ДОДАТОК А

Перелік джерел посилання науковими напрямами керівника та науковців кафедри Програмної інженерії

2. Nazarenko D. S., Afanasieva I. V., Golian N. V. Neural network approach for emotional recognition in text. Bionics of intelligence. 2019. Т. 1, № 92. С. 9–13. URL: https://doi.org/10.30837/bi.2019.1(92).02 (дата звернення: 29.05.2024).

3. Investigation of the deep learning approaches to classify emotions in texts / D. Nazarenko та ін. CEUR workshop proceedings. 2021. Т. 2870. С. 206–224. (дата звернення: 17.03.2024).

16. Shopynskyi M., Golian N., Afanasieva I. Long short-term memory model appliance for generating music compositions. 2020 IEEE international conference on problems of infocommunications. science and technology (PIC S&T), м. Kharkiv, Ukraine, 6–9 жовт. 2020 р. 2020. URL: https://doi.org/10.1109/picst51311.2020.9468088 (дата звернення: 29.05.2024).

ДОДАТОК Б

Слайди презентації

Рисунок Б.1 – Слайд 1 (тема дослідження)

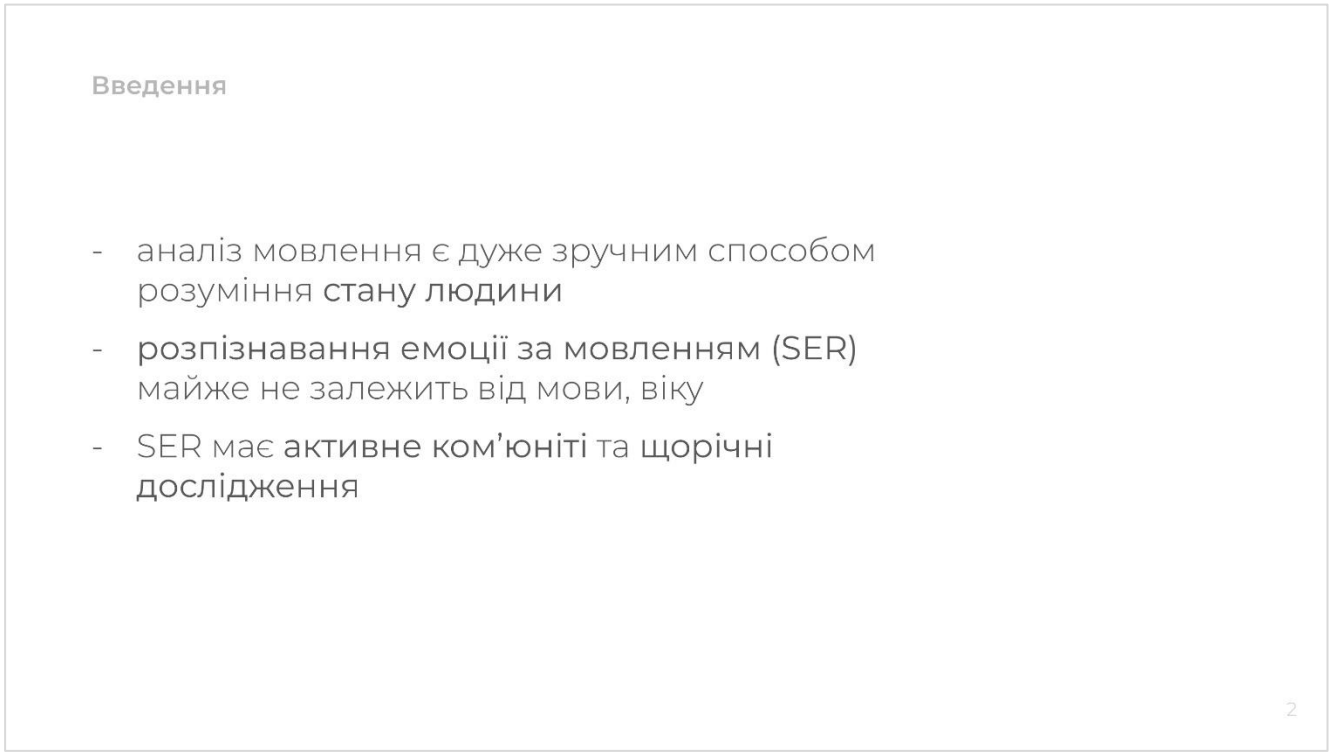


Рисунок Б.2 – Слайд 2 (опис галузі)

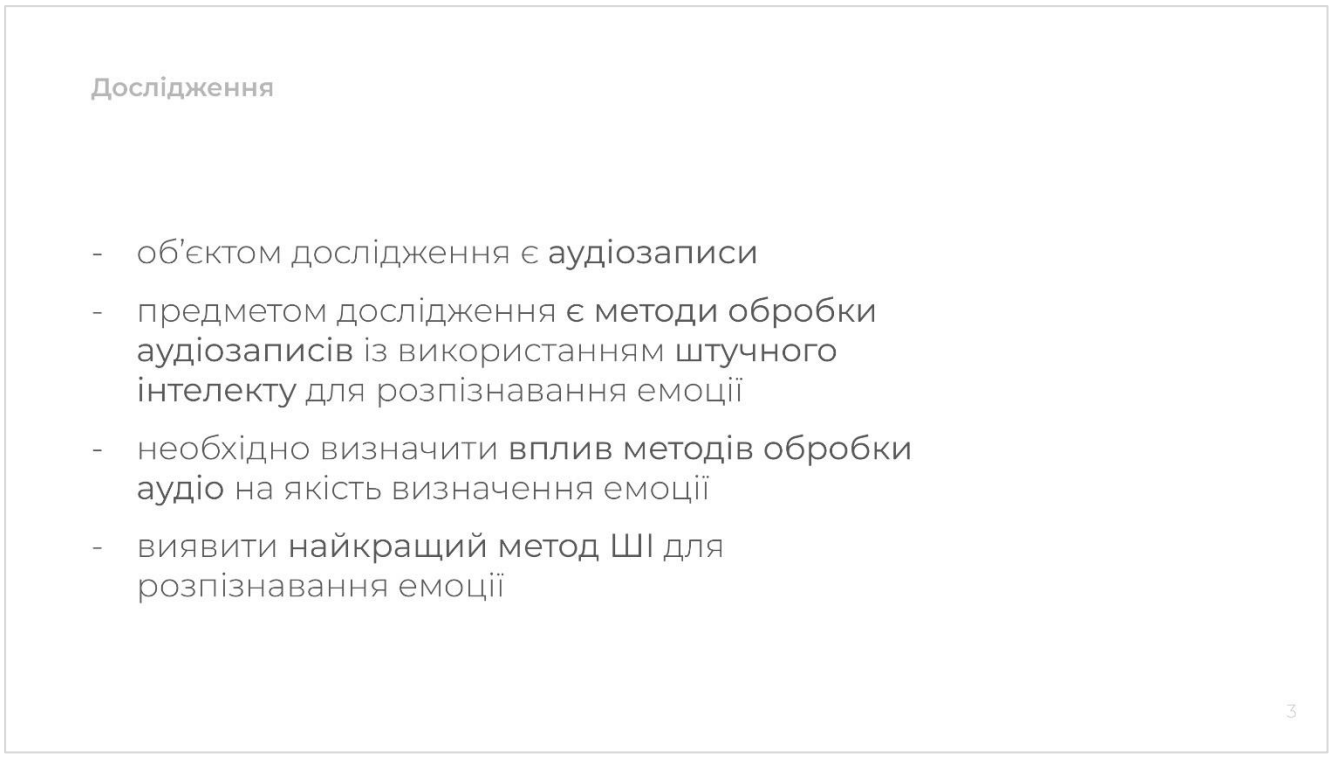


Рисунок Б.3 – Слайд 3 (опис дослідження)

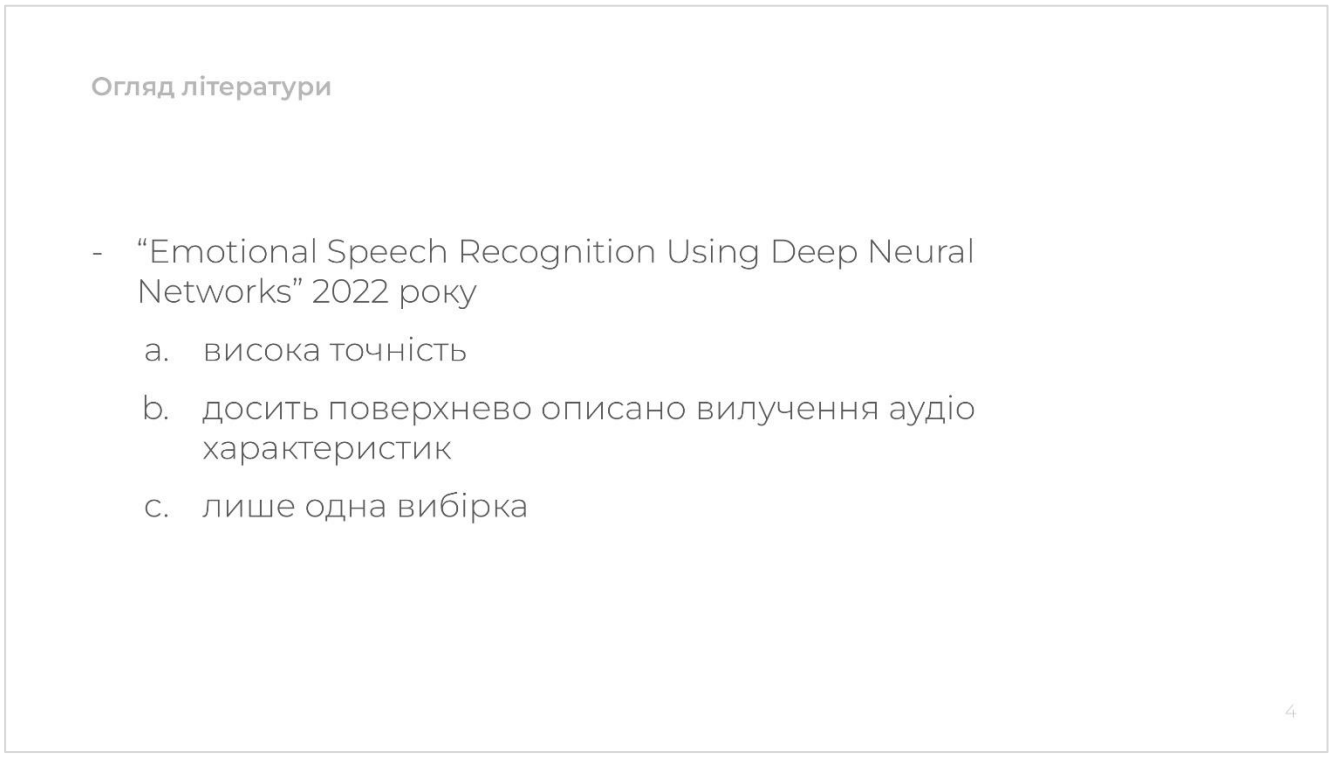


Рисунок Б.4 – Слайд 4 (огляд літератури)

