомагає зосередитись на ключових аспектах мовлення.	Висока складність моделі, що може впливати на швидкість обробки.
Розпізнавання емоцій за допомогою мови: ознаки та моделі класифікації [11]	Огляд різних алгоритмів класифікації надає гнучкість у виборі підходу.	Деякі алгоритми можуть не бути достатньо ефективними в специфічних сценаріях розпізнавання.

Дане дослідження вирізняється (див. таблицю 1.1) передовим підходом до розпізнавання емоцій за допомогою мови, що об'єднує в собі інноваційні

методики збору та обробки даних, а також застосування передових технологій машинного навчання. Особливість нашого підходу полягає в інтеграції аугментації аудіоданих, яка дозволяє збільшити різноманітність навчальних даних без потреби в їх додатковому зборі, що є вирішальним для точності моделювання емоцій. Комбінування цього з ретельною екстракцією ознак, таких як Mel-frequency cepstral coefficients (MFCC), і застосуванням глибоких нейронних мереж з Conv1D та MaxPooling1D шарами, забезпечує високу точність розпізнавання. Такий комплексний підхід дозволяє не тільки точно визначати емоційні стани, але й адаптувати систему під конкретні задачі та умови використання, що ставить наше дослідження на передові позиції у цій галузі.

1.3 Вибір перспективного шляху і постановка задачі дослідження

Актуальність дослідження методу розпізнавання емоцій на основі мовлення обумовлена зростаючою потребою в інтелектуальних системах, здатних розпізнавати й реагувати на емоційний стан людини. У сучасному світі такі системи знаходять широке застосування в різних галузях, включаючи охорону здоров'я, освіту, розважальну індустрію та інтелектуальних помічників. Наприклад, в галузі охорони здоров'я розпізнавання емоцій може сприяти ранньому виявленню та моніторингу психічних розладів, таких як депресія або тривожні розлади. У сфері обслуговування клієнтів, здатність системи розпізнавати емоції з мовлення може значно покращити якість взаємодії з користувачами, адаптуючи відповіді залежно від їхнього емоційного стану.

Зі збільшенням обсягу даних, які щодня генеруються, і вдосконаленням технологій глибокого навчання, стало можливим створення більш точних і ефективних моделей для розпізнавання емоцій. Традиційні методи розпізнавання емоцій на основі мовлення часто обмежені через використання

лише вручну створених ознак або простих моделей класифікації, що не дозволяє повною мірою виявити складні емоційні патерни в мовленні. Інтеграція глибоких нейронних мереж, зокрема згорткових і рекурентних нейронних мереж, дозволяє обробляти великі масиви аудіоданих і автоматично витягувати суттєві ознаки для класифікації емоцій. Це відкриває нові можливості для розвитку інтелектуальних систем, які можуть адаптивно взаємодіяти з користувачами.

Дослідження в цій області має особливу актуальність у контексті розвитку технологій штучного інтелекту, спрямованих на поліпшення взаємодії між людиною і машиною. У той час як більшість сучасних систем розпізнавання емоцій зосереджені на аналізі текстових або візуальних даних, використання мовних сигналів для цієї мети залишається менш розвиненою, хоча й не менш перспективною сферою. Мовлення несе в собі багатий спектр емоційної інформації, яка може бути використана для створення більш інтуїтивних і чутливих систем. Отже, розробка ефективного методу розпізнавання емоцій на основі мовлення з використанням глибокого навчання є актуальним завданням, яке може зробити вагомий внесок у подальший розвиток адаптивних і чутливих до користувача технологій.

Мета дослідження полягає в розробці методу розпізнавання емоцій на основі мовлення з використанням глибоких нейронних мереж, який забезпечить високу точність класифікації емоційних станів на основі аудіосигналів.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких завдань:

1. Провести аналіз проблеми класифікації емоцій та визначити її ключові аспекти в контексті розпізнавання емоцій за допомогою мовлення.

2. Оглянути та проаналізувати існуючі підходи до класифікації емоцій, визначити їхні переваги та недоліки, а також обґрунтувати вибір перспективного підходу для розпізнавання емоцій.

3. Розробити алгоритмічне та інформаційне забезпечення для розділення інтерпретованих та автономних репрезентацій у рамках IG-CNN для розпізнавання емоцій.

4. Визначити та реалізувати обмеження інтерпретованості для навчання моделі на основі набору GeMAPS, включаючи обмеження на евклідову відстань між вивченими та створеними людиною репрезентаціями.

5. Реалізувати групову CNN та забезпечити некорельованість репрезентацій для підвищення ефективності розпізнавання емоцій.

6. Розробити метод розпізнавання емоцій на основі мовлення, включаючи етапи аугментації даних, екстракції ознак, побудови та навчання моделі.

7. Підібрати та підготувати набори даних для експериментів, встановити умови експерименту, включаючи процедури розподілу даних на навчальні, валідаційні та тестові набори.

8. Провести експерименти для оцінки загальної ефективності моделі IG-CNN порівняно з базовими моделями, а також дослідити вплив компонентів моделі та гіперпараметрів на її продуктивність.

9. Здійснити аналіз інтерпретованості моделі IG-CNN, досліджуючи взаємодію між інтерпретованими та автономними репрезентаціями та їх внесок у точність розпізнавання емоцій.

Висновки до розділу 1

1. У цьому розділі було проведено глибокий аналіз проблеми класифікації емоцій за аудіоданими, який включав огляд основних акустичних маркерів, таких як тон, інтенсивність та частотні компоненти, що можуть бути використані для розпізнавання емоційних станів. Розглянуто складності цього завдання, зокрема, варіативність емоційних маркерів між індивідуумами та наявність шумів у записах. Окрему увагу було приділено

сучасним методам машинного та глибинного навчання, зокрема CNN, RNN та їх комбінаціям, які показали значний потенціал у цій галузі. Проте, незважаючи на досягнутий прогрес, було виявлено, що для підвищення точності класифікації емоцій необхідно враховувати соціокультурні та індивідуальні аспекти, а також забезпечити оптимізацію методів екстракції ознак та обробки сигналів.

2. Проведено огляд існуючих підходів для класифікації емоцій, де порівнювалися класичні алгоритми машинного навчання, такі як SVM та k-NN, з сучасними методами глибинного навчання, такими як CNN, RNN та LSTM. Переваги та недоліки кожного методу було докладно розглянуто в контексті їх застосування до завдання розпізнавання емоцій. Особливістю цього дослідження є інтеграція інноваційних методів аугментації даних та глибокого навчання для підвищення точності класифікації емоцій. Представлений підхід забезпечує не лише високу ефективність та точність розпізнавання, але й адаптивність до різних контекстів і завдань, що робить його перспективним для подальших досліджень та практичного застосування в системах обробки природної мови.

2 АЛГОРИТМІЧНЕ ТА ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЗА ДОПОМОГОЮ МОВЛЕННЯ

2.1 Розділення інтерпретованих та автономних репрезентацій у рамках IG-CNN для розпізнавання емоцій

Для отримання глибоких інтерпретованих репрезентацій необхідно використовувати особливості, створені людиною, для спрямування процесу навчання репрезентацій. Визначено набір ключових особливостей, створених людиною, і обмежено інтерпретовані представлення так, щоб вони відповідали цим особливостям. Це нагадує існуючі підходи, де обмеження інтерпретованості використовується для спрямування навчального процесу. Проте, щоб уникнути проблем, спричинених обмеженням інтерпретованості, додано інший набір репрезентацій, які не обмежуються інтерпретованістю та вивчаються автоматично. Завдяки цьому модель здатна самостійно вивчати корисні складні представлення, як у традиційних моделях глибокого навчання. Перший набір представлень назвемо інтерпретованими, а другий – автономними представленнями.

Однак автономні репрезентації можуть негативно впливати на оцінку внеску кожної інтерпретованої репрезентації, оскільки вивчені автономні репрезентації здатні корелювати з інтерпретованими. На жаль, ці два набори репрезентацій мають високу ймовірність кореляції між собою через обмін інформацією в проміжному шарі та однакову навчальну процедуру, орієнтовану на задачу. Основним викликом, таким чином, є створення умов, за яких автономна репрезентація доповнює інтерпретовані репрезентації, а не погіршує їх в рамках системи SER.

Для вирішення цієї проблеми застосовано групову нейронну мережу для розділення процесів навчання двох наборів представлень, щоб уникнути обміну інформацією на проміжному шарі. Зокрема, побудовано дві паралельні нейронні мережі: одну, обмежену для навчання інтерпретованих

представлень, а іншу – для навчання корисних складних представлень. Додатково, накладено штраф на подібність між двома наборами представлень, щоб забезпечити незалежність простору автономних представлень від простору, визначеного інтерпретованими представленнями. Запропоновану IG-CNN продемонстровано на рисунку 2.1.

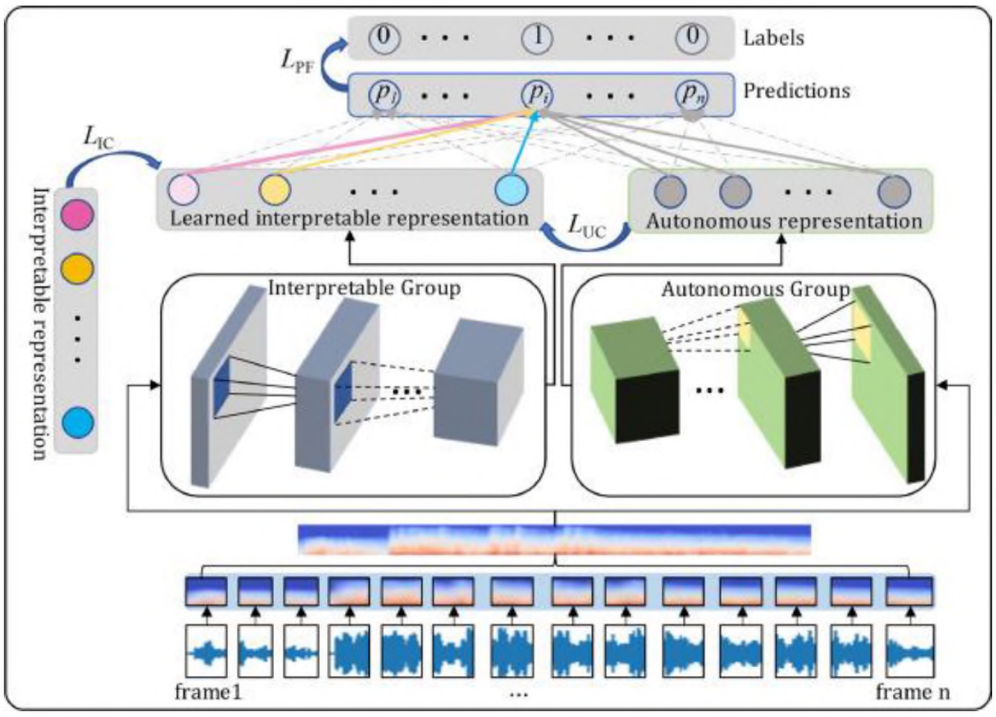


Рисунок 2.1 - Фреймворк інтерпретованої групової згорткової нейронної мережі

2.2 Обмеження інтерпретованості для навчання моделі

Як інтерпретоване представлення було використано набір GeMAPS, запропонований у [14]. Для визначення найбільш значущих ознак відповідно до їх оцінки важливості [17] (Rong та ін., 2009) застосовано алгоритм випадкового лісу. Набір GeMAPS створено з метою забезпечення порівнянності результатів аналізу емоцій у мовленні (SER) за уніфікованими критеріями. Попередні дослідження продемонстрували ефективність GeMAPS, який може бути отриманий за допомогою інструмента openSMILE (Eyben та ін., 2010). У результаті отримано 25

представлень, включаючи такі параметри, як гучність, основна частота (F0), H1-H2 тощо. таблиця 2.1 містить опис цих параметрів, а індекси від f1 до f25 впорядковані за спаданням балів важливості.

Таблиця 2.1 - Огляд параметрів GeMAPS

Індекс	**Опис**	**Індекс**	**Опис**
f1	Середнє значення висхідного нахилу гучності.	f14	Стандартне відхилення індексу Хаммаргера (HI) у дзвінкій області.
f2	Середнє значення спадаючого нахилу гучності.	f15	Середнє значення індексу Хаммаргера (HI) в неозвученій ділянці.
f3	Діапазон гучності між 20-м і 80-м процентилем.	f16	Гучність 50-го процентиля.
f4	Гучність 80-го процентиля.	f17	Середнє значення H1-H2. H1-H2 - це відношення енергії першої та другої гармонік F0.
f5	Стандартне відхилення F0.	f18	Стандартне відхилення мерехтіння.
f6	Середнє значення гучності.	f19	Середній нахил неозвученої області становить від 500 Гц до 1500 Гц.
f7	Середнє значення відношення сумарної енергії від 50-1000 Гц і 1-5 кГц.	f20	Стандартне відхилення довжини неозвученої області.
f8	Стандартне відхилення гучності.	f21	Середньоквадратичне відхилення відношення енергії спектрального піку гармоніки на центральній частоті другої форманти до енергії спектрального піку на F0.
f9	Гучність 20-го процентиля.	f22	Стандартне відхилення спадаючого нахилу гучності.
f10	Стандартне відхилення висхідного нахилу гучності.	f23	Середнє значення джиттера.
f11	Діапазон F0 між 20-м і 80-м процентилем.	f24	Стандартне відхилення H1-A3.
f12	Середнє значення нахилу дзвінкої області між 0 Гц і 500 Гц.	f25	F0 80-го процентиля.
f13	Середнє значення нахилу неозвученої області між 0 Гц і 500 Гц.		

Ці створені людиною репрезентації в таблиці 2.1 мають практичне значення і є зрозумілими для сприйняття. Наприклад, f1-f4, f6, f8, f9 і f10 - це репрезентації, пов'язані з гучністю: у випадках інтенсивних емоційних станів (як-от щастя і гнів) гучність мовлення схильна до суттєвих змін, демонструючи велику дисперсію та

нахил у бік зростання або спаду. Емоційні стани, такі як щастя та гнів, зазвичай супроводжуються вищим середнім рівнем гучності порівняно з сумом. Процес навчання репрезентації було обмежено, встановивши обмеження на евклідову відстань (тобто середньоквадратичну похибку) між вивченими репрезентаціями та визначеними репрезентаціями, створеними людиною, що дозволяє моделі вивчати інтерпретовані репрезентації. Це формально позначається :

$$L_{\mathrm{IC}} = \|\boldsymbol{r}_{\mathrm{I}} - \widehat{\boldsymbol{r}}_{\mathrm{I}}\|_2^2 = \frac{1}{K}\sum_{k=1}^{K}\left(r_{\mathrm{I}k} - \hat{r}_{\mathrm{I}k}\right)^2 , \qquad (2.1)$$

де L_{IC} - середньоквадратична похибка (MSE) втрати обмеження інтерпретованості;

$\boldsymbol{r}_{\mathrm{I}}$ - вектор ознак, що інтерпретуються;

$\widehat{\boldsymbol{r}}_{\mathrm{I}}$ - вектор навчених інтерпретованих представлень;

$\boldsymbol{r}_{\mathrm{I}k}$ - k-й елемент $\boldsymbol{r}_{\mathrm{I}}$;

$\widehat{\boldsymbol{r}_{\mathrm{I}k}}$ - k-й елемент $\widehat{\boldsymbol{r}}_{\mathrm{I}}$;

K - розмірність інтерпретованого представлення.

Рисунок 2.2 ілюструє обмеження інтерпретованості.