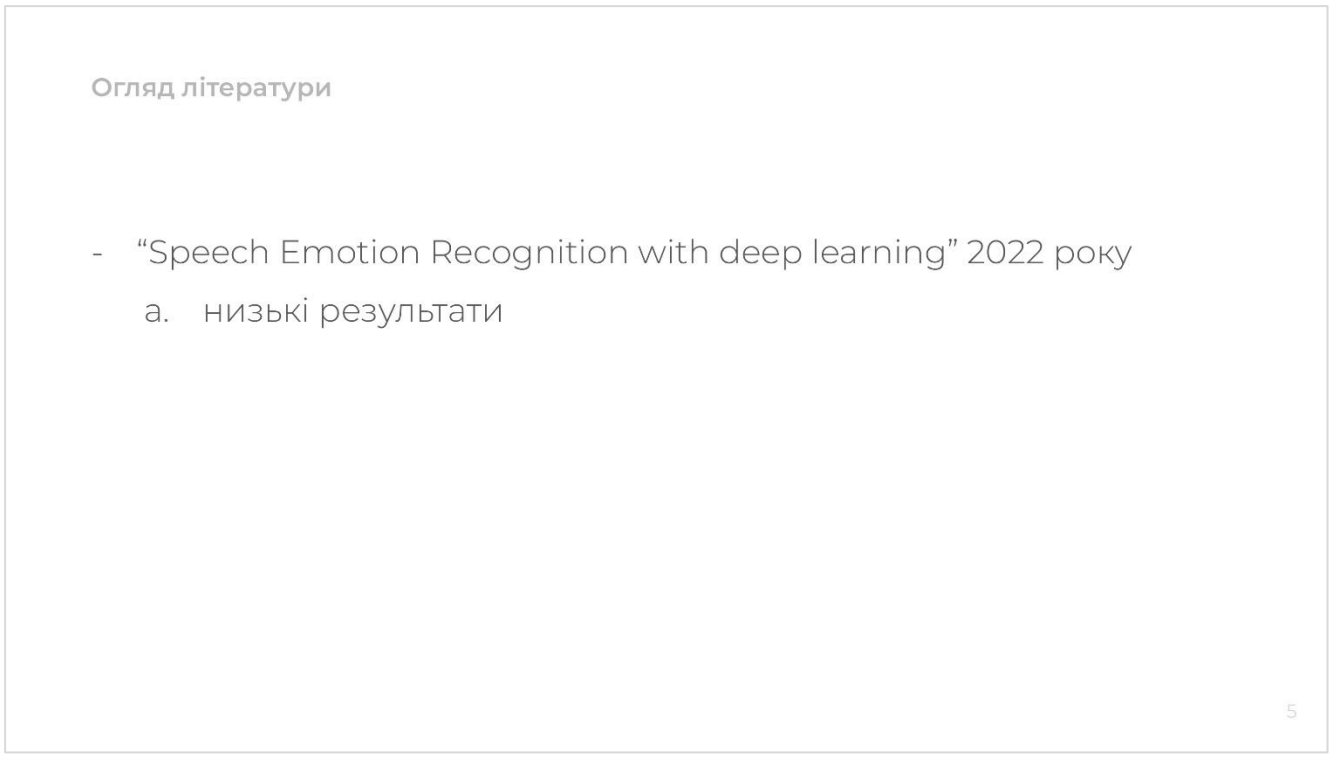


Рисунок Б.5 – Слайд 5 (огляд літератури)

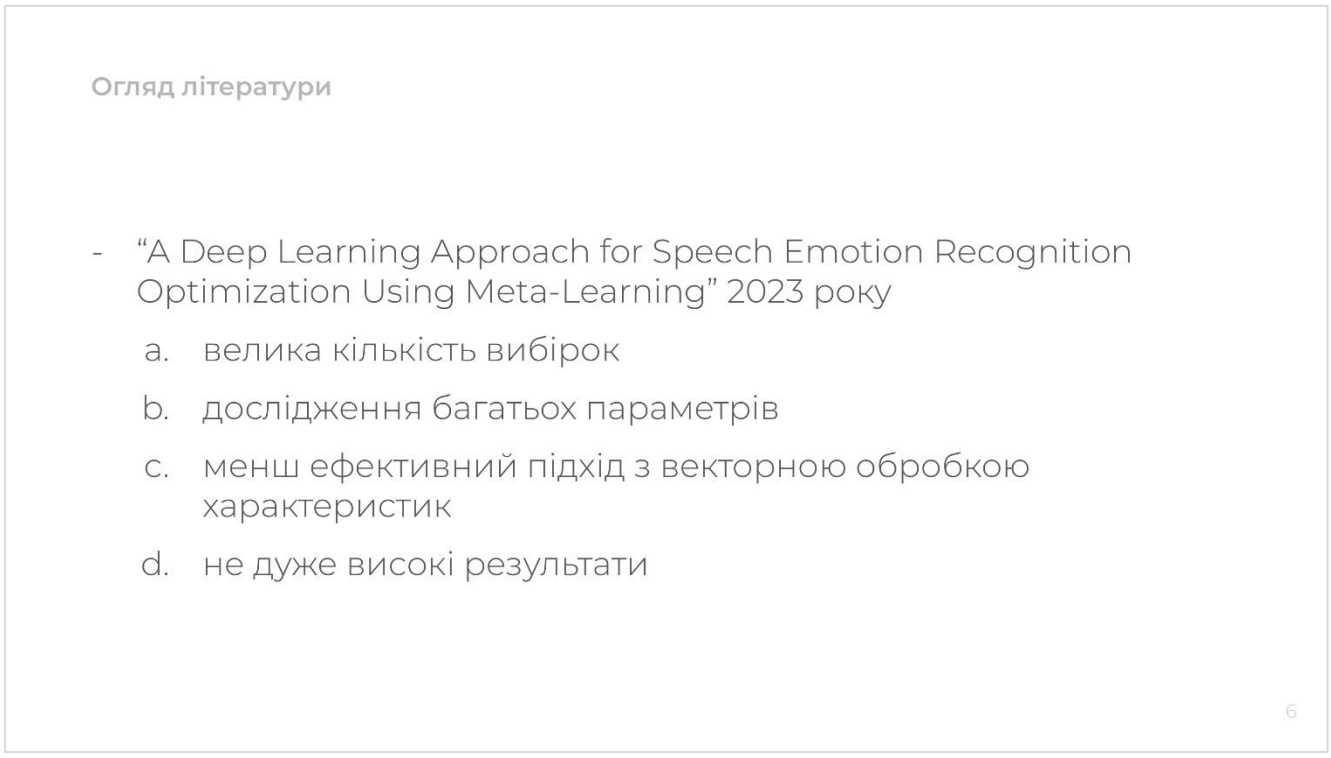


Рисунок Б.6 – Слайд 6 (огляд літератури)

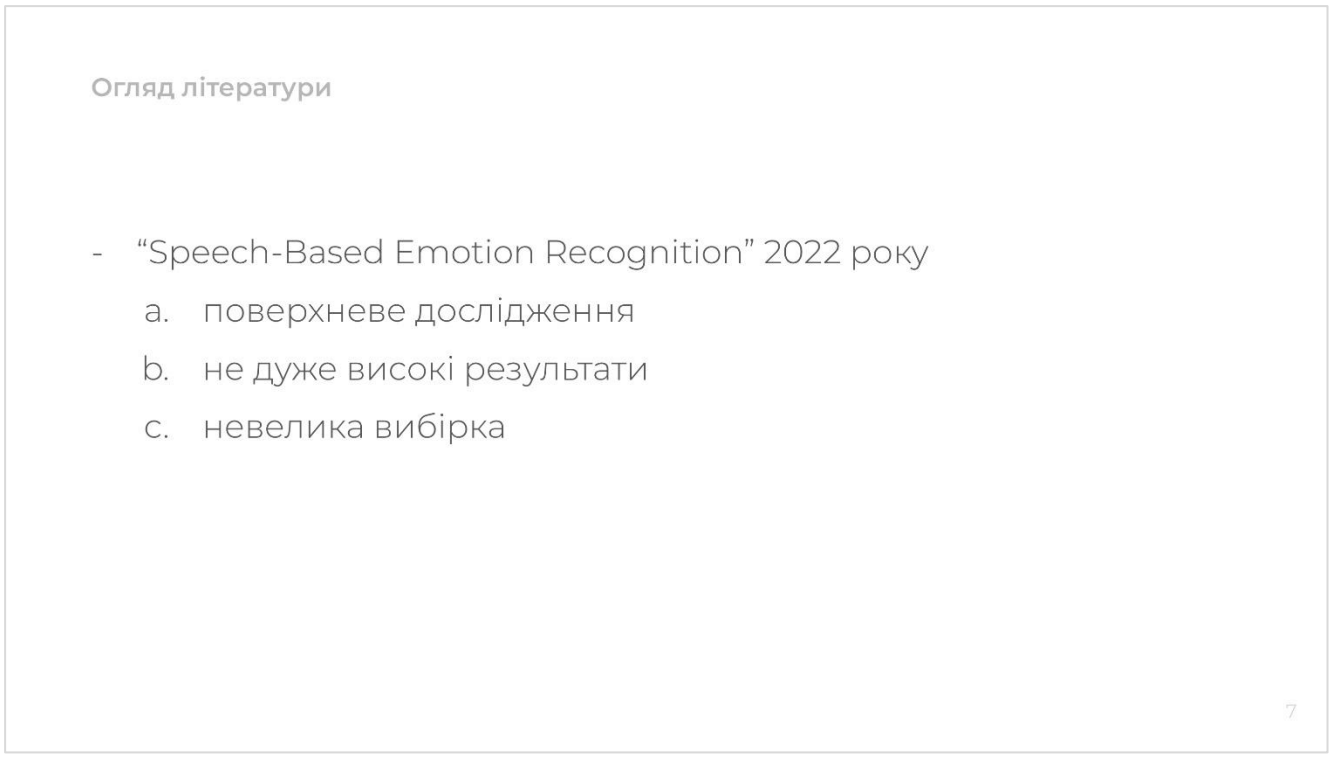


Рисунок Б.7 – Слайд 7 (огляд літератури)