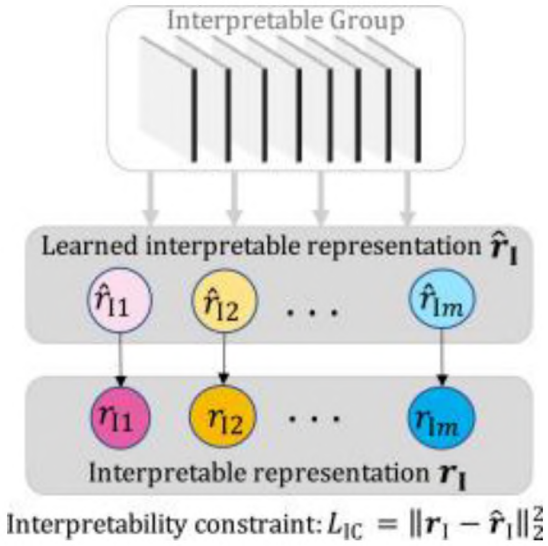


Рисунок 2.2 - Графічне представлення обмеження інтерпретованості

2.3 Групова CNN та некорельованість репрезентацій

Групову згорткову нейронну мережу (CNN) вперше запропоновано для паралельних обчислень шляхом створення окремих згорткових ядер для двох гілок CNN [16]. Подальші дослідження показали, що незалежні набори фільтрів, використані в груповій CNN, здатні знизити взаємозалежність між фільтрами, що дозволяє моделі навчатися менш корельованих представлень (Ioannou та ін., 2017) [15]. У розробленій моделі реалізовано дві групи фільтрів CNN: одну для навчання інтерпретованих репрезентацій та іншу для навчання автономних репрезентацій мовних сигналів. Обидві групи фільтрів мають однакову структуру, але різні параметри, що забезпечує одночасне та незалежне вилучення інтерпретованих та автономних представлень. Взаємодія між цими двома потоками вилучення репрезентацій не додана, оскільки метою є отримання інтерпретованих та автономних репрезентацій, які доповнюють остаточний прогноз з різних аспектів. Детальна схема групової згортки подана на рисунку 2.3.

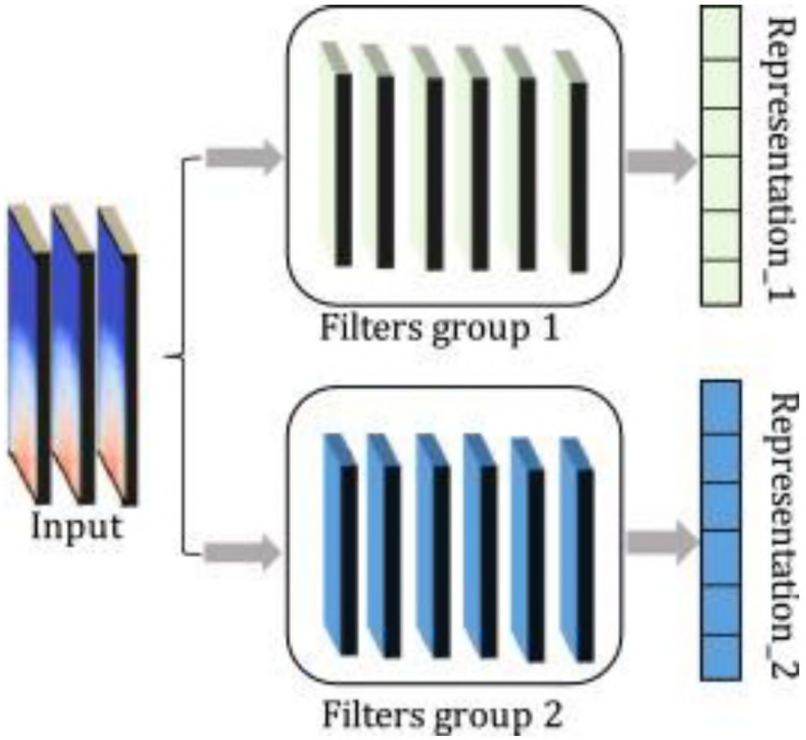


Рисунок 2.3 - Загальна структура компоненти 2-групових згорток

Щоб зробити автономні репрезентації некорельованими з інтерпретованими, було застосовано обмеження некорельованості для керування процесом автономного навчання. Зокрема, очікується, що подібність між автономними та інтерпретованими представленнями буде мінімальною. Для цього використано критерій косинусної подібності для вимірювання кореляції між двома наборами репрезентацій та її обмеження у процесі навчання. Припускаючи, що автономне представлення позначено як r_A, косинусна подібність $L_{\{UC\}}$ між r_A та вивченим інтерпретованим представленням r_I визначається формулою:

$$L_{\mathbf{UC}} \triangleq |\mathbf{Cos}\,(r_{\mathbf{A}}, \hat{r}_{\mathbf{I}})| = \left|\frac{r_{\mathbf{A}} \cdot \hat{r}_{\mathbf{I}}}{r_{\mathbf{A2}} \times \hat{r}_{\mathbf{I2}}}\right| = \left|\frac{\sum_{k=1}^{K} r_{\mathbf{A}k} \times \hat{r}_{\mathbf{I}k}}{\sqrt{\sum_{k=1}^{K} r_{\mathbf{A}k}^2} \times \sqrt{\sum_{k=1}^{K} \hat{r}_{\mathbf{I}k}^2}}\right|, \tag{2.2}$$

де $|\cdot|$ - обчислення абсолютного значення.

Функція загальних втрат повинна відображати ефективність розпізнавання емоцій, якість інтерпретованих репрезентацій та схожість між двома наборами репрезентацій. Отже, сумарна функція :

$$\mathscr{L} = L_{\mathrm{PF}} + \alpha L_{\mathrm{IC}} + \beta L_{\mathrm{UC}}, \tag{2.3}$$

де α - вага для обмеження, що інтерпретується;

β - вага обмеження некореляції.

LPF - це продуктивності втрата яка визначається як перехресна ентропія за формулою:

$$L_{\mathrm{PF}} = -\sum_{j=1}^{J} y_j \log\left(\hat{y}_j\right), \tag{2.4}$$

де y_j - мовленнєва емоція базової істини;

$\hat{y}_J$ - ймовірність j-ї категорії емоції, передбаченої моделлю;

J - номер категорії емоцій.

Підсумовуючи, функція загальних втрат записується:

$$\mathscr{L} = -\sum_{j=1}^{J} y_j \log\left(\hat{y}_j\right) + \alpha \frac{1}{K} \sum_{k=1}^{K} \left(r_{\mathrm{I}k} - \hat{r}_{\mathrm{I}k}\right)^2 + \beta \left| \frac{\sum_{k=1}^{K} r_{\mathrm{A}k} \times \hat{r}_{\mathrm{I}k}}{\sqrt{\sum_{k=1}^{K} r_{\mathrm{A}k}^2} \times \sqrt{\sum_{k=1}^{K} \hat{r}_{\mathrm{I}k}^2}} \right| , \quad (2.5)$$

2.4 Метод розпізнавання емоцій на основі мовлення

Для навчання запропонованої моделі використано загальний рекурентний підхід. Після обчислення градієнтів за допомогою поширених інструментів глибокого навчання (наприклад, TensorFlow) застосовується оптимізатор (такий як Adam або RMSprop) для оновлення параметрів моделі, що сприяє покращенню її здатності представляти та розпізнавати емоції. Такий підхід до навчання моделей є стандартною практикою в науковій літературі з цієї тематики. Псевдокод процесу навчання моделі представлено:

Вхідні дані: Навчальні дані $\{(x_{sj}, r_{Ij}, y_j)\}_{j=1}^{J}$, де x_s - спектрограма мови, $y_j \in \{0, 1, 2, 3\}$
Вихідні дані: Параметри θ

1. Випадково ініціалізуйте θ за допомогою методу Xavier.
2. Поки **поточна_позиція** знаходиться всередині кола, виконуйте:
    1. Для кожної партії в розмірі партії:
        1. Обчисліть прогноз $\hat{y}_j = \mathrm{IG\text{-}CNN}(x_{sj}, r_{Ij})$ для міні-партії.
        2. Обчисліть значення втрат L згідно з Рівнянням (5).
        3. Оновіть параметри: $\theta \leftarrow \theta - \eta \frac{\partial L}{\partial \theta}$.
    2. Кінець для.
3. Кінець поки.

Алгоритм (рисунок 2.4) розпізнавання емоцій за допомогою мови складається з декількох етапів, які можна описати у вигляді наступного покрокового алгоритму:

Крок 1: Імпорт бібліотек. Імпортуйте необхідні бібліотеки для обробки аудіо та моделювання.

Крок 2: Завантаження та підготовка даних.

— Завантажте аудіодані з різних джерел.

— Створіть DataFrame для зберігання шляхів до файлів і відповідних емоцій.

Крок 3: Візуалізація даних

— Відобразіть кількість записів для кожної емоції.

— Візуалізуйте аудіофайли за допомогою waveplot та спектограм.

Крок 4: Аугментація даних. Застосування різних технік аугментації до аудіоданих, такі як додавання шуму, зміна швидкості, зсув часу і зміна висоти звуку.

Крок 5: Екстракція ознак. Витягування корисних ознак з аудіоданих, включаючи Mel-frequency cepstral coefficients (MFCC), Chroma_stft, zero crossing rate, та інші.

Крок 6: Підготовка даних для моделювання

— Нормалізація ознак.

— Кодування міток за допомогою one-hot encoding.

— Розділення даних на навчальні та тестові набори.

Крок 7: Побудова моделі

— Створіть модель глибокого навчання, використовуючи шари, такі як Conv1D, MaxPooling1D, Dense.

— Навчіть модель на навчальних даних.

Крок 8: Оцінка моделі

— Оцініть точність моделі на тестових даних.

— Виведіть confusion matrix та classification report для оцінки результатів.

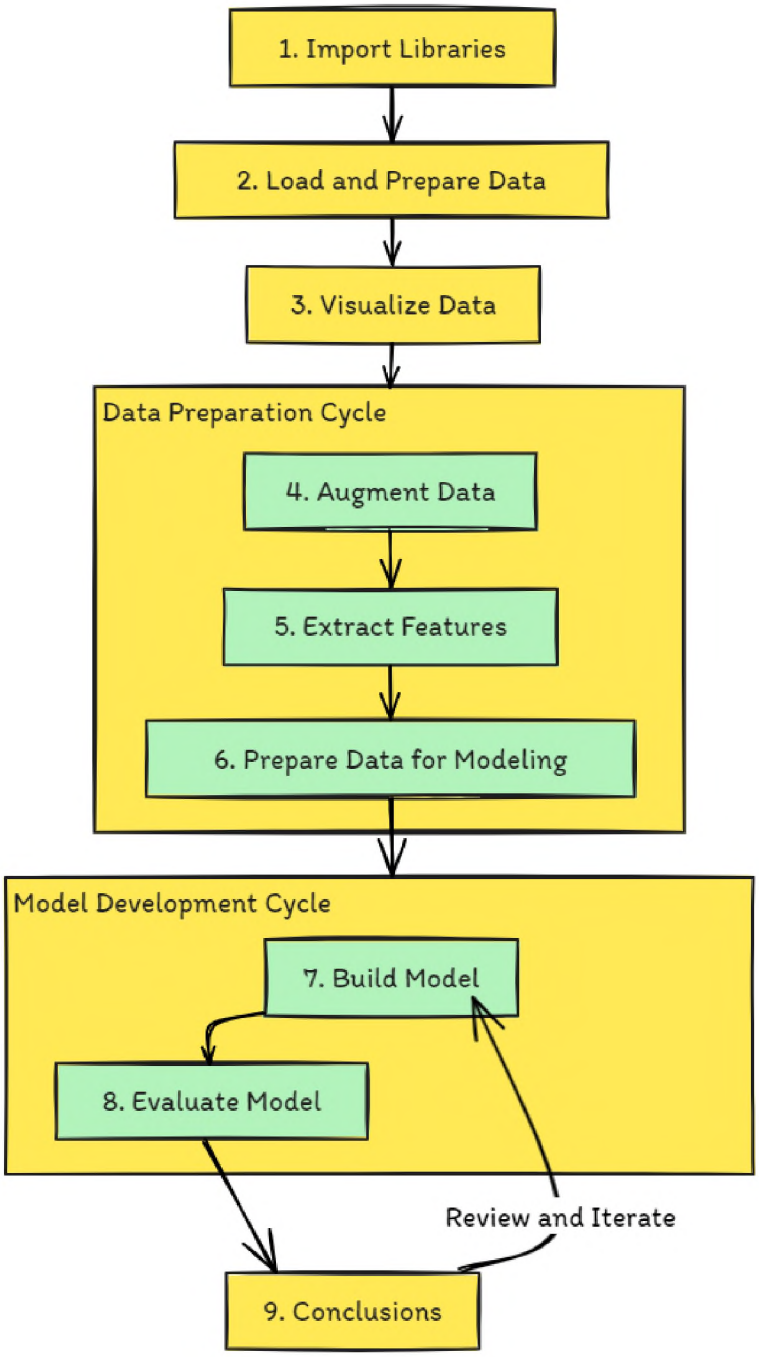


Рисунок 2.4 – Структура методу розпізнавання емоцій за допомогою мовлення

Крок 9: Висновки. Аналізуйте результати, визначте слабкі місця моделі та спробуйте покращити результати, модифікуючи модель або попередню обробку даних.

Цей алгоритм може слугувати основою для розробки систем розпізнавання емоцій на основі мовленнєвих даних, проте в кожному

конкретному випадку можуть знадобитися специфічні модифікації або додаткові етапи.

Висновки до розділу 2

1. Розглянуто алгоритмічне та інформаційне забезпечення розпізнавання емоцій за допомогою мовлення. Запропоновано підхід, що поєднує інтерпретовані та автономні репрезентації у рамках IG-CNN, для ефективного аналізу мовних сигналів. Це дозволяє моделі не тільки інтерпретувати репрезентації, створені людиною, але й автономно вивчати більш складні представлення, які доповнюють інтерпретовані ознаки. Такий підхід забезпечує багаторівневий аналіз емоційного контексту мовлення, зменшуючи ризик погіршення точності через кореляцію між різними наборами репрезентацій. Було застосовано обмеження на некорельованість та групову згортку для забезпечення незалежності між інтерпретованими та автономними представленнями, що сприяє підвищенню загальної ефективності системи розпізнавання емоцій.

2. Описано метод обмеження інтерпретованості для навчання моделі, використовуючи набір ознак GeMAPS. Цей набір забезпечує стандартизований підхід до визначення важливих акустичних маркерів для емоційного розпізнавання. Застосування алгоритму випадкового лісу дозволяє виділити найбільш значущі ознаки та встановити обмеження на їх інтерпретованість. У поєднанні з груповою CNN це допомагає покращити якість навчання моделі та її здатність точно класифікувати емоційні стани за мовними сигналами. Обмеження інтерпретованості в цьому контексті зменшує ризик надмірного узагальнення та підвищує здатність моделі до адекватної інтерпретації емоцій.

3. Окрім цього, було запропоновано комплексний алгоритм розпізнавання емоцій на основі мовлення, який включає покроковий процес від імпорту

бібліотек та завантаження даних до побудови моделі глибинного навчання та її оцінки. Цей алгоритм дозволяє систематизувати підхід до обробки аудіоданих, а також ефективно екстрагувати ознаки та використовувати техніки аугментації для підвищення стійкості моделі. Комбінація методів обробки сигналів, екстракції ознак та глибинного навчання дає можливість створити надійну систему розпізнавання емоцій, яка може бути адаптована під конкретні завдання та умови застосування.

3 ПРОГРАМНО-ТЕХНОЛОГІЧНЕ ЗАБЕЗПЕЧЕННЯ

3.1 Набори даних та умови експерименту

Для тестування методу використано чотири набори даних: IEMOCAP, RAVDESS, eNTERFACE'05 та CREMA-D. Далі подано детальний опис кожного з них.

IEMOCAP: Інтерактивна база даних захоплення емоційних діадних рухів (IEMOCAP) була створена Лабораторією аналізу та інтерпретації мовлення Університету Південної Каліфорнії (Busso та ін., 2008) [20]. Десять акторів були запрошені виконати певні емоційні сценарії та імпровізувати діалоги в умовних ситуаціях для вираження певних емоційних станів, таких як щастя та гнів. Набір даних IEMOCAP охоплює близько дванадцяти годин записаних діалогів, під час яких також реєструвалися рухи обличчя, рук і голови. Ці діалоги були вручну сегментовані на окремі репліки за ходом розмови. Емоційний стан кожного сегмента оцінювався трьома незалежними оцінювачами з дискретного та безперервного аспектів. Для завдання SER використовуються дискретні емоції, і обрано аудіосегменти, що містять емоції гніву, щастя, смутку та нейтрального стану. Відповідно до попередніх досліджень, збуджену емоцію розглядаємо як щастя (Mirsamadi та ін., 2017) [31].