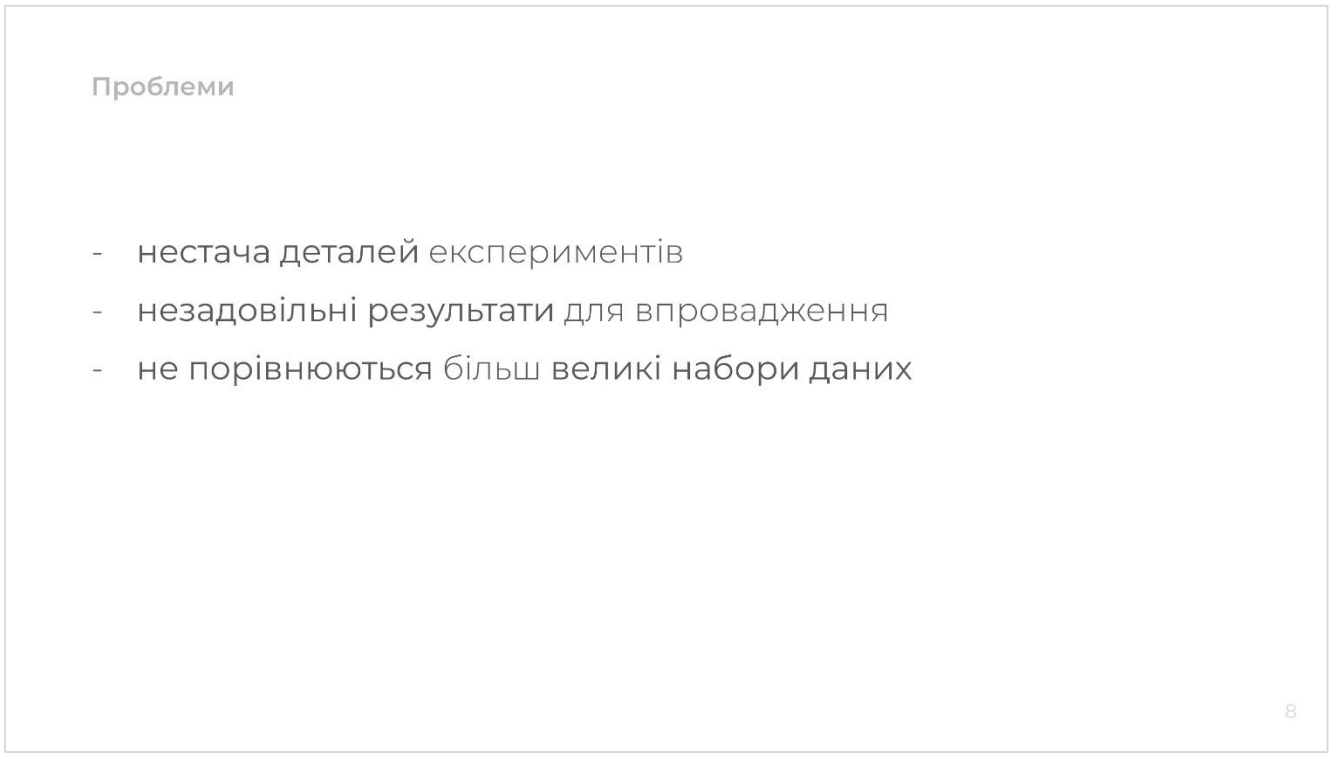


Рисунок Б.8 – Слайд 8 (проблеми галузі)

Рисунок Б.9 – Слайд 9 (задачі дослідження)

Рисунок Б.10 – Слайд 10 (методологія дослідження)

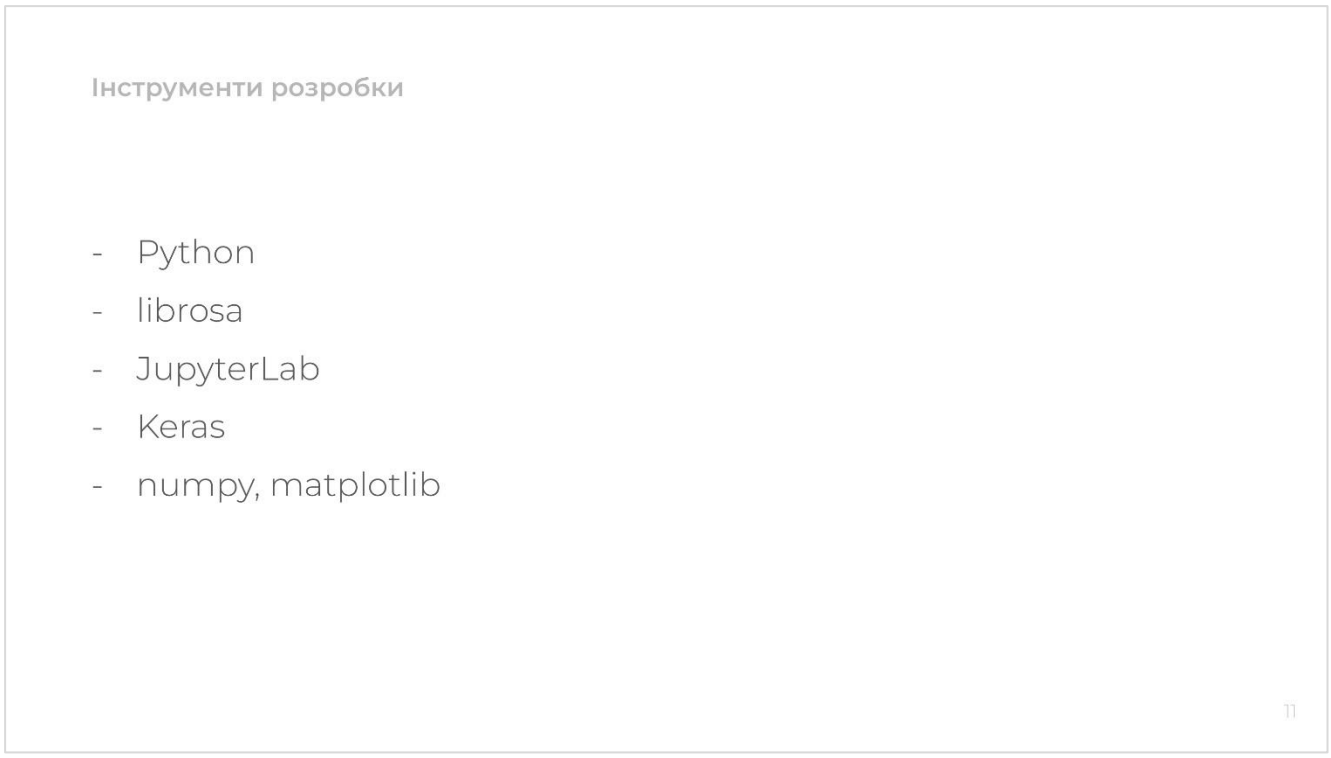


Рисунок Б.11 – Слайд 11 (інструменти розробки)

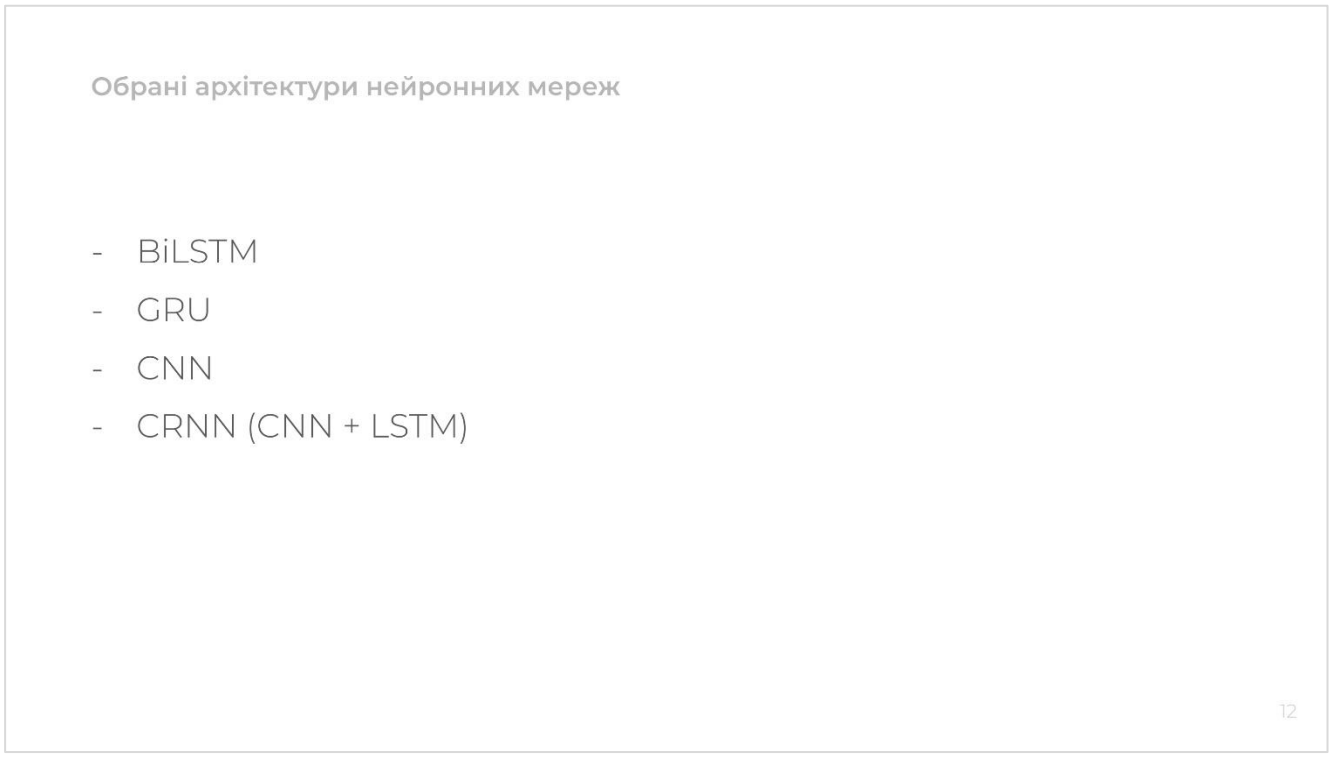
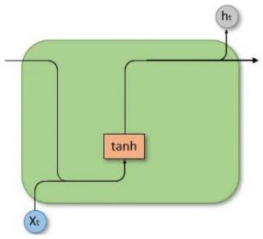


Рисунок Б.12 – Слайд 12 (обрані архітектури нейронних мереж)

RNN

- призначені для роботи з **послідовностями даних**, створюючи цикли для отримання залежностей
- швидко стикаються з проблемою **вибухового** або **зникаючого** градієнтів

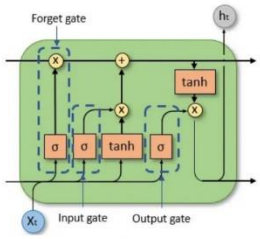


13

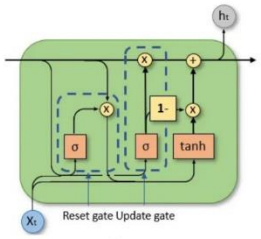
Рисунок Б.13 – Слайд 13 (опис RNN)

LSTM та GRU

- **вдосконалені рекурентні блоки** з так званими «шлюзами»
- краще справляються з проблемами **градієнта**
- можуть мати **двонаправлену** структуру



LSTM



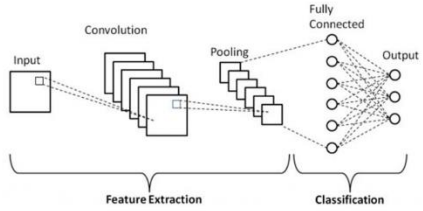
GRU

14

Рисунок Б.14 – Слайд 14 (опис LSTM та GRU)

CNN

- призначені для аналізу матричних/векторних даних
- виділяють патерни за допомогою спеціальних шарів
  a. шари згортки з фільтрами для виділення патерну
  b. шари об'єднання (pooling) для підкреслення патерну



15

Рисунок Б.15 – Слайд 15 (опис CNN)

Набори даних

		SAVEE	RAVDESS	TESS	CREMA-D	IEMOCAP
Total samples		480	1,440	2,800	7,442	10,039
Emotions	Anger	+	+	+	+	+
	Happiness	+	+	+	+	+
	Disgust	+	+	+	+	+
	Fear	+	+	+	+	+
	Sadness	+	+	+	+	+
	Surprise	+	+	+		+
	Neutral	+	+	+	+	+
	Calmness		+			
	Frustration					+
	Excitation					+
Total emotions		7	8	7	6	10
Text variations		15	2	20	12	A lot of
Samples per emotion		~60	195 (96 for neutral)	400	~1,270	Non uniform
Speakers		4 (M)	24 (12 M/12F)	2 (F)	91 (48M/43F)	10 (5M/5F)
Emotion levels		1	2	1	4	A lot of

16

Рисунок Б.16 – Слайд 16 (порівняння наборів даних)

Набори даних

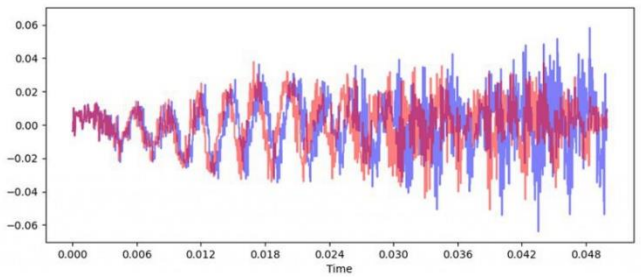
- CREMA-D
  a. рівномірний розподіл за класами
  b. багато акторів і рівнів емоцій
- IEMOCAP
  a. живі розмови
  b. імпровізовані сценарії

17

Рисунок Б.17 – Слайд 17 (обрані набори даних)

Аугментація даних

- додавання шуму
- розтягування у часі
- зміна висоти тону



Підвищення тону (червоний - оригінальний сигнал)

18

Рисунок Б.18 – Слайд 18 (типи аугментації даних)

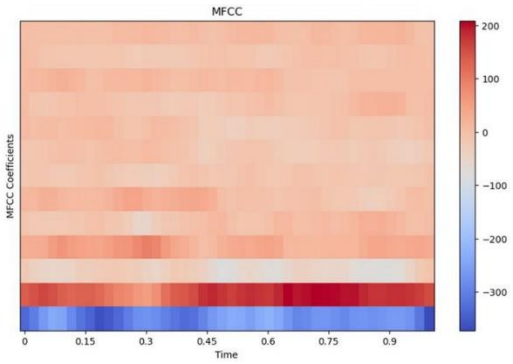
Типи звукових характеристик

- часові (пов’язані з **амплітудою** або силою сигналу)
- частотні (пов’язані з **частотним складом** сигналу)
- часово-частотні (описують зміни **частоти сигналу в часі**)
  a. найбільш інформативні
  b. найчастіше використовуються **MFCCs**

19

Рисунок Б.19 – Слайд 19 (типи звукових характеристик)

MFCCs



Приклад MFCCs у вигляді спектрограми

20

Рисунок Б.20 – Слайд 20 (MFCCs у вигляді спектрограми)

Рисунок Б.21 – Слайд 21 (алгоритм вилучення характеристик)

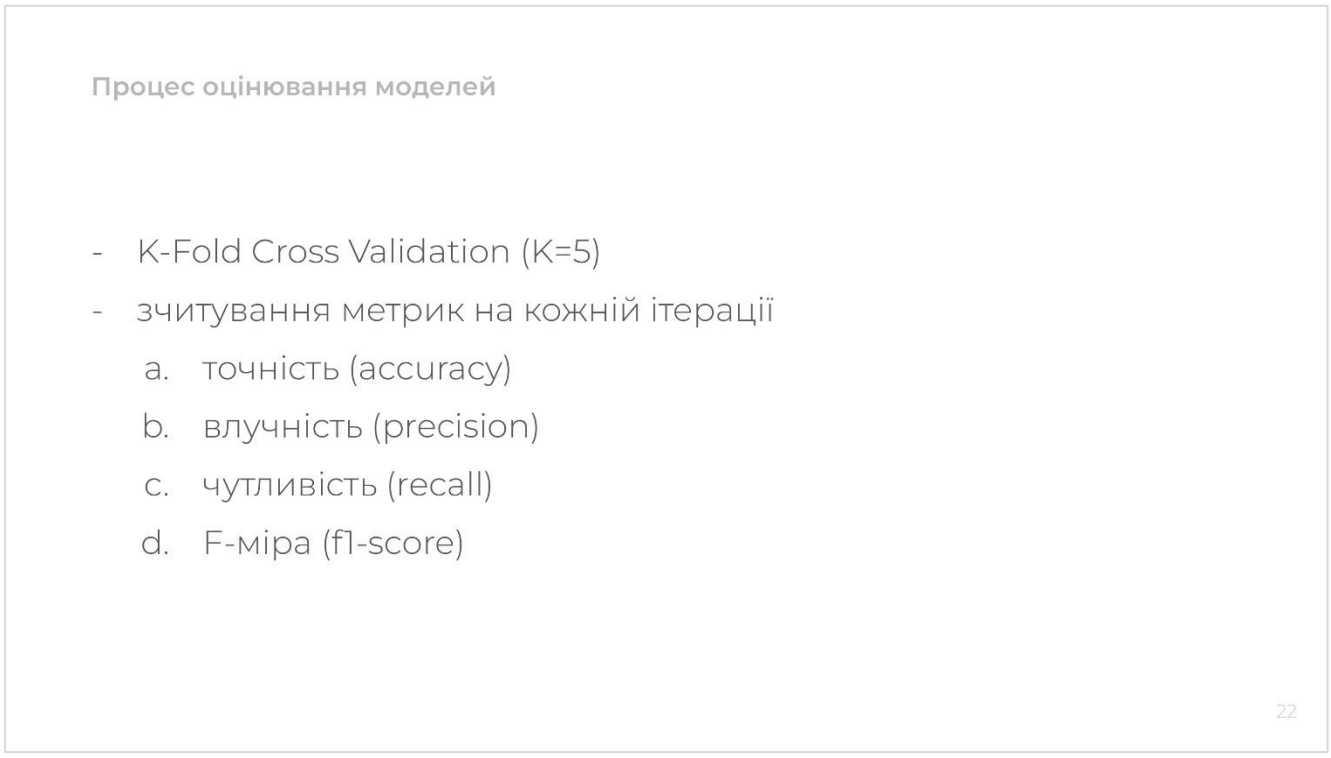


Рисунок Б.22 – Слайд 22 (оцінювання моделей)

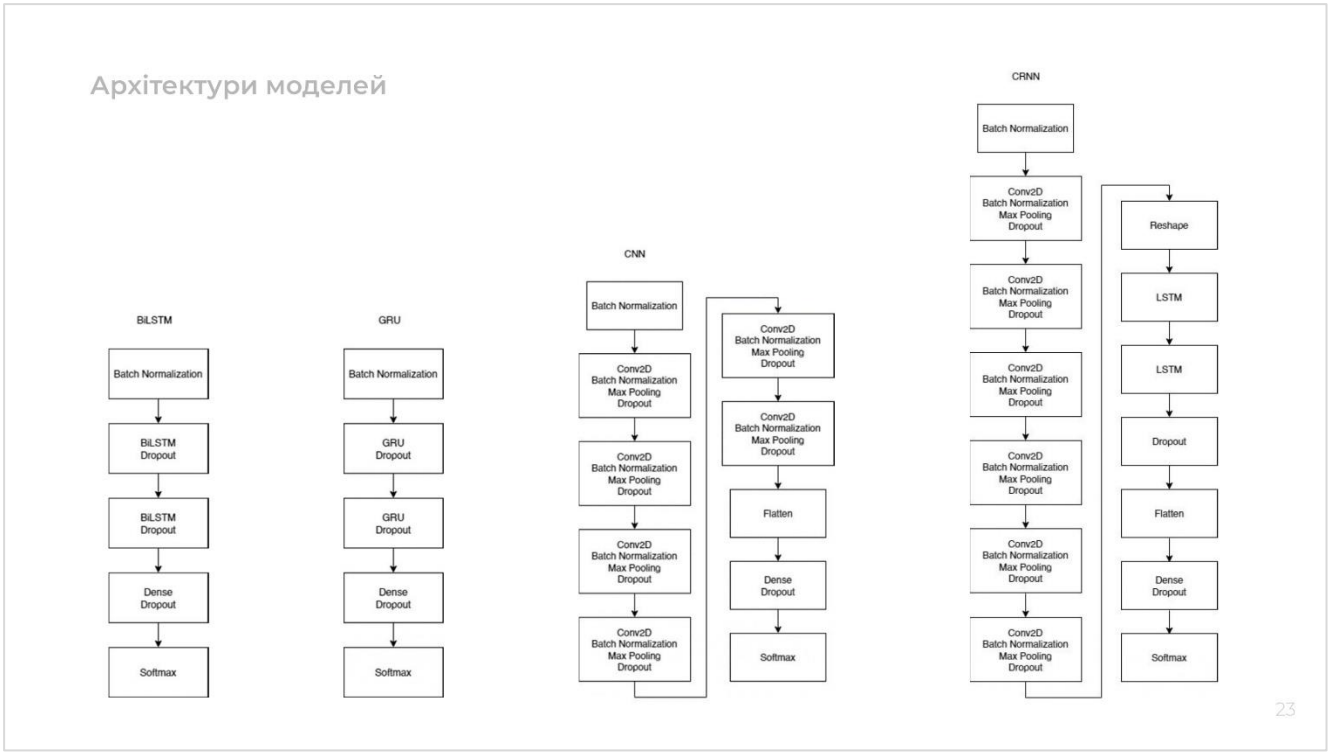


Рисунок Б.23 – Слайд 23 (архітектури моделей)