RAVDESS: Аудіовізуальна база даних емоційного мовлення та пісень Райерсона (RAVDESS) - це перевірена та гендерно збалансована мультимодальна база даних мовлення та пісень (Livingstone & Russo, 2018) [28]. Двадцять чотири професійні актори були запрошені висловити цільові емоції за допомогою двох нейтральних тверджень "Діти розмовляють біля дверей" та "Собаки сидять біля дверей". Мовна частина містить сім емоцій: спокій, щастя, смуток, гнів, страх, здивування, огиду. Кожен запис бази даних був оцінений 10 разів 247 особами, щоб перевірити достовірність

маркування даних. А сімдесят два оцінювачі взяли участь у перевірці надійності за методом "тест-повторне тестування".

eNTERFACE'05: eNTERFACE'05 - це база даних аудіовізуальних емоцій (Martin et al., 2006) [30] включає 42 суб'єкти, кожного з яких попросили прослухати шість послідовних коротких історій. Кожна історія була розроблена таким чином, щоб викликати одну певну емоцію (гнів, огиду, страх, щастя, смуток і здивування). Двоє експертів-людей оцінювали, чи реагували суб'єкти на ситуації з однозначною емоцією.

CREMA-D: Набір даних емоційних мультимодальних акторів з натовпу (CREMA-D) - це аудіо-візуальна база даних, що містить дані про емоційний вираз обличчя та голосу (Cao та ін., 2014) [21]. Дев'яносто одному акторові та актрисі було запропоновано вимовити задане речення під однією цільовою емоцією (щастя, смуток, гнів, страх, відраза та нейтральна), викликаною описом розроблених сценаріїв. Кожен записаний ролик оцінювався щонайменше шість разів 2443 експертами.

Через різницю в довжині аудіороликів для забезпечення однакової тривалості необхідно виконати кілька кроків попередньої обробки. Відповідно до попередніх досліджень, таких як Сатт та ін. (2017) та Tariq та ін. (2019) [33], було виконано наступні три етапи обробки. Спершу видалено аудіозаписи надто короткої тривалості (менше 1 секунди). Далі, аудіокліпи розділено на сегменти однакової тривалості — 3 секунди, з використанням нульового заповнення для сегментів меншої тривалості (Satt et al., 2017; Tariq et al., 2019) [33]. На третьому етапі отримано спектрограму мовлення, застосувавши короткочасне Фур'є-перетворення (STFT) із довжиною кадру 25 мс та зсувом кадру 10 мс до короткого мовного сигналу. Зведену статистику експериментальних наборів даних наведено в Таблиці 3.1.

Кожен набір даних випадковим чином розбивається на навчальний, валідаційний та тестовий набори у співвідношенні 7:1:2, як запропоновано в Jiang et al. (2020) [26] та Cho та ін. (2021) [23]. Щоб виконати справедливих

порівнянь та достовірних оцінок, усі моделі навчалися, перевірялися та тестувалися на одних і тих самих даних.

Таблиця 3.1. Зведена статистика IEMOCAP, RAVDESS, eNTERFACE'05 та CREMA-D

Дані	IEMOCAP	RAVDESS	eNTERFACE'05	CREMA-D
Гнів	2019	192	210	1271
Щастя	3012	192	210	1271
Смуток	2387	192	210	1270
Нейтральний	2635	-	-	1087
Страх	-	192	210	1271
Огида	-	192	210	1271
Сюрприз	-	192	210	-
Спокійно	-	192	-	-
Всього	10053	1344	1257	7441

Експерименти проводилися з використанням TensorFlow v1.14.0 на робочій станції з процесором Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 ГГц і однією графічною картою NVIDIA Titan X. Оптимізатор Adam Оптимізатор з початковою навчання швидкістю $\eta = 0.05$. Коефіцієнти обмеження інтерпретованості та некорельованості задано на рівні $\alpha = 0.2$ та $\beta = 0.7$ для IEMOCAP та RAVDESS, $\alpha=0.2$ та $\beta=0.8$ для eNTERFACE'05, $\alpha = 0.5$ та $\beta = 0.8$ для CREMA-D, відповідно. Коефіцієнти відсіву встановлені на рівні 0,25, 0,25, 0,2 та 0,25 для IEMOCAP, RAVDESS, eNTERFACE'05 та CRENA-D відповідно. А розміри партій встановлено на 256, 256, 64 та 256 для IEMOCAP, RAVDESS, eNTERFACE'05 та CREMA-D відповідно. Далі проаналізуємо всі ці гіперпараметри в частині аналізу чутливості.

3.2 Загальна ефективність

Далі проведено результати застосування запропонованої моделі IG-CNN на тестовому наборі порівняно з десятьма базовими лініями. Запускаємо кожну модель 10 разів і порівнюємо середнє значення метрик з t-тестом для перевірки статистичної значущості. Експериментальні результати представлені в таблиці А.1 (див. додаток А) з якої видно, що всі результати пройшли значущість.

Згідно з результатами, наведеними в таблиці А.1 , ми бачимо, що IG-CNN досягла найкращої продуктивності і значно покращила показники точності, макроточності, макропригадування, макропоказника F1, мікропоказника F1 та зваженого показника F1. Продуктивність більшості моделей DL значно вища, ніж у методів, що не використовують DL. Ці результати вказують на те, що методи ШНМ можуть навчатися більш дискримінативним репрезентаціям на основі необроблених даних, ніж репрезентації, отримані вручну.

3.3 Дослідження впливу компонентів моделі

Було проведено дослідження впливу компонентів моделі (тобто обмеження інтерпретованості, групової згортки та обмеження некорельованості) на загальну продуктивність. Зокрема, з запропонованої моделі окремо вилучено обмеження інтерпретованості, групову згортку та обмеження некорельованості, позначивши три варіанти як IG-CNN_1, IG-CNN_2 та IG-CNN_3 відповідно. Продуктивність кожного варіанту було протестовано за показниками точності, макроточності, макропригадування, макрооцінки F1, мікрооцінки F1 та зваженої оцінки F1. Результати експериментів з абляцією наведені в таблиці Б.1 (див. додаток Б).

З таблиці Б.1 видно, що IG-CNN значно перевершує свої варіанти за всіма метриками, що підтверджує важливість усіх трьох компонентів моделі для

досягнення високої продуктивності. Середньоквадратична похибка (MSE) IG-CNN становить 1.959, 2.876, 3.900 та 1.660 на чотирьох наборах даних, тоді як для варіанту без обмеження інтерпретованості MSE дорівнює 76.237, 73.513, 77.922 та 107.630. Це свідчить про те, що обмеження інтерпретованості сприяє ефективному навчанню моделі інтерпретованим представленням.

Щоб продемонструвати ефект навчання репрезентацій, побудовано функції щільності ядра для порівняння між реальними та навченими інтерпретованими репрезентаціями на прикладі набору даних IEMOCAP. Через обмеженість простору на рисунку 3.2 представлено часткові репрезентації з таблиці 2.1.

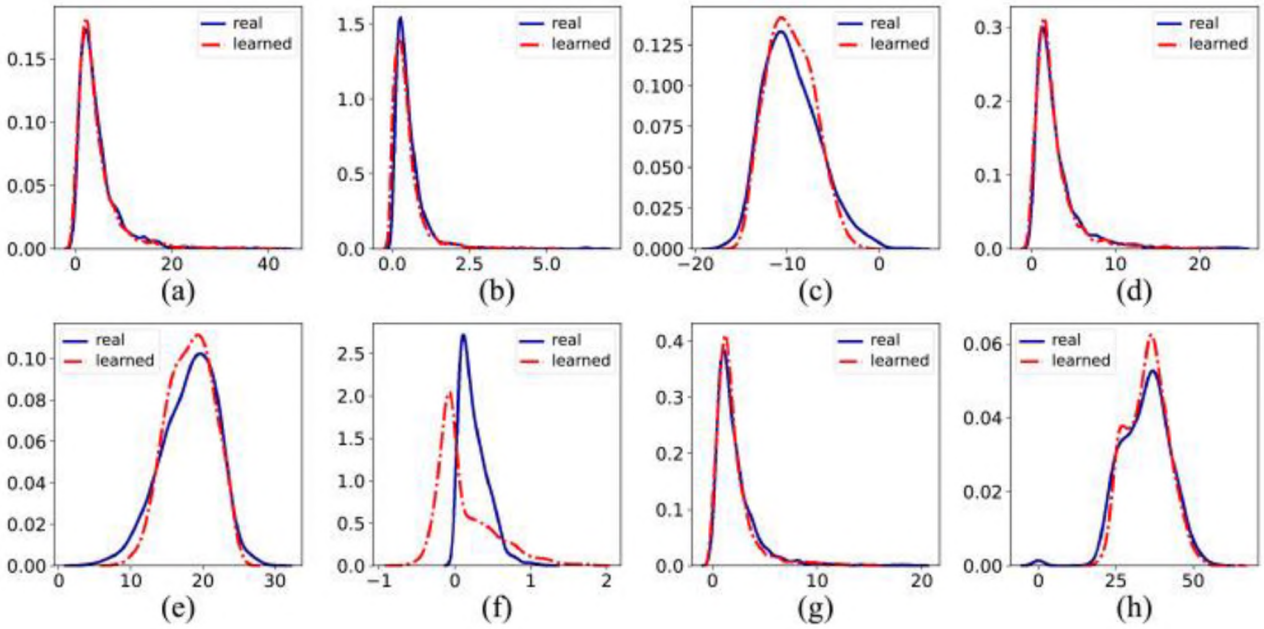


Рисунок 3.2 - Функції щільності ядра реальних та навчених інтерпретованих представлень. (Індекси (a) - (h) відповідають f1, f4, f7, f10, f15, f20, f22 та f25 в таблиці 2.1)

На рисунку 3.2 синя суцільна лінія позначає реальне представлення, отримане за допомогою інструмента openSMILE, а червона пунктирна лінія — навчене представлення за допомогою IG-CNN. Загалом, рисунок 3.2 демонструє узгодженість між цими навченими інтерпретованими представленнями та реальними. Завдяки доданню обмеження на некореляцію, IG-CNN очікувано

вивчає більш корисні та повні представлення, мінімізуючи косинусну подібність між інтерпретованими та автономними представленнями.

Косинусні метрики подібності для IG-CNN становлять 0.0018, 0.0068, 0.0110 та 0.0017 на чотирьох наборах даних, у той час як для варіанту IG-CNN без обмеження на некореляцію ці значення дорівнюють 0.1054, 0.1701, 0.1287 та 0.0407. Крім того, компонент групової згортки використовується для окремого виділення інтерпретованих та автономних представлень, щоб мінімізувати кореляцію між ними. Результати косинусної подібності свідчать, що моделі з обмеженням на некореляцію та груповою згорткою здатні вивчати більш корисні та багаті представлення..

3.4 Аналіз чутливості

Проведено аналіз чутливості для оцінки продуктивності моделі при налаштуванні гіперпараметрів, таких як кількість шарів, розмір ядра, частота відсіву, швидкість навчання та розмір партії. Оскільки вагові коефіцієнти у функції втрат врівноважують три ключові компоненти моделі, додатково було досліджено результати за різних вагових значень. Аналіз чутливості та оптимізація гіперпараметрів також сприяли пом'якшенню проблем надмірного та недостатнього навчання моделі (Srivastava та ін., 2014 [32]; Baydilli & Atila, 2018 [19]).

3.4.1 Згорткові шари та розміри ядер

Запропонована модель IG-CNN є варіантом CNN. Таким чином, згортки шари та згортки ядро та розмір ядра згортки впливатимуть на продуктивність моделі. На рисунку 3.3 наведено аналіз чутливості для різних шарів конфігурацій . Загалом, тришарова згорткова структура є достатньою для

вилучення емоційної інформації на наборах даних IEMOCAP, RAVDESS та eNTERFACE'05, тоді як чотиришарова є більш ефективною для CREMA-D.

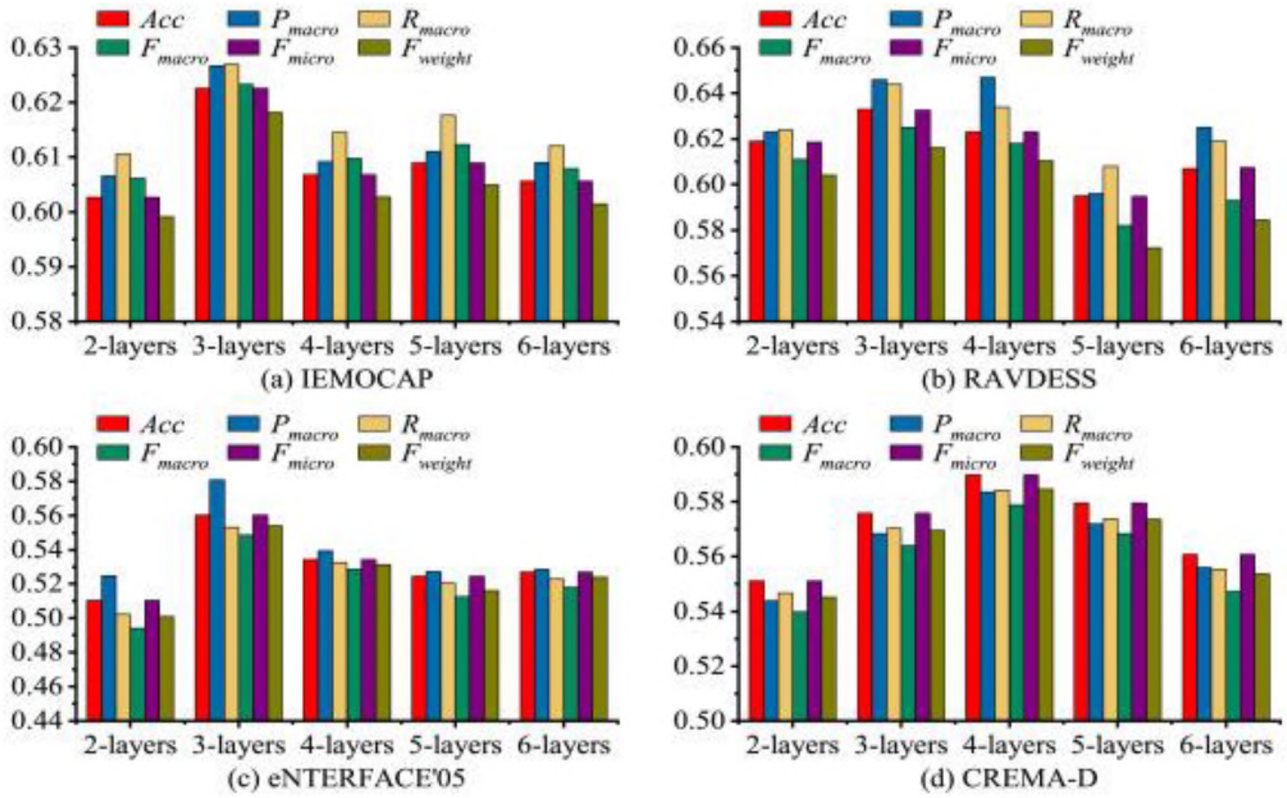


Рисунок 3.3 - Результати роботи IG-CNN з різними конфігураціями шарів згортки згортки шарів

Рисунок 3.4 показує, що ядра згортки розміром (5 × 5), (5 × 5) і (3 × 3) у кожному з цих трьох шарів досягають найкращої продуктивності на IEMOCAP і RAVDESS. Ядро великого розміру використовується для вилучення інформації з нижніх шарів, тоді як ядро малого розміру використовується для агрегування інформації у верхньому шарі. Згорткове ядро розміром (7×7), (7×7) та (7×7) є більш ефективним на eNTERFACE'05 та CREMA-D.