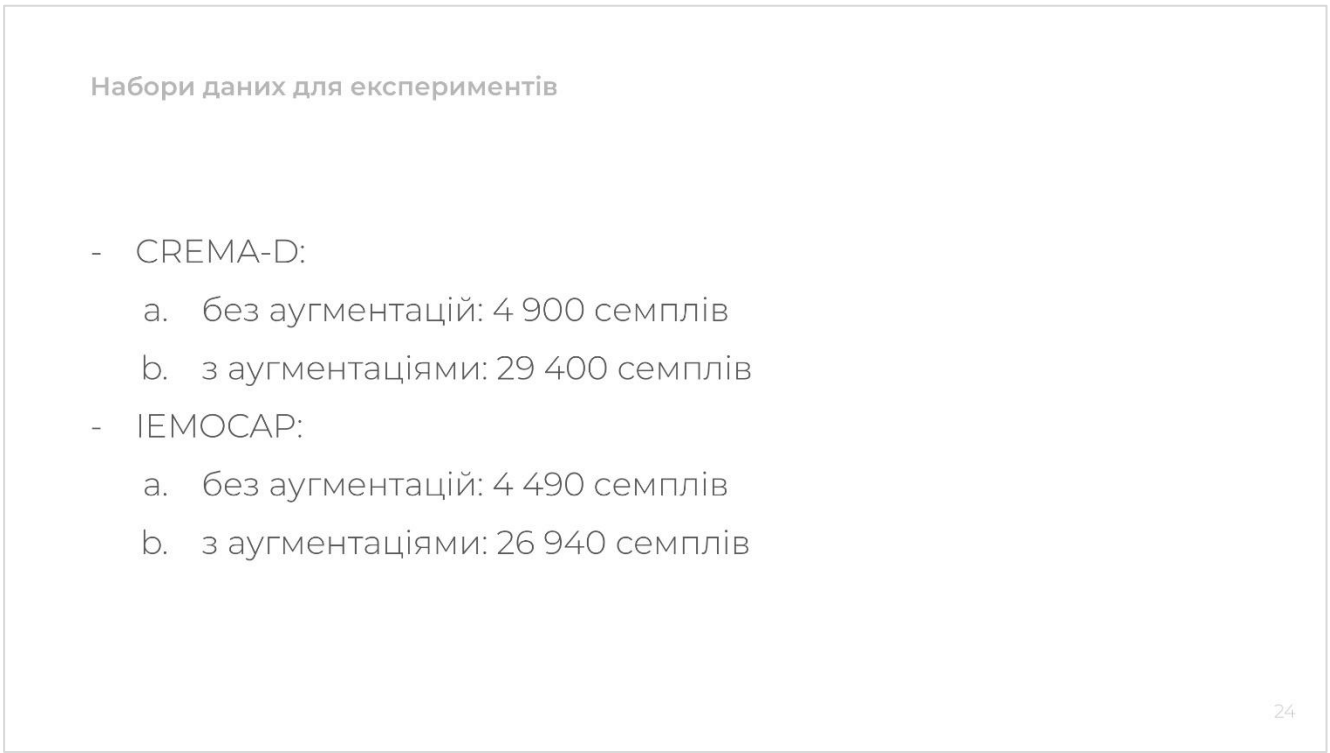


Рисунок Б.24 – Слайд 24 (опис фінальних наборів даних)

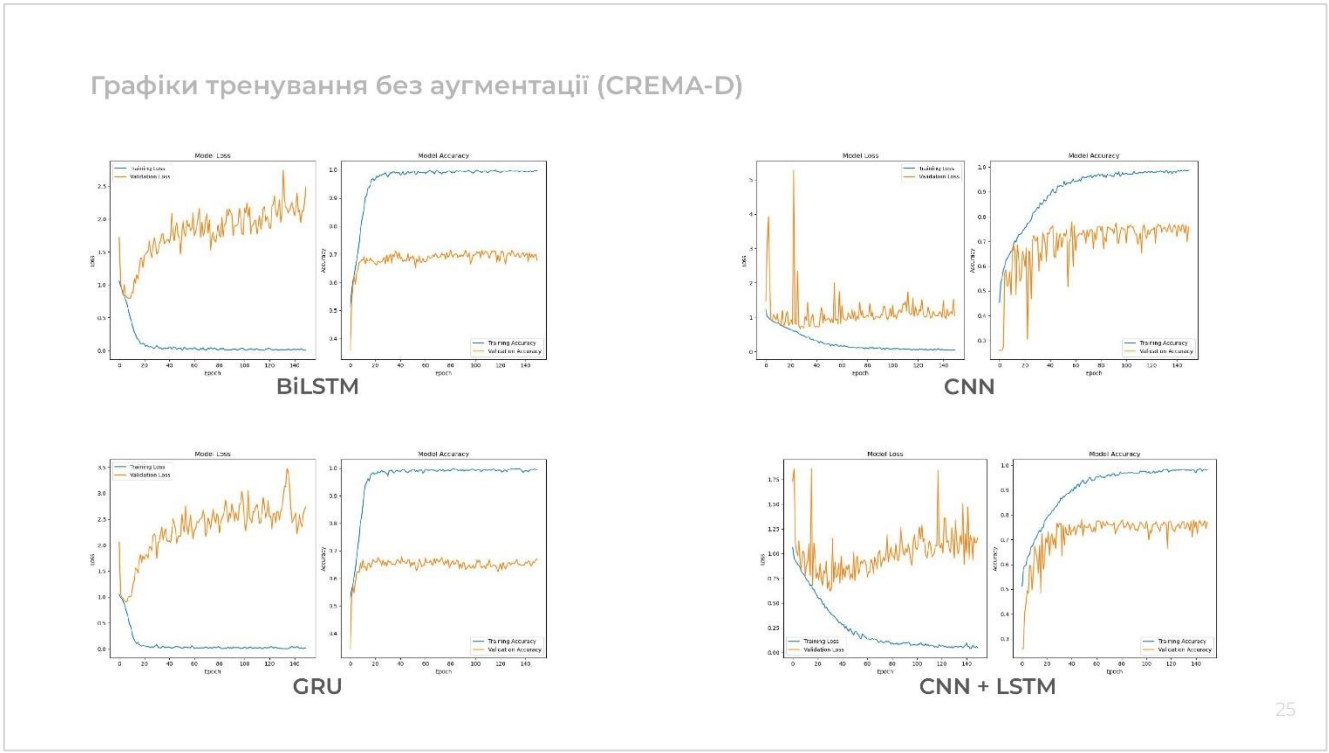


Рисунок Б.25 – Слайд 25 (графіки тренування без агментацій)

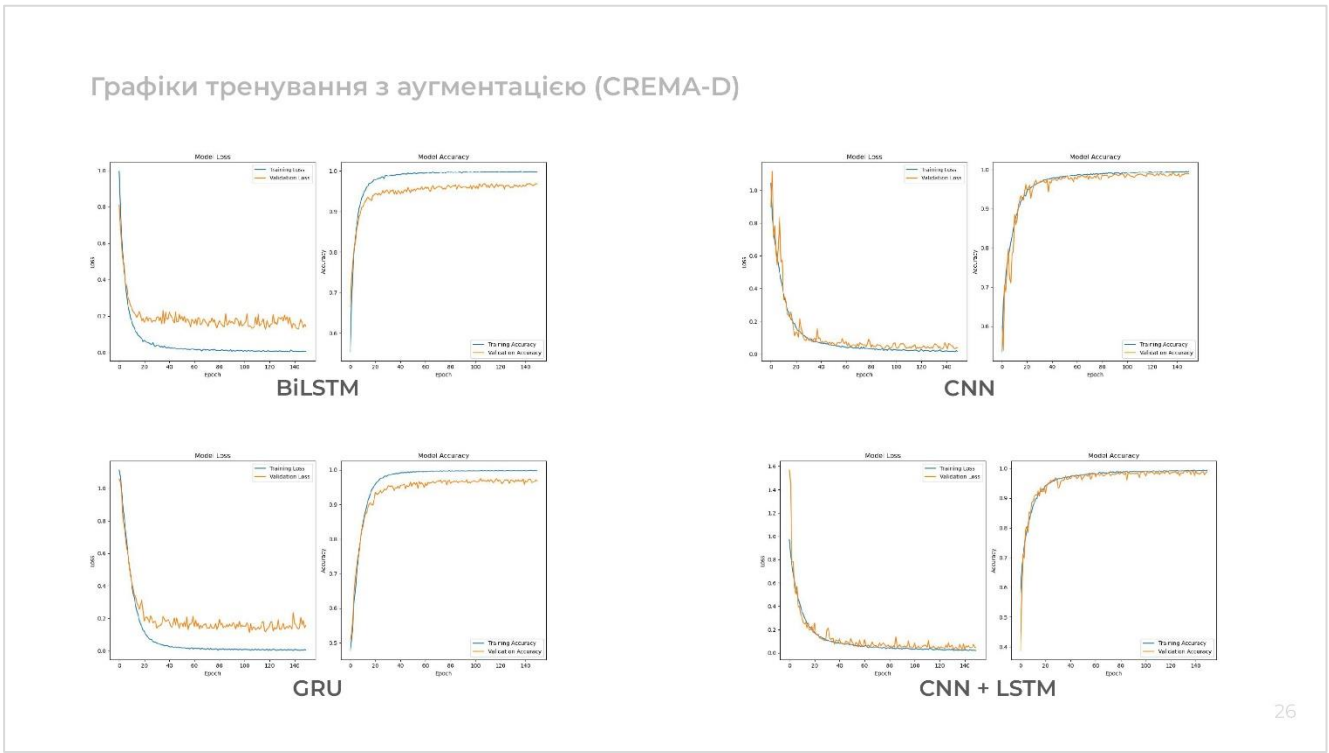


Рисунок Б.26 – Слайд 26 (графіки тренування з аугментаціями)

Результати

Результати навчання моделі BiLSTM на наборі IEMOCAP без аугментації

	precision	recall	f1-score	support
anger	0,701	0,725	0,712	221
happiness	0,346	0,227	0,272	119
neutral	0,637	0,655	0,645	342
sadness	0,632	0,693	0,661	217
accuracy			0,624	
macro avg	0,579	0,575	0,572	898
weighted avg	0,613	0,624	0,616	898

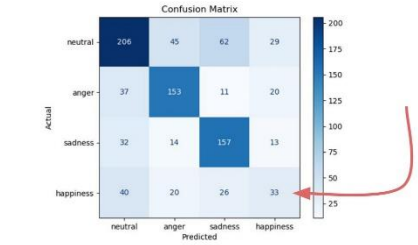
Результати навчання моделі BiLSTM на наборі IEMOCAP з аугментацією

	precision	recall	f1-score	support
anger	0,969	0,962	0,965	1 324
happiness	0,945	0,884	0,913	714
neutral	0,936	0,947	0,942	2 050
sadness	0,927	0,950	0,938	1 301
accuracy			0,943	
macro avg	0,944	0,936	0,940	5 388
weighted avg	0,943	0,943	0,943	5 388

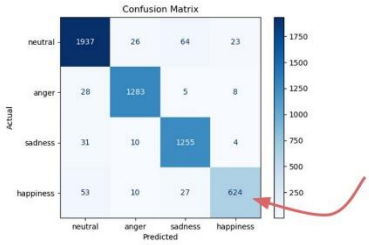
27

Рисунок Б.27 – Слайд 27 (порівняння точності з та без аугментацій)

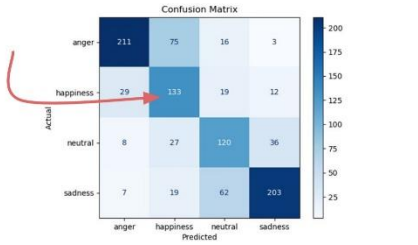
Результати



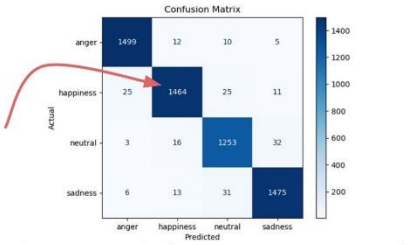
BiLSTM на наборі IEMOCAP без аугментації



BiLSTM на наборі IEMOCAP з аугментацією



BiLSTM на наборі CREMA-D без аугментації



BiLSTM на наборі CREMA-D з аугментацією

28

Рисунок Б.28 – Слайд 28 (матриці помилок для BiLSTM на різних наборах)

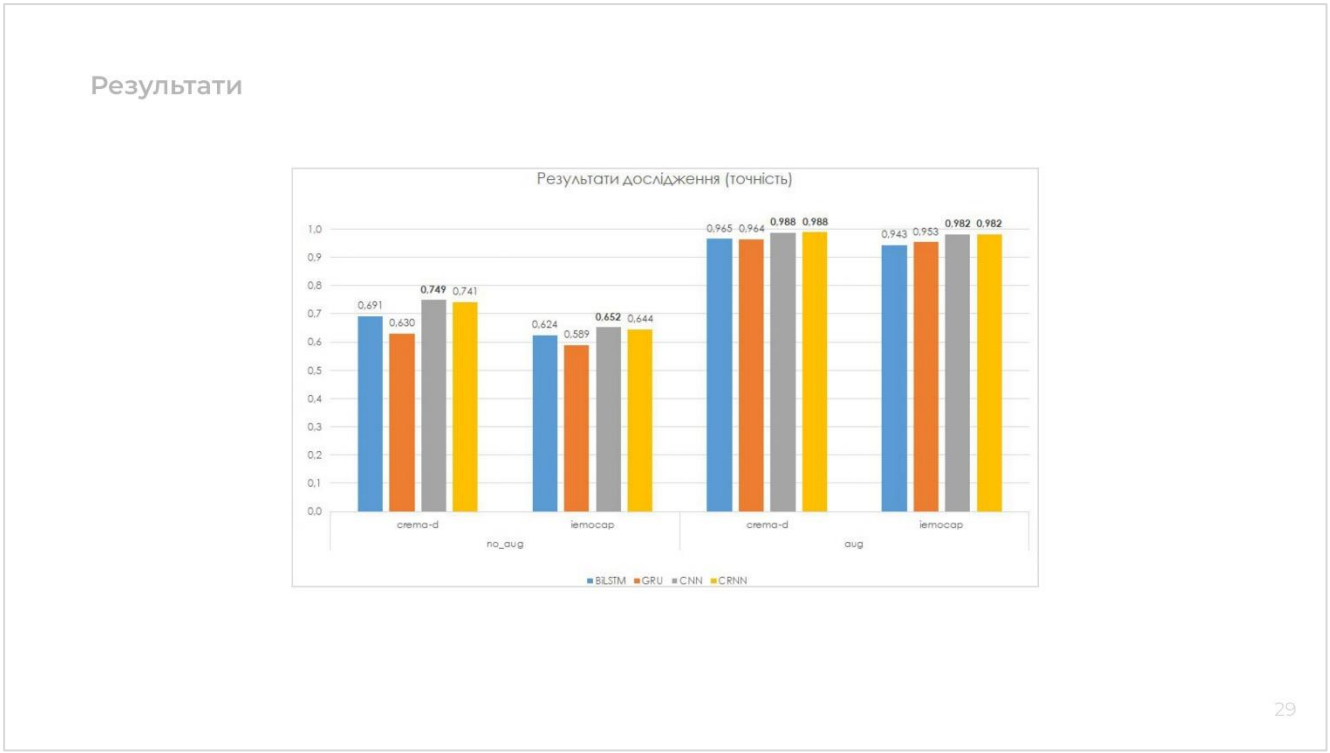


Рисунок Б.29 – Слайд 29 (графік точності моделей на різних наборах даних)

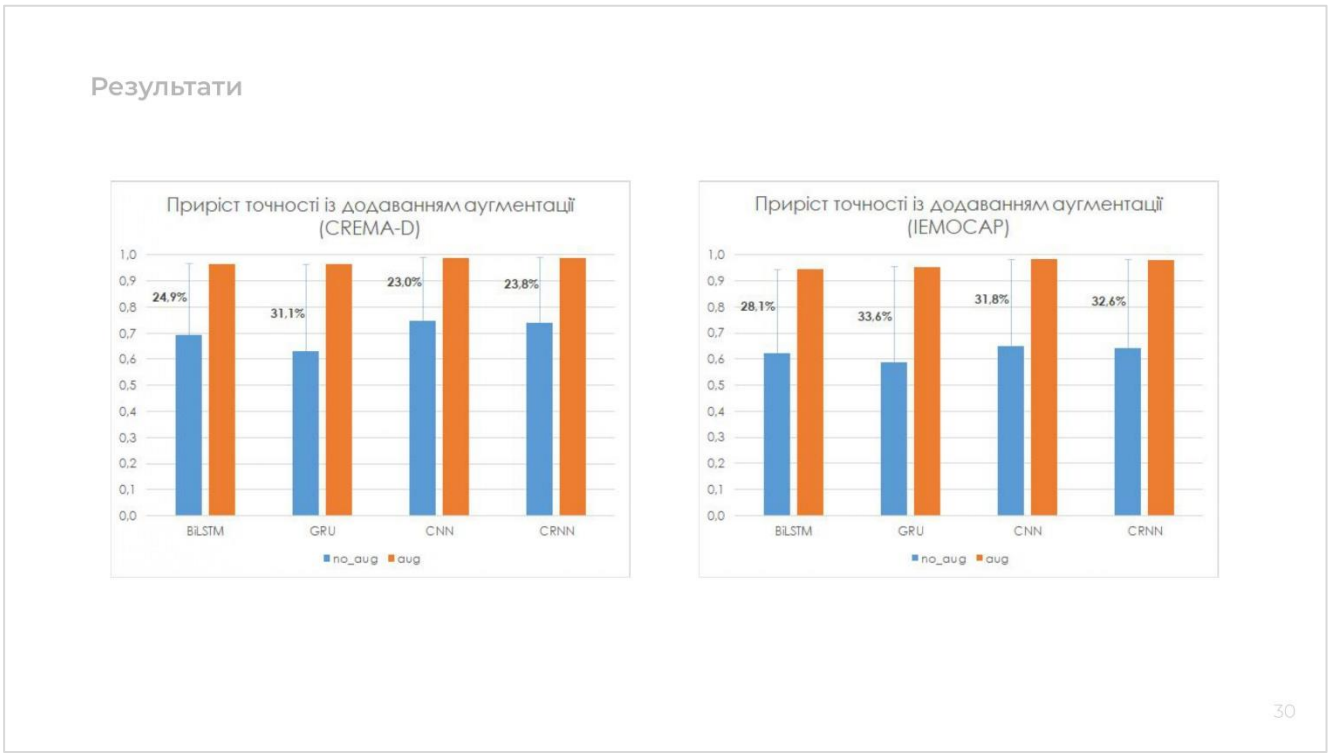


Рисунок Б.30 – Слайд 30 (приріст точності для наборів CREMA-D та IEMOCAP)

Висновки

- **динамічний розмір фрейму** допомагає краще класифікувати
- **MFCCs** є найбільш репрезентативними характеристиками
- актуальними методами ШІ для задачі SER наразі є **нейромережі**
- **CNN** демонструє найякісніші результати
- **CREMA-D** показує вищу ефективність завдяки більш рівномірному розподілу класів
- аугментація даних покращує точність на **25-30%**
- аугментація даних також дозволяє **вирівняти якість** наборів даних з нерівномірним розподілом класів