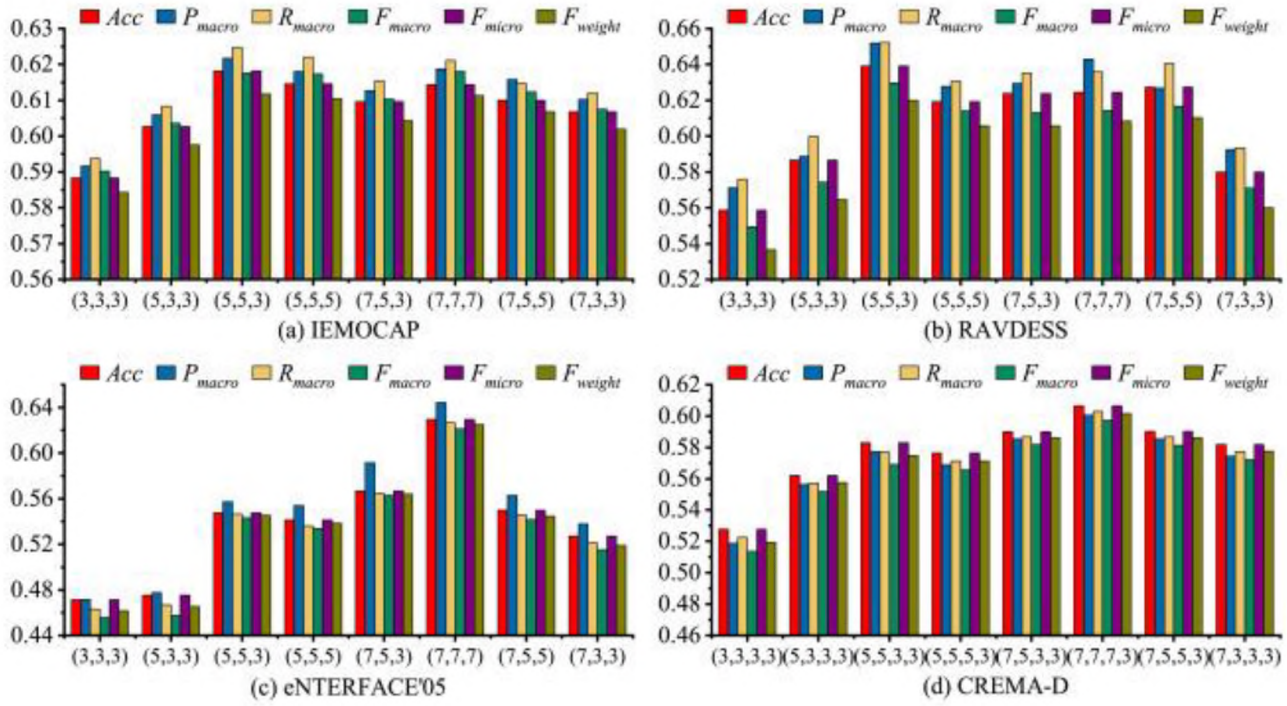
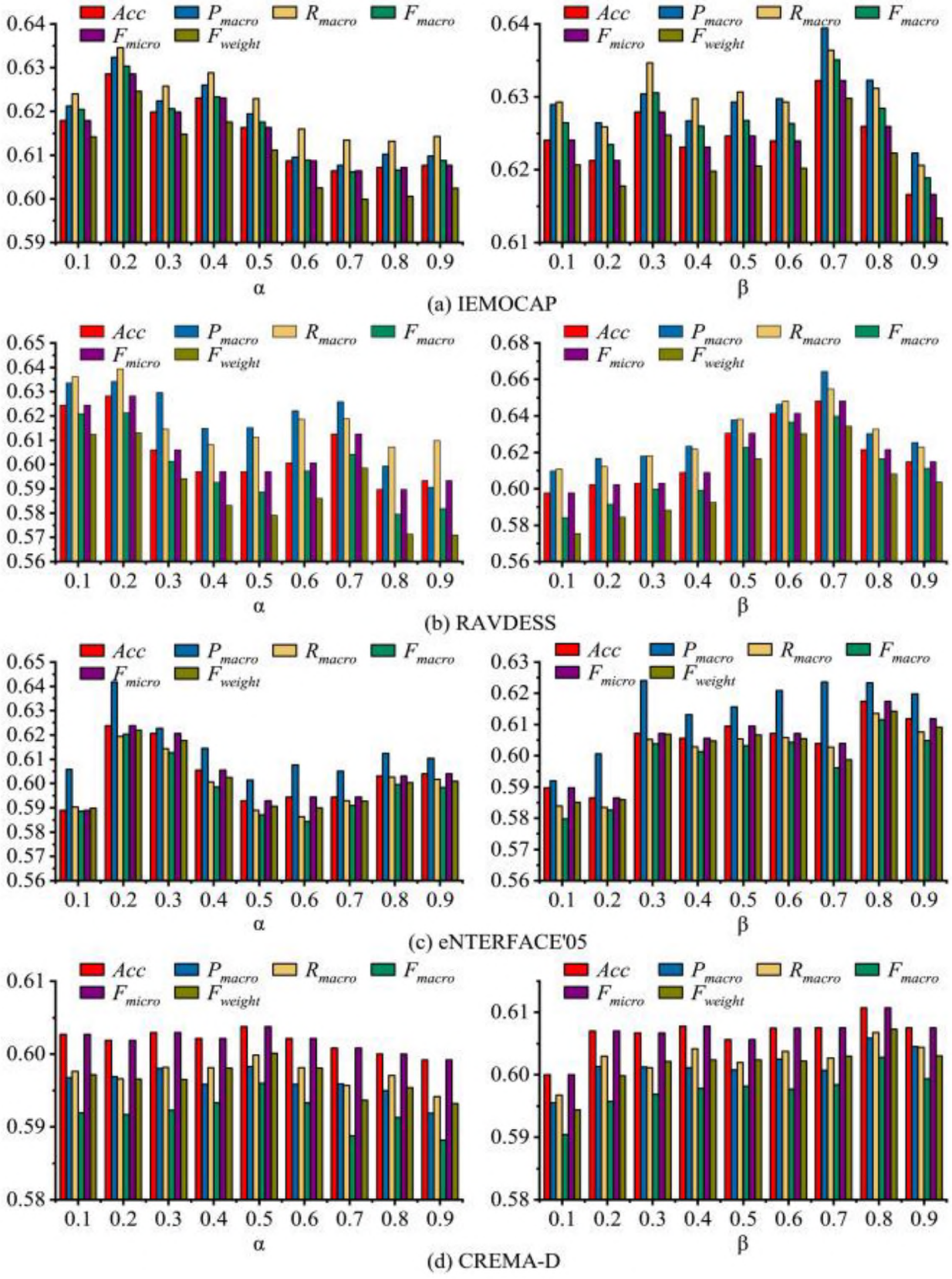


Рисунок 3.4 - Продуктивність IG-CNN з різними конфігураціями згортки ядро розмірів

3.4.2 Інтенсивність інтерпретованості та обмеження некорельованості

Коефіцієнт обмеження інтерпретованості α та коефіцієнт обмеження некореляції β впливають на продуктивність моделі. На рисунку 3.5 показано результати роботи запропонованої моделі при різних значеннях α та β. Видно, що IG-CNN досягає найкращих результатів при парах α-β (α = 0,2, β = 0,7), (α = 0,2, β = 0,7), (α = 0,2, β = 0,8) та (α = 0,5, β = 0,8) на IEMOCAP, RAVDESS, eNTERFACE'05 та CREMA-D відповідно. Здатність IG-CNN виокремлювати інтерпретовані ознаки контролюється α, і α = 0.2, 0.2, 0.2 і 0.5 показує, що IG-CNN може добре запам'ятовувати інтерпретовані ознаки. Ці автономні представлення повинні якомога більше відрізнятися від представлень, які можна інтерпретувати (β = 0.7, 0.7, 0.8 і 0.8), щоб забезпечити отримання більш повної та корисної інформації, яка може бути використана для підтримки рішень прийняття . Це також вказує на те, що

автономні репрезентації також важливі для розпізнавання мовленнєвих емоцій.



Рисунку 3.5 - Результати роботи IG-CNN з різними α та β

3.4.3 Швидкість навчання та розмір партії

Швидкість навчання та розмір партії — це два гіперпараметри, які впливають на процес навчання моделі ШНМ. Зазвичай між швидкістю навчання та розміром партії існує компроміс. Наприклад, правило лінійного масштабування (Goyal et al., 2017) [25] вказує, що при множенні розміру міні-партії на коефіцієнт k, швидкість навчання також має бути помножена на k. Дослідження Він та ін. (2019) [33] виявило статистично негативну кореляцію між здатністю моделі до узагальнення та співвідношенням розміру партії до швидкості навчання. Тому було проведено додатковий аналіз впливу швидкості навчання та розміру партії на продуктивність IG-CNN. На рисунку B.1 (див. додаток B) представлено результати роботи IG-CNN залежно від різних комбінацій швидкості навчання та розміру партії.

Рисунок B.1 показує, що продуктивність IG-CNN спочатку зростає, а потім зменшується при однаковій швидкості навчання зі збільшенням розміру партії та при однаковому розмірі партії зі зменшенням швидкості навчання, відповідно. Нарешті, для всіх наборів даних було обрано швидкість навчання 0.05. Для IEMOCAP, eNTERFACE'05 та CREMA-D обрано розмір партії 256, а для RAVDESS - 64.

3.4.4 Рівень відсіву

Під час навчання DL-моделі відсів використовується як прийом для зменшення перенавчання шляхом випадкового відсіву одиниць з нейронної мережі нейронної мережі (Srivastava та ін., 2014) [32]. Оскільки швидкість відсіву є параметром, що налаштовується, ми провели експеримент з аналізу чутливості, щоб показати, як швидкість відсіву впливає на продуктивність моделі. Результати представлені на рисунку 3.6. Загалом, результати показують, що найкращі результати досягаються при значенні коефіцієнта

відсіву 0,25 для моделей IEMOCAP, RAVDESS та CREMA-D, тоді як для eNTERFACE'05 він дорівнює 0,2.

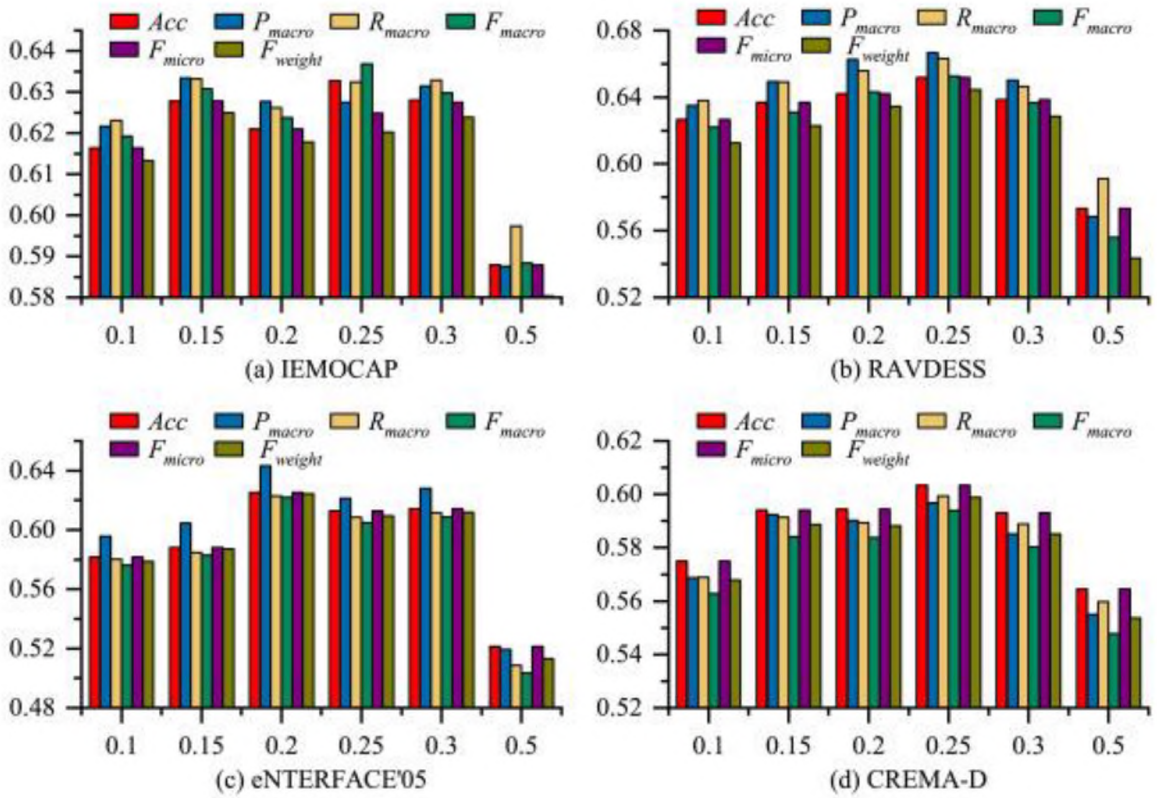


Рисунок 3.6 - Продуктивність IG-CNN з різним відсотком відсіву

3.5 Тестування роботи моделі в різних ситуаціях

Проаналізовано вплив віку, статі та шумових факторів на продуктивність IG-CNN. Слід зазначити, що eNTERFACE'05 не надає конкретної інформації про стать та вік досліджуваних. IEMOCAP і RAVDESS мають лише інформацію про стать, тоді як CREMA-D надає детальну інформацію про вік на додаток до статі. Таким чином, ми провели експериментальний аналіз на CREMA-D. Крім того, ми також включили RAVDESS для гендерного аналізу, оскільки його розмір відносно менший, ніж у IEMOCAP і CREMA-D, але схожий за розміром на eNTERFACE'05.

а) Моделювання гендерних відмінностей

Проведено t-тест на наборах даних RAVDESS і CREMA-D для розпізнавання гендерних емоцій у мовленні. У тестовому наборі RAVDESS міститься 142 сегменти чоловічого мовлення та 127 сегментів жіночого мовлення. CREMA-D містить 766 сегментів чоловічого мовлення та 723 сегменти жіночого мовлення. У таблиці 3.1 міститься результати аналізу. Результати показують, що наша модель краще працює на жіночих сегментах мовлення. Це логічно, оскільки вираження емоцій у жіночому мовленні є більш очевидним, що може полегшити розпізнавання моделі.

Таблиця 3.1 - Т-тест гендерних відмінностей у показниках RAVDESS та CREMA-D

Дані	**Стать**	Ак	Pmacro	Rmacro	Fmacro	Fmicro	Fweight
RAVDESS	Чоловік	0.609	0.618	0.603	0.586	0.609	0.600
	Жінка	0.688	0.691	0.684	0.672	0.688	0.687
	p-значення	0.0024	0.0017	0.0022	0.0009	0.0024	0.0009
CREMA-D	Чоловік	0.561	0.572	0.569	0.556	0.561	0.551
	Жінка	0.594	0.600	0.597	0.595	0.594	0.593
	p-значення	0.0075	0.0299	0.0119	0.0045	0.0075	0.0032

b) Загальна ефективність на різних вікових групах

Оскільки база даних CREMA-D містить інформацію про вік, перевірено чутливість запропонованої моделі до різних вікових груп на CREMA-D. 91 актор набору даних CREMA-D був віком від 20 до 74 років із середнім віком 36 років. Розділено їх на 9 груп з п'ятирічним інтервалом. Оскільки актори-сенсори досить рідкісні, остання група складається з акторів віком 61, 62, 66 і 74 роки. Кількість мовленнєвих аудіозаписів у групах з першої по дев'яту становить 455, 153, 220, 141, 135, 91, 82, 114 і 98 відповідно. Таблиці 3.2 та Г.1 (див. додаток Г) представлено результати кожної групи за показниками точності, макроточності, макропригадування, макрооцінки F1, мікрооцінки F1 та зваженої оцінки F1, а також p-значення Т-тесту між різними групами.

Результати показують, що наша модель краще працює у віковому діапазоні від 51 року до 60 років. Попередні дослідження показали, що люди похилого віку більш чутливі до емоцій і більш схильні до вираження сильних емоцій (Levenson et al., 1991 [27]; Makita et al., 2021 [29]), що підтверджує наші результати. Причому у віковому діапазоні від 20 до 45 років не спостерігається значної різниці в показниках.

Таблиця 3.2 - Загальна ефективність на різних вікових групах CREMA-D

Групи	Ак	Pmacro	Rmacro	Fmacro	Fmicro	Fweight
20-25	0.584	0.587	0.589	0.580	0.584	0.581
26-30	0.575	0.596	0.605	0.576	0.575	0.565
31-35	0.580	0.582	0.587	0.579	0.580	0.578
36-40	0.560	0.606	0.572	0.563	0.560	0.553
41-45	0.575	0.618	0.588	0.576	0.575	0.569
46-50	0.525	0.530	0.507	0.506	0.525	0.519
51-55	0.632	0.638	0.632	0.622	0.632	0.625
56-60	0.642	0.674	0.631	0.638	0.642	0.640
60-	0.496	0.512	0.513	0.493	0.496	0.488

c) Робота моделі в різних середовищах

Також проаналізовано здатність IG-CNN розпізнавати мовленнєві емоції в шумному середовищі на наборах даних RAVDESS та CREMA-D. Було додано три типи шуму, а саме адитивний білий гаусівський шумшум салону автомобіля та шум балаканини, до сегментів чистого мовлення (Chenchah & Lachiri, 2016 [22]; Bandela & Kumar, 2021 [18]). Співвідношення шуму сигналу signal сигнал/шум співвідношення (SNR) були на різних рівнях, включаючи 30 дБ, 35 дБ, 40 дБ, 45 дБ і 50 дБ. Таблиці Д.1 та Д.2 (див додаток Д) наведено результати експериментів у шумному середовищі

порівняно з чистими мовними сегментами на RAVDESS і CREMA-D відповідно.