31

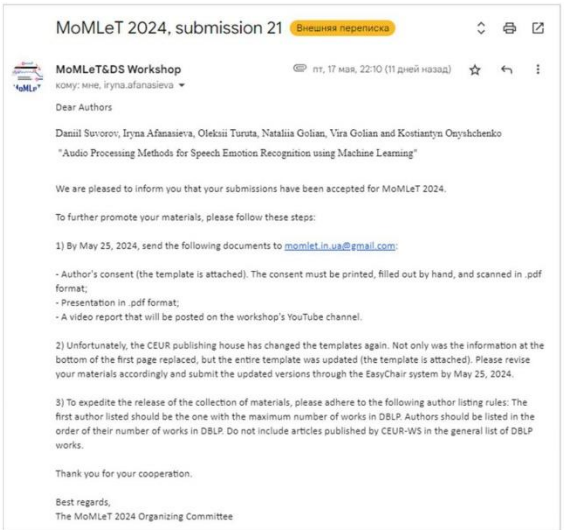
Рисунок Б.31 – Слайд 31 (висновки дослідження)

Публікація результатів

ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ АУДІОЗАПИСІВ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЗНАЧЕННЯ ЕМОЦІЙНОГО СТАНУ

RESEARCH OF AUDIO RECORDING PROCESSING METHODS USING AI TO DETECT EMOTIONAL STATE

Публікація тез у “Znanstvena misel journal”



MoMLeT 2024, submission 21 Внешняя переписка

MoMLeT&DS Workshop пт, 17 мая, 22:10 (11 дней назад)

кому: мне, iryna.afanasieva

Dear Authors

Daniil Suvorov, Iryna Afanasieva, Oleksii Turuta, Nataliia Golian, Vira Golian and Kostiantyn Onyshchenko

"Audio Processing Methods for Speech Emotion Recognition using Machine Learning"

We are pleased to inform you that your submissions have been accepted for MoMLeT 2024.

To further promote your materials, please follow these steps:

1) By May 25, 2024, send the following documents to momlet.in.ua@gmail.com:

- Author's consent (the template is attached). The consent must be printed, filled out by hand, and scanned in .pdf format;
- Presentation in .pdf format;
- A video report that will be posted on the workshop's YouTube channel.

2) Unfortunately, the CEUR publishing house has changed the templates again. Not only was the information at the bottom of the first page replaced, but the entire template was updated (the template is attached). Please revise your materials accordingly and submit the updated versions through the EasyChair system by May 25, 2024.

3) To expedite the release of the collection of materials, please adhere to the following author listing rules: The first author listed should be the one with the maximum number of works in DBLP. Authors should be listed in the order of their number of works in DBLP. Do not include articles published by CEUR-WS in the general list of DBLP works.

Thank you for your cooperation.

Best regards,
The MoMLeT 2024 Organizing Committee

Виступ на конференції “MoMLeT Workshop 2024” із публікацією статті

32

Рисунок Б.32 – Слайд 32 (публікація результатів)

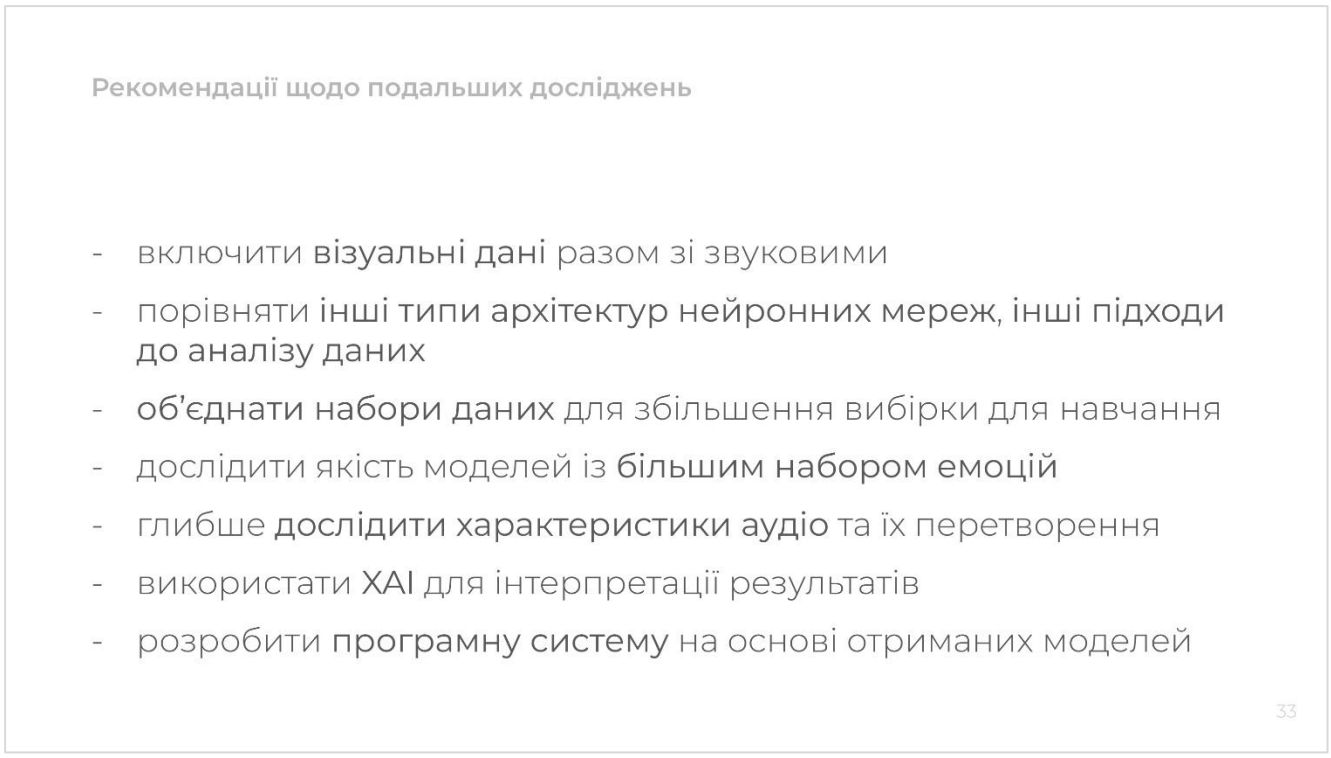


Рисунок Б.33 – Слайд 33 (рекомендації щодо подальших досліджень)

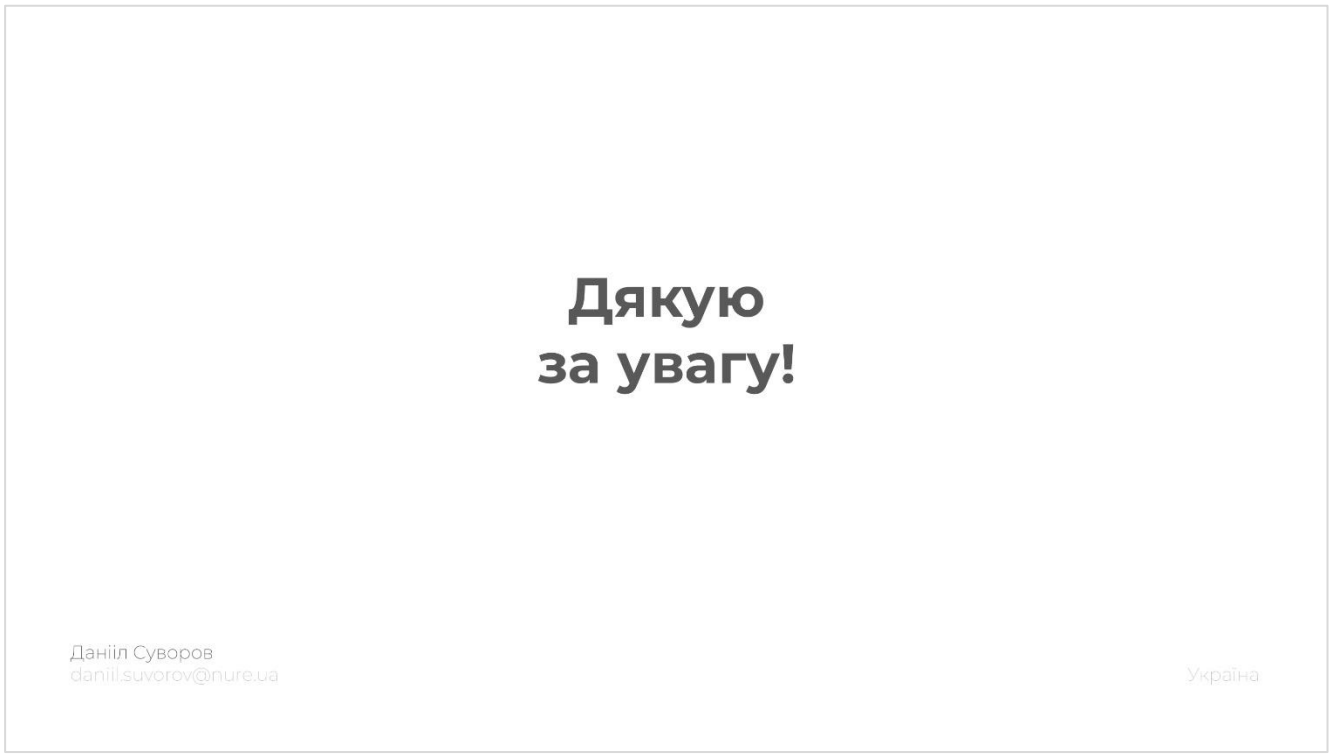
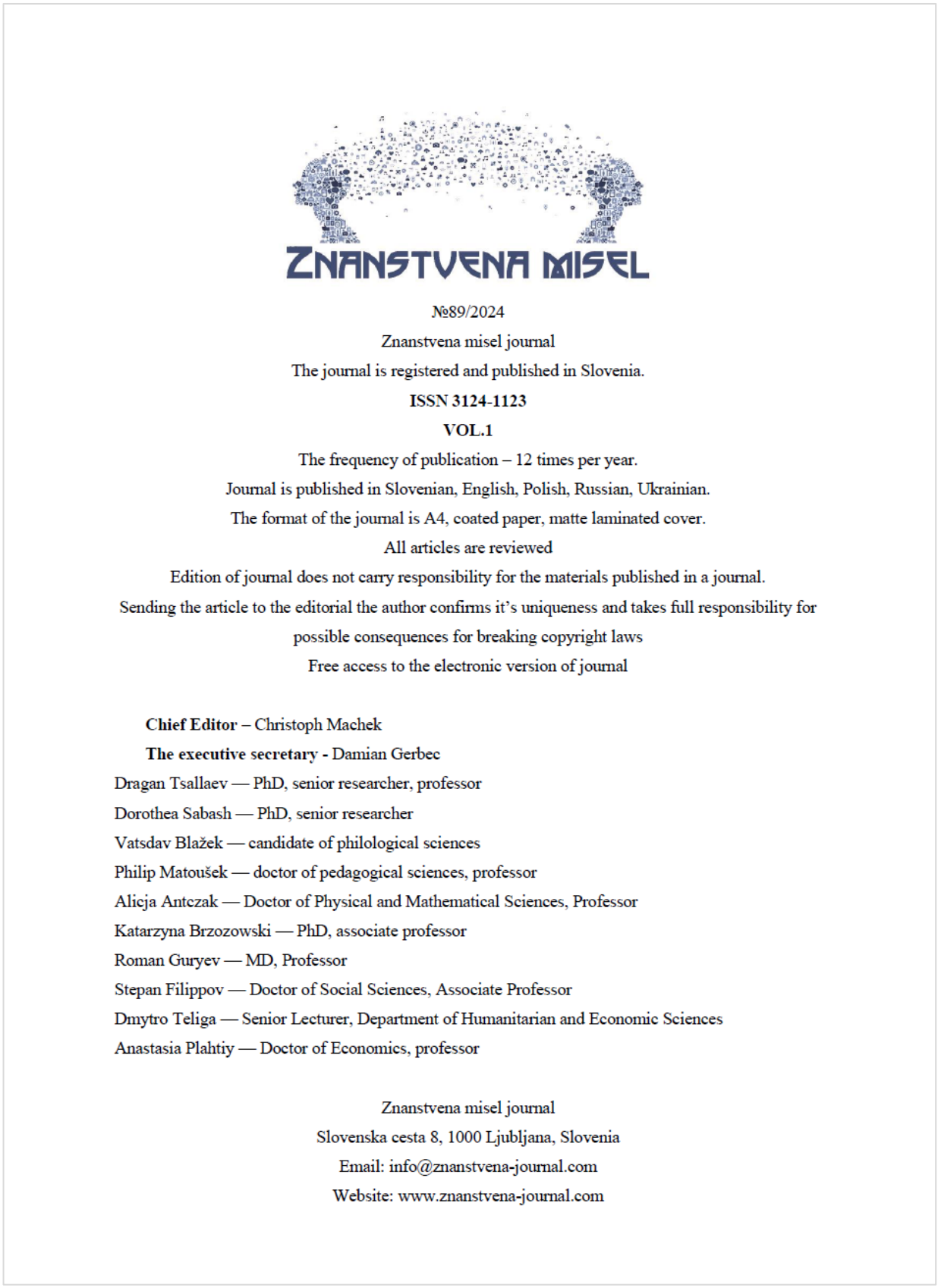