Результати показують, що розроблена модель є стійкою до білого гаусівського шуму та шуму в салоні автомобіля, коли SNR перевищує 40 дБ. Результати також вказують на те, що модель є стійкою на наборах даних RAVDESS та CREMA-D до балакучого шуму, коли SNR перевищує 50 дБ та 35 дБ відповідно. Згідно з Bandela and Kumar (2021) [18] у звичайних реальних ситуаціях, таких як керування автомобілем та скупчення людей, найпоширенішими типами шумів є шум автомобіля, балаканина та інші. Відношення сигнал/шум мовних сигналів у таких шумових контекстах зазвичай перевищує 40 дБ за умови використання мікрофонів, оснащених сучасними технологіями шумозаглушення (Chen et al., 2021) [22]. Це означає, що наша модель має вищу працездатність і надійність у таких реальних ситуаціях.

3.6 Аналіз інтерпретованості

Інтерпретованість IG-CNN на основі інтерпретованих репрезентацій (див. таблицю 2.1) та вивчених коефіцієнтів прийняття рішень. Візьмемо для прикладу IEMOCAP, на рисунку 3.7 наведено теплову карту-схему вивчених коефіцієнтів прийняття рішень з різних моделей.

На рисунку 3.7 (a) та (б) представлені коефіцієнти логістичної регресії та одношарової повністю зв'язаної мережі з функцією softmax на основі інтерпретованих репрезентацій. У порівнянні з рисунком 3.7 (c)-(f), більшість репрезентацій на рисунку 3.7 (a) та (b) мають незначний вплив на прогнозування. Автономні репрезентації можуть доповнювати інтерпретовані репрезентації, покращуючи розпізнавання емоцій у мовленні (SER). З рисунка 3.7 (c), що представляє модель без обмеження інтерпретованості, видно, що хоча численні репрезентації є корисними для

прогнозування, їм бракує семантичного змісту (тобто модель без обмеження інтерпретованості є неінтерпретованою). Рисунок 3.7 (d) та (e) показує коефіцієнти моделей без обмеження некорельованості та без групової згортки відповідно, тоді як рисунок 3.7 (f) демонструє модель, яка включає обидва ці компоненти. На теплових картах видно, що лише декілька репрезентацій мають значний вплив на прогнозування, як показано на Рисунку 3.7 (d) та (f). Це свідчить про те, що обмеження некорельованості сприяє вилученню більш корисних та численних репрезентацій, які доповнюють інтерпретовані та автономні репрезентації. Аналогічно, групова згортка також є ефективною для зменшення кореляції між інтерпретованими та автономними репрезентаціями, що підтверджується порівнянням Рисунка 3.7 (e) та (f).

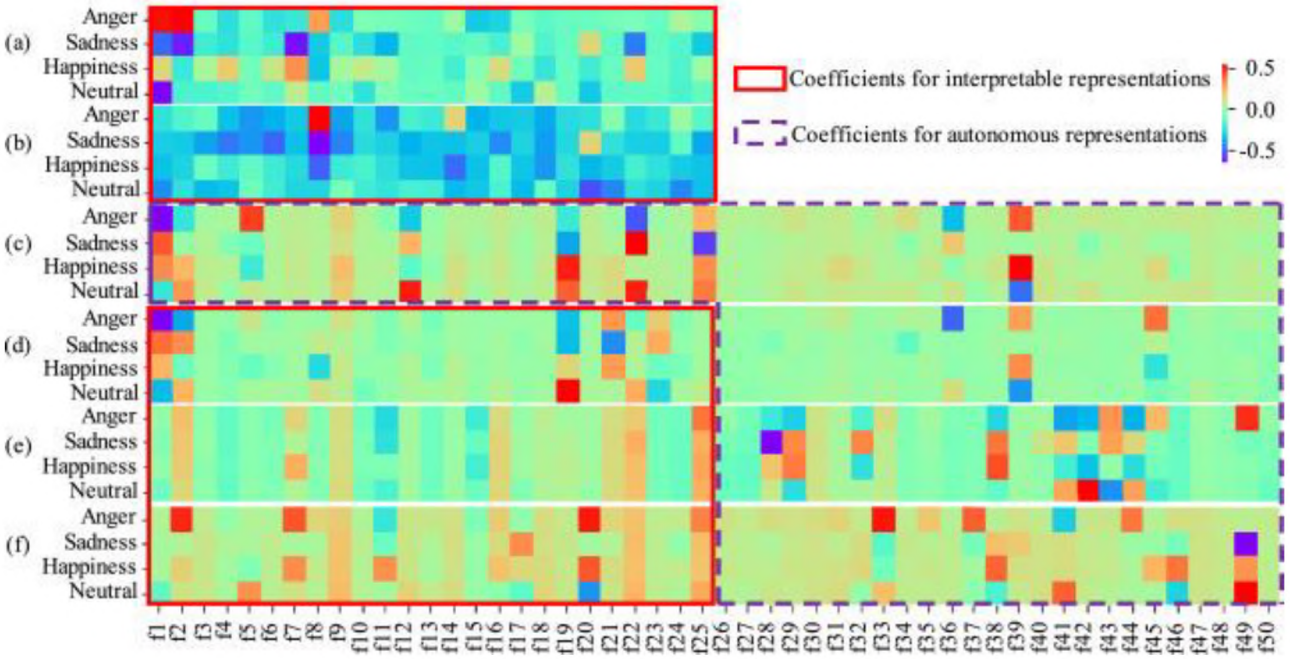


Рисунок 3.7 - Коефіцієнти прийняття рішень різних моделей. (а) логістична регресія(б) простий softmax, (в) без обмеження інтерпретованості, (г) без обмеження некорельованості, (д) без групової згортки та (е) IG-CNN

Оскільки запропонована модель може виокремлювати інтерпретовані репрезентації, інтерпретованість моделі може бути задана коефіцієнтами цих

репрезентацій (див. таблицю 2.1) з точки зору сприйняття. Репрезентації f1, f2, f3, f4, f6, f8, f9, f10, f16 і f22 відносяться до гучності. Інтуїтивно, коли люди говорять з емоціями гніву або щастя, вони кричать або посміхаються з високою і мінливою гучністю. У рисунку 3.7 (f), f2 і f8 є позитивно значущими для гніву і щастя порівняно з сумом і нейтральним станом. f9 є значно корисним для смутку і нейтрального стану, оскільки f9 є 20-м процентилем гучності, що вказує на низький рівень гучності. Навпаки, f16 є значущим для гніву і щастя, оскільки це 50-й процентиль гучності.

Більш високе співвідношення альфа пов'язане з високою активністю (Waaramaa et al., 2010) [34], як показано на рисунку 3.7 (f), f7 є значущим для емоцій гніву та щастя з високою активністю. Чим більше значення f12, тим менший діапазон зміни висоти тону, отже, f12 є значущим для нейтрального стану. Представленість f17 пов'язана з якістю голосу, і існуючі дослідження вказують на те, що голос з придихом зазвичай має вищий показник f17 (Gobl & Chasaide, 2010) [25]. Наш результат підтверджує, що f17 є значущим для смутку. Особливістю f20 є стандартне відхилення тривалості невимовлених сегментів, і вона є значущою для гніву та щастя, оскільки ці дві емоції іноді перебувають у стані сильного збудження з переривчастим мовленням. І навпаки, f20 негативно пов'язаний з нейтральним станом при плавному мовленні.

Інтерпретованість залежить від коефіцієнтів інтерпретованих репрезентацій, а результати узгоджуються з людським сприйняттям емоцій від мовлення. Наприклад, репрезентації гучності роблять значний внесок, оскільки це найбільш безпосереднє відчуття слух. Також були інтерпретовані характеристики, пов'язані з якістю голосу, такі як дихання голосу.

Висновки до розділу 3

1. Проведено комплексний аналіз програмно-технологічного забезпечення, необхідного для розпізнавання емоцій на основі мовлення. Дослідження охоплювало чотири різні набори даних: IEMOCAP, RAVDESS, eNTERFACE'05 та CREMA-D, кожен з яких має свої особливості та вимоги до попередньої обробки. Було розроблено експериментальну платформу на основі TensorFlow, що дозволило провести оптимізацію гіперпараметрів та оцінити ефективність запропонованої моделі IG-CNN.

2. Запропонована модель IG-CNN показала найкращі результати за всіма метриками, такими як точність, макроточність, макропригадування, макропоказник F1, мікропоказник F1 та зважений показник F1. Модель значно перевершила свої варіанти, які не мали обмежень інтерпретованості, некорельованості або групової згортки, що підтверджує важливість цих компонентів у підвищенні загальної продуктивності. Результати експериментів також продемонстрували здатність методів глибинного навчання навчатися більш дискримінативним репрезентаціям на основі необроблених даних.

3. Проведений аналіз чутливості підтвердив, що такі гіперпараметри, як кількість шарів, розмір ядра, частота відсіву, швидкість навчання та розмір партії, значно впливають на продуктивність моделі. Було виявлено, що певні комбінації параметрів, наприклад, пари коефіцієнтів обмеження інтерпретованості та некореляції, забезпечують оптимальне розпізнавання емоцій. Крім того, дослідження показало, що IG-CNN є стійкою до різних типів шуму, включаючи білий гаусівський шум та шум балаканини, особливо при високому співвідношенні сигнал/шум.

4. Також було проведено тестування моделі в різних умовах, включаючи вплив віку, статі та шумових факторів на продуктивність розпізнавання. Результати показали, що модель краще розпізнає емоції у жіночих сегментах

мовлення та у вікових групах від 51 до 60 років. Окрім цього, аналіз інтерпретованості продемонстрував, що модель може виокремлювати інтерпретовані репрезентації, узгоджуючи їх з людським сприйняттям емоцій у мовленні, що робить IG-CNN не лише ефективною, але й інтерпретованою системою для розпізнавання емоцій.

ВИСНОВКИ

1. Підсумовуючи результати даного дослідження, можна стверджувати, що запропонована модель IG-CNN є ефективним та інтерпретованим підходом для розпізнавання емоцій на основі мовлення. Завдяки поєднанню інтерпретованих та автономних репрезентацій, модель здатна навчатися складним та багатошаровим представленням емоційних станів, що дозволяє досягти високих показників точності та ефективності. Використання чотирьох різних наборів даних (IEMOCAP, RAVDESS, eNTERFACE'05 та CREMA-D) дозволило провести всебічну оцінку моделі, продемонструвавши її здатність до узагальнення та адаптації до різноманітних умов і сценаріїв.

2. Одним із ключових аспектів запропонованої моделі є її здатність використовувати інтерпретовані репрезентації, отримані з набору ознак GeMAPS, які є зрозумілими для людини. Це забезпечує не лише високу точність розпізнавання емоцій, але й робить модель інтерпретованою, дозволяючи аналізувати вплив різних ознак на процес прийняття рішень. Використання обмеження на некореляцію між інтерпретованими та автономними репрезентаціями допомогло мінімізувати взаємний вплив цих двох наборів ознак, що сприяло покращенню загальної продуктивності моделі.

3. Аналіз ефективності запропонованої моделі показав, що IG-CNN перевершує інші базові моделі за всіма основними метриками, включаючи макроточність, макропригадування та макропоказник F1. Це свідчить про те, що запропоновані компоненти моделі, такі як обмеження інтерпретованості, групова згортка та обмеження некорельованості, відіграють важливу роль у підвищенні якості розпізнавання емоцій. Проведені дослідження впливу компонентів моделі підтвердили, що кожен з цих елементів суттєво впливає на здатність моделі навчатися більш корисним та багатим репрезентаціям.

4. Ретельний аналіз гіперпараметрів, таких як кількість шарів, розмір ядра, частота відсіву та швидкість навчання, показав, що належна оптимізація цих параметрів може значно вплинути на продуктивність моделі. Аналіз чутливості продемонстрував, що певні конфігурації, наприклад, використання трьохшарової згорткової структури та комбінація певних розмірів ядер, є оптимальними для різних наборів даних. Крім того, вибір оптимальних коефіцієнтів обмеження інтерпретованості та некореляції дозволив досягти максимального розділення інтерпретованих та автономних репрезентацій, що призвело до підвищення точності класифікації.

5. Також було проведено дослідження продуктивності моделі в різних ситуаціях, включаючи вплив віку, статі та шумових факторів. Результати показали, що IG-CNN демонструє стабільну продуктивність у широкому спектрі умов, зокрема, вона більш ефективна у розпізнаванні емоцій у жіночому мовленні та в певних вікових групах. Крім того, модель показала високу стійкість до шумів різного типу, особливо при високому співвідношенні сигнал/шум, що є важливим для її застосування в реальних умовах.

6. Важливою перевагою запропонованої моделі є її інтерпретованість, яка була оцінена на основі аналізу інтерпретованих репрезентацій та вивчених коефіцієнтів прийняття рішень. Було виявлено, що модель може виділяти значущі акустичні маркери, пов'язані з емоційним станом, наприклад, гучність, частоту та інші характеристики, які інтуїтивно відповідають людському сприйняттю емоцій. Це робить модель не тільки високоефективною, але й зрозумілою, що важливо для практичного застосування та подальшого вдосконалення.

Таким чином, дослідження продемонструвало, що запропонована модель IG-CNN є потужним інструментом для розпізнавання емоцій на основі мовлення. Вона поєднує в собі високу ефективність, інтерпретованість та стійкість до різних факторів, що робить її перспективною для

використання в різних сферах, включаючи системи обробки природної мови, людсько-комп'ютерну взаємодію та автоматичний аналіз емоцій. Подальші дослідження можуть бути спрямовані на розширення можливостей моделі та її адаптацію до різноманітних контекстів і застосувань.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. (2018). Applications of support vector machine (SVM) learning in cancer genomics. Cancer genomics & proteomics, 15(1), 41-51.
2. Ezz-Eldin, M., Khalaf, A. A., Hamed, H. F., & Hussein, A. I. (2021). Efficient feature-aware hybrid model of deep learning architectures for speech emotion recognition. IEEE Access, 9, 19999-20011.
3. Najadat, H., & Abushaqra, F. (2018). Multimodal sentiment analysis of Arabic videos. Journal of Image and Graphics, 6(1), 39-43.
4. Doğdu, C., Kessler, T., Schneider, D., Shadaydeh, M., & Schweinberger, S. R. (2022). A comparison of machine learning algorithms and feature sets for automatic vocal emotion recognition in speech. Sensors, 22(19), 7561.
5. Hu, Z., Wang, L., Luo, Y., Xia, Y., & Xiao, H. (2022). Speech Emotion Recognition Model Based on Attention CNN Bi-GRU Fusing Visual Information. Engineering Letters, 30(2).
6. Zheng, C., Wang, C., & Jia, N. (2019). An ensemble model for multi-level speech emotion recognition. Applied Sciences, 10(1), 205.
7. Trinh Van, L., & Le, D. T. T., Le Xuan, T., Castelli, E.: Emotional speech recognition using deep neural networks. Sens. 22 (4), 1414 (2022).
8. Jain, M., Narayan, S., Balaji, P., & Bhowmick, A. (2020). *Speech Emotion Recognition Using Support Vector Machine*. ArXiv preprint arXiv:2002.07590. Retrieved from https://arxiv.org/abs/2002.07590
9. Farooq, M., Hussain, F., Baloch, N. K., Raja, F. R., & Yu, H. (2020). *Impact of Feature Selection Algorithm on Speech Emotion Recognition Using Deep Convolutional Neural Network*. Sensors, 20(21), 6008. MDPI. Retrieved from https://www.mdpi.com/1424-8220/20/21/6008
10. Koduru, A., Valiveti, H. B., & Budati, A. K. (2020). *Feature Extraction Algorithms to Improve the Speech Emotion Recognition Rate*. International

Journal of Speech Technology, Springer. Retrieved from https://link.springer.com/article/10.1007/s10772-020-09672-4

11. Mirsamadi, S., Barsoum, E., & Zhang, C. (2017). *Automatic Speech Emotion Recognition Using Recurrent Neural Networks with Local Attention*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. Retrieved from https://ieeexplore.ieee.org/abstract/document/7952552/
12. Chen, L., Mao, X., Xue, Y., & Cheng, L. L. (2012). *Speech Emotion Recognition: Features and Classification Models*. Digital Signal Processing, 22(5), 1154-1160. Elsevier. Retrieved from https://www.sciencedirect.com/science/article/pii/S1051200412001133
13. Eyben, F., Wöllmer, M., & Schuller, B. (2010). Opensmile: the Munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on multimedia (pp. 1459–1462).
14. Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., & Narayanan, S. S. (2016). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing, 7(2), 190–202.
15. Ioannou, Y., Robertson, D., Cipolla, R., & Criminisi, A. (2017). Deep roots: Improving CNN efficiency with hierarchical filter groups. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 1231–1240).
16. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In 26th conference on neural information processing systems (NeurIPS 2012) (pp. 1097–1105).
17. Rong, J., Li, G., & Chen, Y.-P. (2009). Acoustic feature selection for automatic emotion recognition from speech. Information Processing and Management, 45(3), 315–328.

18. Bandela, S. R., & Kumar, T. K. (2021). Unsupervised feature selection and NMF de-noising for robust speech emotion recognition. Applied Acoustics, 172, Article 107645.

19. Baydilli, Y., & Atila, U. (2018). Understanding effects of hyper-parameters on learning: A comparative analysis. In Proceedings of the international conference on advanced technologies, computer engineering and science (pp. 11–13).

20. Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. Language Resources and Evaluation, 42, 335–359.

21. Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkova, A., & Verma, R. (2014). Crema-d: Crowd-sourced emotional multimodal actors dataset. IEEE Transactions on Affective Computing, 5(4), 377–390.

22. Chen, Y.-C., Lo, S.-C., Wang, S., Wang, Y.-J., Wu, M., & Fang, W. (2021). On the PZT/Si unimorph cantilever design for the signal-to-noise ratio enhancement of piezoelectric MEMS microphone. Journal of Micromechanics and Microengineering, 31(10), Article 105003.

23. Cho, Y., Park, B., Lee, S. M., Lee, K. H., Seo, J. B., & Kim, N. (2021). Optimal number of strong labels for curriculum learning with convolutional neural network to classify pulmonary abnormalities in chest radiographs. Computers in Biology and Medicine, 136, Article 104750.

24. Gobl, C., & Chasaide, A. (2010). Voice source variation and its communicative functions. In J. L. W. J. Hardcastle, & F. E. Gibbon (Eds.), The handbook of phonetic sciences (2nd ed., pp. 378–423).

25. Goyal, P., Dollár P., Girshick R., Noordhuis P., Wesolowski L., Kyrola A., Tulloch A., Jia Y. and He K. (2017). Accurate, large minibatch sgd: Training imagenet in 1 h. arXiv Preprint, 12.

26. Jiang, L., Li, R., Wu, W., Qian, C., & Loy, C. C. (2020). Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) (pp. 2886–2895). Virtual Online.

27. Levenson, R. W., Carstensen, L. L., Friesen, W. V., & Ekman, P. (1991). Emotion, physiology, and expression in old age. Psychology and Aging, 6(1), 28–35.

28. Livingstone, S. R., & Russo, F. A. (2018). The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLoS One, 13(5), Article e0196391.

29. Makita, M., Mas-Bleda, A., Stuart, E., & Thelwall, M. (2021). Ageing, old age and older adults: A social media analysis of dominant topics and discourses. Ageing and Society, 41(2), 247–272.

30. Martin, O., Kotsia, I., Macq, B., & Pitas, I. (2006). The eNTERFACE'05 audio-visual emotion database. In 22nd international conference on data engineering workshops (ICDEW'06) (p. 8).

31. Mirsamadi, S., Barsoum, E., & Zhang, C. (2017). Automatic speech emotion recognition using recurrent neural networks with local attention. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 2227–2231).

32. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The journal of Machine Learning Research, 15(1), 1929–1958.

33. Tariq, Z., Shah, S. K., & Lee, Y. (2019). Speech emotion detection using IoT based deep learning for health care. In 2019 IEEE international conference on big data (Big Data) (pp. 4191–4196).

34. Waaramaa, T., Laukkanen, A.-M., Airas, M., & Alku, P. (2010). Perception of emotional valences and activity levels from vowel segments of continuous speech. Journal of Voice, 24(1), 30–38.

35. Загальні методичні рекомендації з підготовки, оформлення, захисту та оцінювання кваліфікаційних робіт здобувачів вищої освіти першого (бакалаврського) і другого (магістерського) рівнів. Тернопіль: ЗУНУ, 2024. 83 с.

36. Комар М.П., Саченко А.О., Васильків Н.М., Загородня Д.І. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за другим (магістерським) рівнем вищої освіти. – Тернопіль: ЗУНУ, 2024. – 32 с.

Додаток А

Порівняння ефективності з базовими методами на тестовому наборі

Таблиця А.1. Порівняння ефективності з базовими методами на тестовому наборі

Дані	**Категорія бенчмарку**	**Модель**	АкАкАк	PmacroPmacro	RmacroRmacro	FmacroFmacro	FmicroFmicro	FweightFweight
IEMOCAP		IG-CNN	**0.620**	**0.624**	**0.628**	**0.622**	**0.620**	**0.617**
	Традиційне машинне навчання	LR	0.529***	0.544***	0.531***	0.535***	0.529***	0.530***
		SVM	0.551***	0.555***	0.553***	0.552***	0.551***	0.549***
		kNN	0.505***	0.502***	0.513***	0.505***	0.505***	0.499***
		Наївний Байєс	0.479***	0.483***	0.492***	0.479***	0.479***	0.469***
		Дерево рішень	0.517***	0.523***	0.524***	0.522***	0.517***	0.515***
	Глибоке навчання	CNN-1D	0.537***	0.545***	0.546***	0.543***	0.537***	0.533***
		CNN-2D	0.583***	0.587***	0.594***	0.588***	0.583***	0.583***
		BLSTM	0.545***	0.550***	0.553***	0.550***	0.545***	0.542***
		DeepRNN	0.601***	0.612**	0.604***	0.605***	0.601***	0.599***
		CLSTM	0.585***	0.592***	0.591***	0.589***	0.585***	0.583***
RAVDESS		IG-CNN	**0.648**	**0.644**	**0.639**	**0.632**	**0.648**	**0.641**

	Традиційне машинне навчання	LR	0.563***	0.567***	0.585***	0.569***	0.554***	0.553***
		SVM	0.430***	0.455***	0.438***	0.424***	0.402***	0.406***
		kNN	0.385***	0.415***	0.405***	0.387***	0.421***	0.409***
		Наївний Байєс	0.437***	0.480***	0.466***	0.417***	0.465***	0.427***
		Дерево рішень	0.445***	0.449***	0.462***	0.449***	0.468***	0.473***
	Глибоке навчання	CNN-1D	0.486***	0.501***	0.504***	0.452***	0.526***	0.513***
		CNN-2D	0.621*	0.616*	0.634*	0.608*	0.625*	0.618*
		BLSTM	0.553***	0.534***	0.537***	0.529***	0.553***	0.546***
		DeepRNN	0.611***	0.580***	0.584***	0.576***	0.611***	0.605***
		CLSTM	0.629**	0.622**	0.614**	0.611*	0.629**	0.631*
eNTERFACE'05		IG-CNN	**0.654**	**0.658**	**0.652**	**0.649**	**0.654**	**0.653**
	Традиційне машинне навчання	LR	0.421***	0.396***	0.417***	0.403***	0.421***	0.410***
		SVM	0.345***	0.343***	0.334***	0.336***	0.345***	0.351***
		kNN	0.310***	0.284***	0.299***	0.286***	0.310***	0.397***
		Наївний Байєс	0.361***	0.349***	0.355***	0.314***	0.361***	0.321***
		Дерево рішень	0.358***	0.334***	0.344***	0.327***	0.358***	0.339***

	Глибоке навчання	CNN-1D	0.411***	0.400***	0.398***	0.383***	0.411***	0.396***
		CNN-2D	0.550***	0.545***	0.549***	0.542***	0.550***	0.546***
		BLSTM	0.436***	0.428***	0.429***	0.421***	0.436***	0.432***
		DeepRNN	0.477***	0.472***	0.479***	0.467***	0.477***	0.472***
		CLSTM	0.611**	0.611***	0.608**	0.604**	0.611**	0.610**
CREMA-D		IG-CNN	**0.580**	**0.585**	**0.585**	**0.579**	**0.580**	**0.577**
	Традиційне машинне навчання	LR	0.473***	0.473***	0.477***	0.468***	0.473***	0.465***
		SVM	0.426***	0.423***	0.433***	0.420***	0.426***	0.416***
		kNN	0.369***	0.348***	0.380***	0.343***	0.369***	0.337***
		Наївний Байєс	0.360***	0.339***	0.381***	0.308***	0.360***	0.299***
		Дерево рішень	0.422***	0.425***	0.430***	0.411***	0.422***	0.408***
	Глибоке навчання	CNN-1D	0.524***	0.524***	0.530***	0.520***	0.524***	0.517***
		CNN-2D	0.534***	0.536***	0.539***	0.530***	0.534***	0.527***
		BLSTM	0.462***	0.456***	0.466***	0.456***	0.462***	0.453***
		DeepRNN	0.508***	0.510***	0.512***	0.506***	0.508***	0.504***
		CLSTM	0.568*	0.570*	0.574*	0.567*	0.568*	0.564*

Примітка: *: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$. ЛР: логістична регресія.

Додаток Б

Порівняння продуктивності з варіантами IG-CNN

Таблиця Б.1. Порівняння продуктивності з варіантами IG-CNN

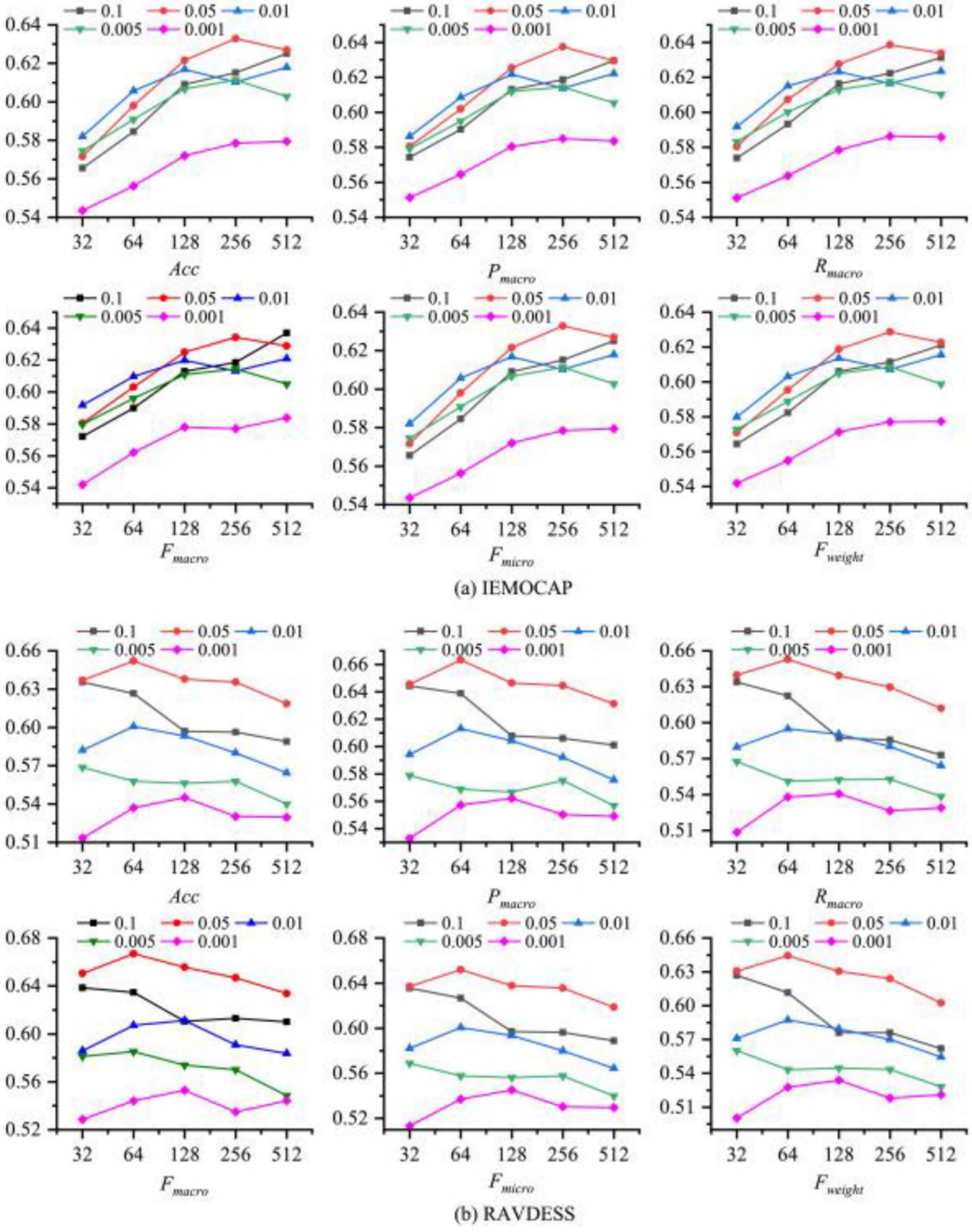
Дані	**Категорія бенчмарку**	**Модель**	Ак	Pmacro	Rmacro	Fmacro	Fmicro	Fweight
IEMOCAP		IG-CNN	**0.620**	**0.624**	**0.628**	**0.622**	**0.620**	**0.617**
	Варіанти IG-CNN	IG-CNN_1	0.587***	0.591***	0.599***	0.592***	0.610*	0.608*
		IG-CNN_2	0.610**	0.617*	0.618**	0.614**	0.587***	0.584***
		IG_CNN_3	0.608***	0.613**	0.617***	0.611**	0.608***	0.605**
RAVDESS		IG-CNN	**0.648**	**0.644**	**0.639**	**0.632**	**0.648**	**0.641**
	Варіанти IG-CNN	IG-CNN_1	0.613***	0.604***	0.601***	0.595***	0.613***	0.606***
		IG-CNN_2	0.585***	0.566***	0.575***	0.559***	0.585***	0.571***
		IG_CNN_3	0.628**	0.623*	0.622**	0.612*	0.628*	0.6211*
eNTERFACE'05		IG-CNN	**0.654**	**0.658**	**0.652**	**0.649**	**0.654**	**0.653**
	Варіанти IG-CNN	IG-CNN_1	0.634*	0.633*	0.628*	0.624*	0.634*	0.633*
		IG-CNN_2	0.627*	0.628*	0.622*	0.619*	0.627*	0.626*
		IG_CNN_3	0.632*	0.638*	0.629*	0.624*	0.633*	0.630*
CREMA-D		IG-CNN	**0.580**	**0.585**	**0.585**	**0.579**	**0.580**	**0.577**

	Варіанти IG-CNN	IG-CNN_1	0.559*	0.564*	0.565*	0.555*	0.559*	0.552*
		IG-CNN_2	0.563*	0.564*	0.567*	0.562*	0.563*	0.559*
		IG_CNN_3	0.565*	0.568*	0.572*	0.563*	0.565*	0.560*

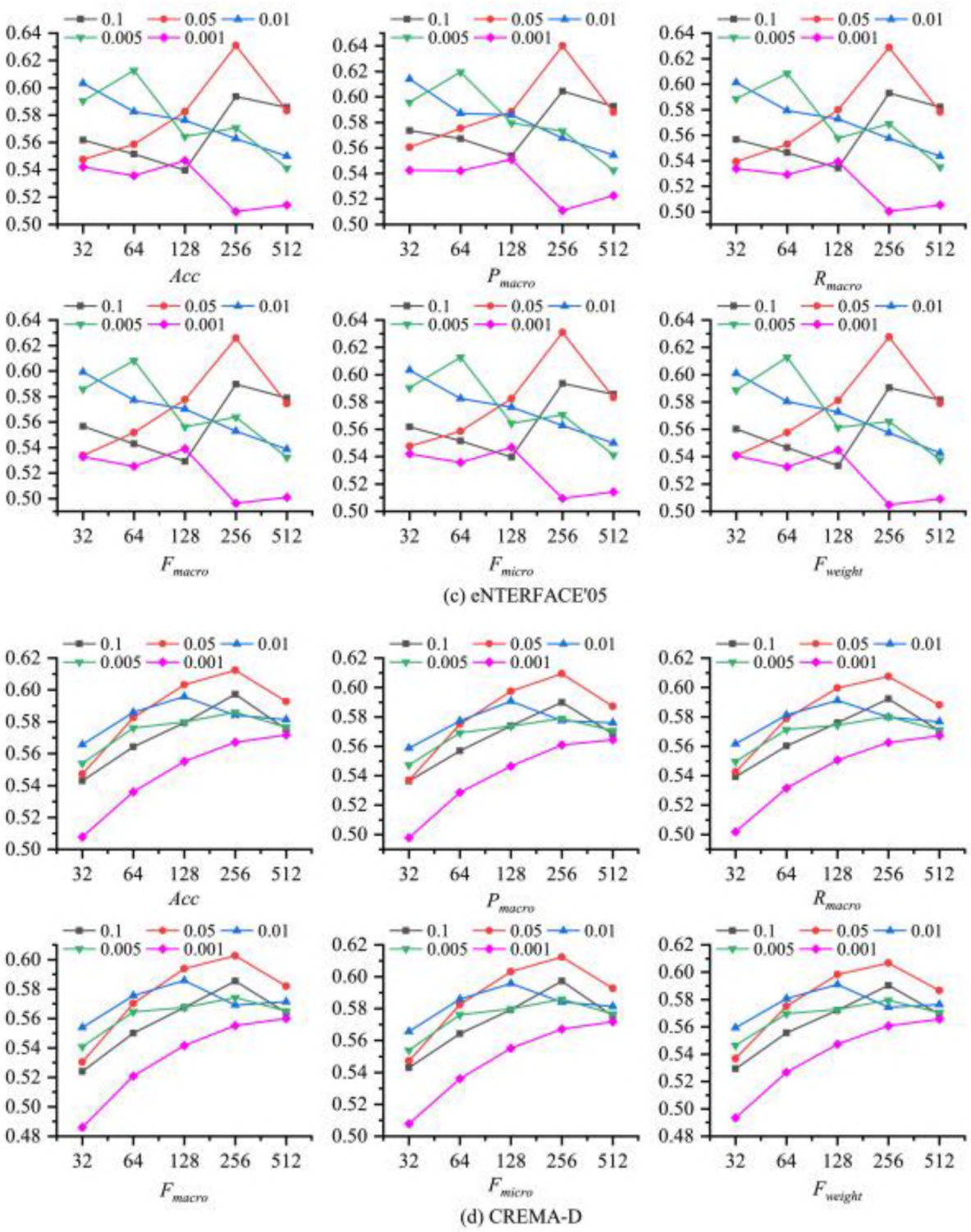
Примітка: *: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$.

Додаток В

Продуктивність IG-CNN при комбінаціях навчання швидкості та розміру партії



Рисунку В.1 - Продуктивність IG-CNN при комбінаціях навчання швидкості та розміру партії



Рисунку В.1 - Продуктивність IG-CNN при комбінаціях навчання швидкості та розміру партії (продовження)

Додаток Г

Т-критерій вікової специфічності CEP на CREMA-D

Таблиця Г.1 - Т-критерій вікової специфічності CEP на CREMA-D

Метрики	**Вік**	**20-25**	**26-30**	**31-35**	**36-40**	**41-45**	**46-50**	**51-55**	**56-60**
Точність	26-30	**0.3111**							
	31-35	**0.6439**	**0.5759**						
	36-40	**0.2178**	**0.4661**	**0.2342**					
	41-45	**0.4412**	**0.9722**	**0.6550**	**0.4958**				
	46-50	**0.0573**	0.0498	**0.0998**	**0.3643**	**0.0656**			
	51-55	0.0255	0.0177	0.0206	0.0573	0.0195	0.0130		
	56-60	0.0030	0.0091	0.0051	0.0011	0.0083	0.0153	**0.6375**	
	60-	0.0001	0.0017	0.0002	0.0156	0.0043	**0.3386**	0.0003	0.0001
Макропрецизійність	26-30	**0.2556**							
	31-35	**0.5855**	**0.1100**						
	36-40	**0.2629**	**0.5758**	**0.1277**					
	41-45	**0.1214**	**0.1450**	**0.0507**	**0.6224**				
	46-50	0.0406	0.0162	**0.0687**	**0.0743**	0.0129			
	51-55	0.0486	0.0423	0.0290	**0.2716**	**0.3103**	0.0148		
	56-60	0.0033	0.0115	0.0032	0.0047	0.0429	0.0082	**0.2401**	
	60-	0.0005	0.0008	0.0019	0.0021	0.0042	**0.4864**	0.0009	0.0005

Відкликання макросу	26-30	**0.0925**							
	31-35	**0.8479**	0.0489						
	36-40	**0.3395**	**0.1207**	**0.3059**					
	41-45	**0.9289**	**0.1149**	**0.9575**	**0.4419**				
	46-50	0.0213	0.0060	0.0320	**0.1228**	0.0119			
	51-55	0.0159	0.0936	0.0279	**0.0691**	0.0244	0.0053		
	56-60	0.0122	**0.1606**	0.0375	0.0083	0.0450	0.0115	**0.9841**	
	60-	0.0005	0.0013	0.0015	0.0366	0.0054	**0.8511**	0.0001	0.0013
Оцінка Macro F1	26-30	**0.6452**							
	31-35	**0.8855**	**0.7504**						
	36-40	**0.3260**	**0.5109**	**0.3048**					
	41-45	**0.7654**	**0.9987**	**0.8442**	**0.5242**				
	46-50	0.0236	0.0110	0.0343	**0.1454**	0.0180			
	51-55	0.0436	0.0498	**0.0525**	**0.0899**	**0.0689**	0.0095		
	56-60	0.0052	0.0167	0.0123	0.0010	0.0138	0.0087	**0.5243**	
	60-	0.0001	0.0012	0.0004	0.0125	0.0053	**0.6240**	0.0003	0.0004
Оцінка Micro F1	26-30	**0.3111**							
	31-35	**0.6439**	**0.5759**						
	36-40	**0.2178**	**0.4661**	**0.2342**					
	41-45	**0.4412**	**0.9722**	**0.6550**	**0.4958**				
	46-50	**0.0573**	0.0498	**0.0998**	**0.3643**	**0.0656**			

	51-55	0.0255	0.0177	0.0206	**0.0573**	0.0195	0.0130		
	56-60	0.0030	0.0091	0.0051	0.0011	0.0083	0.0153	**0.6375**	
	60-	0.0001	0.0017	0.0002	0.0156	0.0043	**0.3386**	0.0003	0.0001
Зважений бал F1	26-30	**0.1076**							
	31-35	**0.7499**	**0.2165**						
	36-40	**0.1437**	**0.5667**	**0.1467**					
	41-45	**0.3794**	**0.7715**	**0.4837**	**0.4667**				
	46-50	0.0381	0.0401	**0.0680**	**0.3496**	**0.0575**			
	51-55	0.0652	0.0309	**0.0510**	**0.0691**	**0.0551**	0.0177		
	56-60	0.0045	0.0085	0.0072	0.0004	0.0077	0.0116	**0.5714**	
	60-	0.0001	0.0012	0.0001	0.0091	0.0060	**0.2620**	0.0011	0.0002

Примітка: Жирний шрифт у таблиці означає, що *p-value* > 0,05.

Додаток Д

Продуктивність моделі при різних шуму рівнях

Таблиця Д.1 - Продуктивність моделі при різних шуму рівнях на RAVDESS

Порожня комірка	Порожня комірка	Ак	Pmacro	Rmacro	Fmacro	Fmicro	Fweight
Чистий		0.648	0.644	0.639	0.632	0.648	0.641
Адитивний білий гаусівський шум	SNR = 50	0.642	0.645	0.636	0.624	0.642	0.634
	SNR = 45	0.636	0.631	0.627	0.617*	0.636	0.629
	SNR = 40	0.641	0.639	0.636	0.629	0.641	0.637
	SNR = 35	0.629*	0.625*	0.620*	0.612*	0.629*	0.622*
	SNR = 30	0.633**	0.628**	0.625*	0.617**	0.633**	0.627**
Шум у салоні автомобіля	SNR = 50	0.633	0.633	0.625	0.615	0.633	0.625
	SNR = 45	0.632	0.629*	0.622	0.609*	0.632	0.620
	SNR = 40	0.638	0.626*	0.628	0.617	0.638	0.629
	SNR = 35	0.578***	0.586***	0.575***	0.560***	0.578***	0.570***
	SNR = 30	0.589***	0.586***	0.580***	0.568***	0.589***	0.583***
Дзюрчання	SNR = 50	0.644	0.645	0.635	0.630	0.644	0.640
	SNR = 45	0.627**	0.621*	0.618**	0.608**	0.627**	0.619**

	SNR = 40	0.620**	0.616***
	SNR = 35	0.602***	0.607***
	SNR = 30	0.606***	0.606***

Примітка: *: $p < 0,05$, **: $p < 0,01$, ***: $p < 0,001$.

	0.609*	0.696***	0.620**	0.609**
	0.592***	0.583***	0.602***	0.595***
	0.594***	0.582***	0.606***	0.596***

Таблиця Д.2 -Продуктивність моделі при різних шуму рівнях на CREMA-D

Порожня комірка	Порожня комірка	Ак	Pmacro	Rmacro	Fmacro	Fmicro	Fweight
Чистий		0.580	0.585	0.585	0.579	0.580	0.577
Адитивний білий гаусівський шум	SNR = 50	0.578	0.580	0.583	0.577	0.578	0.574
	SNR = 45	0.577	0.580	0.582	0.577	0.577	0.574
	SNR = 40	0.572*	0.579	0.577	0.570*	0.572*	0.567*
	SNR = 35	0.561**	0.566***	0.566**	0.559**	0.561**	0.556**
	SNR = 30	0.560**	0.564**	0.567**	0.559*	0.560**	0.555*
Шум у салоні автомобіля	SNR = 50	0.580	0.581	0.585	0.579	0.580	0.577
	SNR = 45	0.580	0.583	0.586	0.580	0.580	0.577
	SNR = 40	0.579	0.581	0.585	0.577	0.579	0.575
	SNR = 35	0.567*	0.567*	0.572*	0.566	0.567*	0.563*
	SNR = 30	0.567**	0.571*	0.573*	0.566**	0.567**	0.563**
Дзюрчання	SNR = 50	0.580	0.582	0.585	0.580	0.580	0.577
	SNR = 45	0.579	0.581	0.585	0.579	0.579	0.575
	SNR = 40	0.580	0.584	0.585	0.580	0.580	0.577
	SNR = 35	0.580	0.582	0.586	0.579	0.580	0.576
	SNR = 30	0.569*	0.572**	0.574*	0.569*	0.569*	0.565*

Примітка: *: $p < 0{,}05$, **: $p < 0{,}01$, ***: $p < 0{,}001$.

Додаток Е

Апробація отриманих результатів

ЗБІРНИК НАУКОВИХ ПРАЦЬ

З МАТЕРІАЛАМИ V МІЖНАРОДНОЇ НАУКОВОЇ КОНФЕРЕНЦІЇ

11 ЖОВТНЯ 2024 РІК

М. ЛУЦЬК, УКРАЇНА

**«СТРАТЕГІЧНІ НАПРЯМКИ РОЗВИТКУ НАУКИ:
ФАКТОРИ ВПЛИВУ ТА ВЗАЄМОДІЇ»**

УДК 082:001
С 83

Організація, від імені якої випущено видання:
ГО «Міжнародний центр наукових досліджень»
Номер запису організації в Єдиному реєстрі громадських об'єднань: 1499141

Голова оргкомітету: Сотник С.Г.

Верстка: Білоус Т.В.
Дизайн: Бондаренко І.В.

Рекомендовано до видання Вченою Радою Інституту науково-технічної інтеграції та співпраці. Протокол № 57 від 10.10.2024 року.

*Конференцію зареєстровано Державною науковою установою у сфері управління Міністерства освіти і науки «Український інститут науково-технічної експертизи та інформації» в базі даних науково-технічних заходів України на поточний рік та бюлетені «План проведення науковиx, науково-технічних заходів в Україні» (**Посвідчення № 351 від 12.06.2024**).*

*Збірник наукових праць з матеріалами конференції видано офіційно суб'єктом видавничої справи зі **Свідоцтвом ДК № 7860 від 22.06.2023***

Матеріали конференції знаходяться у відкритому доступі на умовах ліцензії Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

С 83 **Стратегічні напрямки розвитку науки: фактори впливу та взаємодії**: збірник наукових праць з матеріалами V Міжнародної наукової конференції, м. Луцьк, 11 жовтня, 2024 р. / Міжнародний центр наукових досліджень. — Вінниця: ТОВ «УКРЛОГОС Груп, 2024. — 254 с.

ISBN 978-617-8440-16-9
DOI 10.62731/mcnd-11.10.2024

Викладено матеріали учасників V Міжнародної наукової конференції «Стратегічні напрямки розвитку науки: фактори впливу та взаємодії», яка відбулася 11 жовтня 2024 року у місті Луцьк.

УДК 082:001

© Колектив учасників конференції, 2024
© ГО «Міжнародний центр наукових досліджень», 2024
© ТОВ «УКРЛОГОС Груп», 2024

ISBN 978-617-8440-16-9

СЕКЦІЯ XI.
ХІМІЯ, ХІМІЧНА ТА БІОІНЖЕНЕРІЯ

ЗАСТОСУВАННЯ НЕОРГАНІЧНИХ СПОЛУК ДЛЯ НАДАННЯ ВОЛОКНИСТИМ МАТЕРІАЛАМ БАКТЕРИЦИДНИХ ВЛАСТИВОСТЕЙ У МЕДИЧНИХ І ПРОФІЛАКТИЧНИХ ЦІЛЯХ
Качківський В.В., Попович Т.А. **130**

СЕКЦІЯ XII.
ЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

ЗАСТОСУВАННЯ ТЕХНОЛОГІЇ БЕЗДРОТОВОЇ ЗАРЯДКИ ЗА ДОПОМОГОЮ РЕКТЕН ДЛЯ ЖИВЛЕННЯ СИСТЕМ РЕЗ
Сокіркаєв Д.В. **135**

СЕКЦІЯ XIII.
ЕКОЛОГІЯ ТА ТЕХНОЛОГІЇ ЗАХИСТУ НАВКОЛИШНЬОГО СЕРЕДОВИЩА

ІМПЛЕМЕНТАЦІЯ ЄВРОПЕЙСЬКОГО ДОСВІДУ ЕКОЛОГІЧНОГО РЕГУЛЮВАННЯ ВИКОРИСТАННЯ ТЕРИТОРІЙ ПРИРОДНО-ЗАПОВІДНОГО ФОНДУ
Голян В.М., Іванців В.В. **138**

СЕКЦІЯ XIV.
КОМП'ЮТЕРНА ТА ПРОГРАМНА ІНЖЕНЕРІЯ

HARDWARE SUPPORT OF INFORMATION SYSTEMS
Pavlyk H.V. **142**

СЕКЦІЯ XV.
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СИСТЕМИ

ПОБУДОВА ГНУЧКИХ СТРАТЕГІЙ УПРАВЛІННЯ ЗМІНАМИ В ML-ПРОЄКТАХ
Науково-дослідна група:
Висоцький А.В., Левандівський Н.Б., Вівюрка Н.М., Сулима Б.Я. **151**

ПРИКЛАДНІ АСПЕКТИ ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ОБРОБКИ ТЕКСТОВИХ ДАНИХ
Науково-дослідна група:
Колівушко Е., Лучка С., Кондратюк Г., Кравчук Б., Тарасюк С. **154**

ПРИКЛАДНІ АСПЕКТИ ВИКОРИСТАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ ОБРОБКИ ТЕКСТОВИХ ДАНИХ

НАУКОВО-ДОСЛІДНА ГРУПА:

Колівушко Едуард
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Лучка Святослав
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Кондратюк Георгій
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Кравчук Богдан
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Тарасюк Софія
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Науковий керівник: Ліп'яніна-Гончаренко Христина
канд. техн. наук, доцент,
доцент кафедри інформаційно-обчислювальних систем та управління
Західноукраїнський національний університет, Україна

Машинне навчання (ML) кардинально змінило підходи до обробки текстових даних, дозволяючи автоматизувати складні завдання, такі як класифікація тексту, аналіз емоцій та виявлення фейкових новин. У зв'язку зі зростанням обсягів текстової інформації в Інтернеті, особливо на соціальних платформах, машинне навчання стало ключовим інструментом для швидкої і точної обробки цих даних. За статистикою, кожного дня публікується понад 2,5 мільярда текстових записів на платформах соціальних мереж [5], що робить застосування машинного навчання надзвичайно актуальним.

Існує декілька основних методів машинного навчання, які активно використовуються для обробки тексту. Серед них найпопулярніші — наївний Баєс, метод опорних векторів (SVM), логістична регресія та дерева рішень. Проте з розвитком глибинного навчання в останні роки стали особливо популярними рекурентні нейронні мережі (RNN) та трансформери (наприклад, BERT), що забезпечують високий рівень точності в завданнях класифікації тексту і прогнозуванні наступних слів у реченнях [3].

Попередня обробка тексту є критично важливою для підготовки даних до аналізу. Найбільш ефективними методами є стемінг, лематизація та видалення стоп-слів. Стемінг перетворює слова до їх основних форм, що допомагає зменшити варіативність даних, тоді як лематизація враховує контекст і надає точніші результати. Видалення стоп-слів допомагає зменшити розмір тексту, видаляючи непотрібні слова, які не несуть суттєвої інформації для моделі [4].

Якість попередньої обробки тексту безпосередньо впливає на результати моделей машинного навчання. Якщо дані погано підготовлені, моделі можуть працювати неефективно через надмірну кількість шуму або неправильні структури даних. Наприклад, неправильне видалення стоп-слів або відсутність лематизації може призвести до того, що модель не зможе точно розпізнати ключові патерни тексту, що знижує точність класифікації [2]. Таким чином, ретельна обробка тексту є важливим етапом у процесі машинного навчання.

Векторизація тексту — це процес перетворення текстових даних у числові представлення, які можуть бути використані моделями машинного навчання. Найпоширеніші підходи включають Bag of Words (BoW), TF-IDF і word embeddings (наприклад, Word2Vec, GloVe). BoW і TF-IDF є простими методами, що використовуються для побудови векторів на основі частоти слів. Однак word embeddings дозволяють моделі навчитися семантичним зв'язкам між словами, що значно покращує якість аналізу тексту [3].

Вибір техніки векторизації залежить від характеру задачі. Якщо основний фокус на простій класифікації тексту, такі методи, як TF-IDF або Bag of Words, можуть бути достатніми. Однак для задач, що вимагають врахування контексту та семантики (наприклад, аналіз емоцій або прогнозування наступних слів), більш ефективними будуть word

embeddings або трансформери на основі BERT. Для задач, де точність є критично важливою, варто віддавати перевагу більш складним підходам [1].

Аналіз емоцій є одним з основних застосувань машинного навчання для обробки тексту, особливо у сфері маркетингу, соціальних мереж та підтримки клієнтів. Алгоритми машинного навчання, такі як нейронні мережі та методи класифікації, можуть використовуватися для визначення тональності тексту, що дозволяє автоматично аналізувати настрої користувачів. Успішні кейси використання цих методів вже застосовуються у великих компаніях, таких як Amazon і Google, для аналізу зворотного зв'язку клієнтів [2].

У результаті дослідження виявлено, що машинне навчання є ефективним інструментом для обробки текстових даних, особливо при правильній попередній обробці та виборі підходящої техніки векторизації. Для більш складних завдань, таких як аналіз емоцій або семантична класифікація, рекомендується використовувати сучасні моделі глибинного навчання, що базуються на embeddings або трансформерах. Подальші дослідження можуть зосередитися на оптимізації процесу векторизації та покращенні якості моделей на багатомовних даних.

Список використаних джерел:

1. Vaswani, A., Shazeer, N., Parmar, N., et al. "Attention Is All You Need." *Advances in Neural Information Processing Systems*, 2017.
2. Pang, B., & Lee, L. "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval*, 2008.
3. Mikolov, T., Chen, K., Corrado, G., & Dean, J. "Efficient Estimation of Word Representations in Vector Space." *arXiv preprint arXiv:1301.3781*, 2013.
4. Manning, C., Raghavan, P., & Schütze, H. "Introduction to Information Retrieval." *Cambridge University Press*, 2008.
5. Statista. "Number of social media users worldwide from 2010 to 2021." [Online]. Available: www.statista.com.

ЗБІРНИК НАУКОВИХ ПРАЦЬ

З МАТЕРІАЛАМИ III МІЖНАРОДНОЇ НАУКОВОЇ КОНФЕРЕНЦІЇ

20 ВЕРЕСНЯ 2024 РІК

М. ОДЕСА, УКРАЇНА

**«ІНТЕЛЕКТУАЛЬНИЙ РЕСУРС СЬОГОДЕННЯ:
НАУКОВІ ЗАДАЧІ, РОЗВИТОК ТА ЗАПИТАННЯ»**

УДК 082:001
І-57

Організація, від імені якої випущено видання:
ГО «Міжнародний центр наукових досліджень»
Номер запису організації в Єдиному реєстрі громадських об'єднань: 1499141

Голова оргкомітету: Сотник С.Г.

Верстка: Бабич Ю.В.
Дизайн: Бондаренко І.В.

Рекомендовано до видання Вченою Радою Інституту науково-технічної інтеграції та співпраці. Протокол № 54 від 19.09.2024 року.

*Конференцію зареєстровано Державною науковою установою у сфері управління Міністерства освіти і науки «Український інститут науково-технічної експертизи та інформації» в базі даних науково-технічних заходів України на поточний рік та бюлетені «План проведення наукових, науково-технічних заходів в Україні» (**Посвідчення № 348 від 12.06.2024**).*

*Збірник наукових праць з матеріалами конференції видано офіційно суб'єктом видавничої справи зі **Свідоцтвом ДК № 7860 від 22.06.2023***

Матеріали конференції знаходяться у відкритому доступі на умовах ліцензії Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

І-57 **Інтелектуальний ресурс сьогодення: наукові задачі, розвиток та запитання**: збірник наукових праць з матеріалами III Міжнародної наукової конференції, м. Одеса, 20 вересня, 2024 р. / Міжнародний центр наукових досліджень. — Вінниця: ТОВ «УКРЛОГОС Груп, 2024. — 322 с.

ISBN 978-617-8440-13-8
DOI 10.62731/mcnd-20.09.2024

Викладено матеріали учасників III Міжнародної наукової конференції «Інтелектуальний ресурс сьогодення: наукові задачі, розвиток та запитання», яка відбулася 20 вересня 2024 року у місті Одеса.

УДК 082:001

© Колектив учасників конференції, 2024
© ГО «Міжнародний центр наукових досліджень», 2024
© ТОВ «УКРЛОГОС Груп», 2024

ISBN 978-617-8440-13-8

СЕКЦІЯ X.
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СИСТЕМИ

ОГЛЯД КОНЦЕПЦІЙ ТА МЕТОДІВ МАШИННОГО НАВЧАННЯ: ТЕОРІЯ, ЗАСТОСУВАННЯ ТА ВИКЛИКИ
Кондратюк Г., Кравчук Б., Тарасюк С., Матвійчук М. **177**

СЕКЦІЯ XI.
ТРАНСПОРТ ТА ТРАНСПОРТНІ ТЕХНОЛОГІЇ

INNOVATIVE TECHNOLOGIES AND REGULATORY MEASURES TO REDUCE ENVIRONMENTAL RISKS IN THE SHIPPING INDUSTRY
Sagaydak O., Kucherenko V., Kotenko O., Prokhorov V. **181**

СЕКЦІЯ XII.
ФІЗИКО-МАТЕМАТИЧНІ НАУКИ

ПЕРЕХІД МЕТАЛ-ІЗОЛЯТОР У МОНОКРИСТАЛАХ $Y_{1-x}Pr_xBa_2Cu_3O_{7-\delta}$ З РІЗНИМ ВМІСТОМ ПРАЗЕОДИМУ
Камчатна С. М., Ярчук Д. Ф., Чепурін О. Г., Іноземцев М. М., Вовк Р. В. **187**

СЕКЦІЯ XIII.
СОЦІОЛОГІЯ ТА СТАТИСТИКА

СІМ'Я ЯК СОЦІАЛЬНИЙ ІНСТИТУТ. ФУНКЦІЇ СІМ'Ї
Олійник М. В. **191**

СЕКЦІЯ XIV.
ФІЛОЛОГІЯ ТА ЖУРНАЛІСТИКА

FROM BEATS TO WORDS: HOW MUSIC NORMALIZES CREATIVE AND IMPLICIT ENGLISH LANGUAGE
Hulei T. **196**

ФУНКЦІОНУВАННЯ НЕФОРМАЛЬНОЇ ЛЕКСИКИ У СУЧАСНИХ ПІСНЯХ
Борисова К. М. **205**

АВТОР І ТЕКСТ У РОМАНІ У. ЕКО «ІМ'Я ТРОЯНДИ»
Кулик К. В. **209**

ЗАСОБИ ПЕРЕКЛАДУ РЕАЛІЙ В ІСТОРИЧНОМУ ДЕТЕКТИВІ (НА МАТЕРІАЛІ РОМАНУ К. ДЖ. СЕНСОМА «DISSOLUTION»)
Марченко В. В. **212**

СЕКЦІЯ X.
ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СИСТЕМИ

ОГЛЯД КОНЦЕПЦІЙ ТА МЕТОДІВ МАШИННОГО НАВЧАННЯ: ТЕОРІЯ, ЗАСТОСУВАННЯ ТА ВИКЛИКИ

Кондратюк Георгій
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Кравчук Богдан
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Тарасюк Софія
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Матвійчук Микола
здобувач вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

Науковий керівник: Ліп'яніна-Гончаренко Христина
канд. тех. наук, доцент, доцент кафедри інформаційно-обчислювальних систем та управління
Західноукраїнський національний університет, Україна

Машинне навчання є однією з ключових технологій, що трансформують різні галузі, від медицини до інженерії та промисловості. З розвитком обчислювальної потужності та зростанням обсягів даних, методи машинного навчання стають все більш складними та ефективними. Однак, щоб повністю зрозуміти та застосувати ці методи, необхідно володіти глибоким знанням теоретичних концепцій та практичних підходів. У цьому огляді зібрано ключові публікації, які досліджують різні аспекти машинного навчання, починаючи від основ статистичного навчання і закінчуючи використанням машинного

навчання в різних галузях, таких як медицина, телекомунікації та автоматизація. Метою цього огляду є узагальнення та систематизація знань у цій сфері для кращого розуміння сучасних методів машинного навчання.

У публікації "Machine Learning Methods: An Overview" [1] автор R. Muhamedyev надає загальний огляд методів машинного навчання, їх класифікацію та застосування. Стаття розглядає різні категорії алгоритмів, включаючи навчання з учителем і без учителя, а також надає практичні приклади їх використання. Особлива увага приділяється аналізу переваг та обмежень різних підходів, що робить цю роботу корисною для початківців у цій сфері.

A.F.A.H. Alnuaimi та T.H.K. Albaldawi у своїй роботі [2] представляють фундаментальні концепції статистичного навчання та класифікації в машинному навчанні. Автори висвітлюють основні методи та алгоритми, зокрема регресію та класифікацію, і пояснюють їх роль у вирішенні складних завдань. Ця стаття акцентує увагу на важливості статистичних підходів у машинному навчанні та їх впливі на точність моделей.

У статті "Logical Approaches to Machine Learning" [3] автор P. Flach досліджує логічні підходи до машинного навчання. Ця публікація підкреслює важливість логіки та математичних основ для створення ефективних алгоритмів навчання. Вона розглядає різні типи логічних моделей та їхню здатність вирішувати складні завдання, забезпечуючи глибше розуміння теоретичних аспектів машинного навчання.

У публікації [4] автори S. Albahra та співавтори досліджують застосування методів машинного навчання та штучного інтелекту в патології та лабораторній медицині. Стаття надає огляд ключових алгоритмів та їх застосувань у медичному аналізі, демонструючи, як ці методи можуть покращити діагностику та підвищити ефективність медичних процедур.

H.M. El Misilmani та T. Naous у своїй роботі [5] розглядають використання машинного навчання в проєктуванні антен. Вони описують різні концепції машинного навчання та алгоритми, які можуть бути застосовані для оптимізації антен. Стаття надає практичні приклади використання машинного навчання у галузі телекомунікацій, демонструючи його ефективність та потенціал для інженерних задач.

У публікації [6] автори D. Kreuzberger, N. Kühl, та S. Hirschl досліджують концепцію Machine Learning Operations (MLOps). Вони

надають визначення MLOps, пояснюють його архітектуру та важливість для ефективного розгортання моделей машинного навчання в промислових додатках. Ця стаття є ключовою для розуміння того, як інтегрувати машинне навчання у виробниче середовище.

J.G. Carbonell, R.S. Michalski, та T.M. Mitchell у своїй класичній роботі [7] надають фундаментальний огляд концепцій машинного навчання. Вони пропонують критерії для класифікації та порівняння методів машинного навчання, висвітлюючи різні аспекти, такі як представлення знань та алгоритми навчання. Ця стаття є однією з перших, яка систематизувала знання у цій галузі.

У публікації "An Overview of Machine Learning Techniques in Constraint Solving" [8] автори A. Popescu та інші розглядають, як машинне навчання може використовуватися для оптимізації процесів вирішення обмежень. Вони аналізують різні підходи та методи, зосереджуючись на використанні машинного навчання для автоматизації та покращення процесу вирішення задач з обмеженнями.

T.O. Ayodele у своїй праці [9] надає таксономічний аналіз машинного навчання, організовуючи його на основі різних підходів до навчання. Стаття розглядає різні категорії алгоритмів та підходів до машинного навчання, пропонуючи вичерпний огляд методів та їх застосувань у різних галузях.

У статті "An Overview of Privacy in Machine Learning" [10] автор E. De Cristofaro аналізує концепції конфіденційності в машинному навчанні. Ця робота досліджує, як забезпечити конфіденційність даних під час навчання моделей, і пропонує огляд методів, які спрямовані на захист даних. Стаття є важливим внеском у розуміння безпеки та етики в машинному навчанні.

Огляд літератури показує, що машинне навчання є багатогранною дисципліною, яка включає різноманітні методи, підходи та галузі застосування. Від статистичного та логічного навчання до спеціалізованих практичних застосувань, таких як медична діагностика або проєктування антен, машинне навчання демонструє значний потенціал для вирішення складних задач. При цьому важливим є не лише розвиток нових алгоритмів, але й розуміння таких аспектів, як конфіденційність даних, етика та інтеграція машинного навчання у виробничі середовища. Ці публікації допомагають розширити розуміння ключових концепцій машинного навчання та сприяють розвитку цієї

динамічної області, відкриваючи нові перспективи для подальших досліджень та інновацій.

Список використаних джерел:

1. R. Muhamedyev. Machine Learning Methods: An Overview. 2015. https://www.academia.edu/download/54752237/CMNT_2015_Machine_Learning_ns_87art02_ReviewPaper.pdf
2. A.F.A.H. Alnuaimi, T.H.K. Albaldawi. Concepts of Statistical Learning and Classification in Machine Learning: An Overview. 2024. https://www.bio-conferences.org/articles/bioconf/pdf/2024/16/bioconf_iscku2024_00129.pdf
3. P. Flach. Logical Approaches to Machine Learning - An Overview. 1992. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=03b54d20a9d1b68f01c959be0d1266076803f54d
4. S. Albahra, T. Gorbett, S. Robertson, G. D'Aleo. Artificial Intelligence and Machine Learning Overview in Pathology & Laboratory Medicine. 2023. https://www.sciencedirect.com/science/article/pii/S0740257023000138
5. H.M. El Misilmani, T. Naous. Machine Learning in Antenna Design: An Overview on Machine Learning Concept and Algorithms. 2019. https://hilalelmisilmani.wordpress.com/wp-content/uploads/2019/10/hpcs_acme_2019_paper_4.pdf
6. D. Kreuzberger, N. Kühl, S. Hirschl. Machine Learning Operations (MLOps): Overview, Definition, and Architecture. 2023. https://ieeexplore.ieee.org/iel7/6287639/6514899/10081336.pdf
7. J.G. Carbonell, R.S. Michalski, T.M. Mitchell. An Overview of Machine Learning. 1983. https://www.sciencedirect.com/science/article/pii/B9780080510545500054
8. Popescu, S. Polat-Erdeniz, A. Felfernig, M. Uta. An Overview of Machine Learning Techniques in Constraint Solving. 2022. https://link.springer.com/content/pdf/10.1007/s10844-021-00666-5.pdf
9. T.O. Ayodele. Machine Learning Overview. 2010. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=72bfa0a326e4385ac46916e840e9fdc73a98b9fb
10. E. De Cristofaro. An Overview of Privacy in Machine Learning. 2020. https://arxiv.org/pdf/2005.08679