Рисунок Б.34 – Слайд 34 (подяка)

ДОДАТОК В

Апробація у вигляді тез у журналі «Znanstvena misel journal»



ZNANSTVENA MISEL

№89/2024

Znanstvena misel journal

The journal is registered and published in Slovenia.

ISSN 3124-1123

VOL.1

The frequency of publication – 12 times per year.

Journal is published in Slovenian, English, Polish, Russian, Ukrainian.

The format of the journal is A4, coated paper, matte laminated cover.

All articles are reviewed

Edition of journal does not carry responsibility for the materials published in a journal.

Sending the article to the editorial the author confirms it's uniqueness and takes full responsibility for possible consequences for breaking copyright laws

Free access to the electronic version of journal

Znanstvena misel journal

Slovenska cesta 8, 1000 Ljubljana, Slovenia

Email: info@znanstvena-journal.com

Website: www.znanstvena-journal.com

Рисунок В.1 – Титульна сторінка журналу

CONTENT

AGRICULTURAL SCIENCES

Barbaryan A., Ghazaryan R., Alikhanyan N., Nersisyan H., Khachatryan N.
YIELD AND SOWING QUALITIES OF ALFALFA SEEDS UNDER DIFFERENT METHODS OF SOWING AND FERTILIZATION UNDER THE CONDITIONS OF THE ARARAT VALLEY.....3

ARTS

Harutyunyan M.
LEVERAGING GARNI ROYAL BATH MOSAIC AS KEY CULTURAL EVENT ORNAMENTS.....8

CHEMISTRY

Mammadova M., Ibrahimzada S.
ISOMERIZATION OF N-BUTANE WITH THE PARTICIPATION OF CATALYSTS OF SULFATED ZIRCONIUM DIOXIDE.....14

Aliyev S., Mammadzada A.
INFLUENCE OF DEPRESSANT ADDITIVES ON CRUDE OIL FUEL 19

Mamedova N., Nabiyeva N.
SYNTHESIS AND STUDY OF PROPERTIES OF DERIVATIVES OF NATURAL AND SYNTHETIC PETROLEUM ACIDS.....22

EARTH SCIENCES

Danylyan A.
COUTERACTIN THE SHARP RISE OF THE WORLD OCEAN.....27

HISTORICAL SCIENCES

Steblii N., Dovhan P.,
CROSSBODY WEAPONS FROM «MALE GORODYSHCHE» IN BUSK 31

MEDICAL SCIENCES

Grygoryan R.
EXTENDING THE UNDERSTANDING OF HEALTH MECHANISMS: INVERSE RELATIONSHIPS BETWEEN WORSENING OF CELLS' METABOLISM AND ARTERIAL PRESSURE.....35

Yakovets K., Yakovets R., Chornenka Zh.
REMOTE COMPLICATIONS OF OTITIS.....44

PEDAGOGICAL SCIENCES

Jafarova S., Rakhimova L., Takhirova G.
ICT: TRANSFORMİNG EDUCATİON İN THE DİGİTAL AGE.....49

TECHNICAL SCIENCES

Koval R., Yemelianenko S.
FIRE RISK RESEARCH AND MANAGEMENT OF HOTELS 53

Asgarzada S., Namazova M.
DETERMINING THE OPTIMAL PROCESSING VOLUME OF OIL THAT SATISFIES THE DEMAND FOR RAW MATERIALS OF THE PETROCHEMICAL INDUSTRY 57

Suvorov D., Afanasieva I., Onyshchenko K.
RESEARCH OF AUDIO RECORDING PROCESSING METHODS USING AI TO DETECT EMOTIONAL STATE 60

Рисунок В.2 – Зміст журналу

ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ АУДІОЗАПИСІВ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЗНАЧЕННЯ ЕМОЦІЙНОГО СТАНУ

Суворов Д.С.
Харківський національний університет радіо електроніки, студент
Афанасьєва І.В.
Харківський національний університет радіо електроніки, кандидат технічних наук, доцент
Онищенко К.Г.
Харківський національний університет радіо електроніки, старший викладач

RESEARCH OF AUDIO RECORDING PROCESSING METHODS USING AI TO DETECT EMOTIONAL STATE

Suvorov D.,
Kharkiv National University of Radio Electronics, student
Afanasieva I.,
Kharkiv National University of Radio Electronics, Candidate of Technical Sciences, Associate Professor
Onyshchenko K.
Kharkiv National University of Radio Electronics, Senior Lecturer
DOI: 10.5281/zenodo.11049575

Анотація

Наразі існує задача розпізнавання емоційного стану людини за його мовленням (звуковому наданні). Це може бути застосовано у медичних цілях, правоохоронних органах, системах розумних будинків та подібних. Сучасні технології дозволяють швидко та досить точно вирішувати подібні задачі із використанням штучного інтелекту, зокрема нейронних мереж згорткового та рекурентного типів. Побудова власних мереж цих типів дозволила досягти точності розпізнавання емоційного стану вище 95% із використанням набору даних CREMA-D. У статті також розглянуто позитивний вплив аугментації даних для підвищення точності розпізнавання.

Abstract

Currently, there is a task of recognizing a person's emotional state from their speech (audio representation). This can be used for medical purposes, law enforcement, smart home systems, and so on. Modern technologies make it possible to solve such tasks quickly and accurately enough using artificial intelligence, in particular convolutional and recurrent neural networks. Building our own networks of these types allowed us to achieve an accuracy of emotional state recognition above 95% using the CREMA-D dataset. The article also discusses the positive impact of data augmentation to improve recognition accuracy.

Ключові слова: аудіо, емоції, машинне навчання, мовлення, нейронні мережі, розпізнавання, штучний інтелект, python, tensorflow.

Keywords: audio, emotions, machine learning, speech, neural networks, recognition, artificial intelligance, python, tensorflow.

Поточне дослідження можна віднести до дослідження, яке відомо у світі машинного навчання як Speech Emotion Recognition (SER). Ця задача має вже достатньо досліджень із використанням різних методів штучного інтелекту, які можна проаналізувати перед проведенням власних експериментів.

Блог dataiku [1] поділився своїми результатами у SER із використанням звичайної повнозв'язної мережі. Результати тестування прибизно 60%.

Наступна робота [2] використовувала SVM (метод опорних векторів) – для задачі SER. Використовувався аудіовізуальний набір даних RML, який містить 720 екземплярів даних на кількох мовах 6 людських емоцій. Результати розпізнавання більше за попередні – близько 74%.

Наступна стаття [3] комплексно розглядає різні методи нейронних мереж та набори даних. Загалом, із використанням того ж набору даних CREMA-D вони змогли отримати точність близько 83%.

Найбільш ефективними серед проаналізованих робіт методами виявилися наступні:

- згорткові нейронні мережі;
- рекурентні нейронні мережі.

CNN (згорткова нейрона мережа) [4] – це тип глибокої нейронної мережі, спеціально розроблений для обробки та аналізу структурованих матриць даних, таких як зображення (хоча часто такі мережі використовуються для зображень, вони також є ефективні і для обробки аудіо, оскільки аудіо сигнал можна зобразити графічно, наприклад спектрограмою).

RNN (рекурентна нейрона мереже) [5] – це клас нейронних мереж, призначений для роботи з послідовностями даних, де інформація передається в часі. Основна ідея полягає в тому, щоб в мережі були зв'язки, які створюють цикли, дозволяючи інформації з попередніх кроків часу впливати на поточний стан мережі.

CREMA-D (Crowd-sourced Emotional Mutimodal Actors Dataset) [6] – це набір даних, створений для вивчення та розробки систем визначення.

Рисунок В.3 – Перша сторінка тез

Цей набір даних містить досить рівномірний розподіл екземплярів класів (1 270 для кожної з 6 емоцій), а також 91 актор зачитує 12 варіантів речення, що робить цей набір один з найбільш репрезентативних.

Незважаючи на те, що типів характеристик (features), які можна використовувати як вхідні дані до нейронної мережі, дуже багато і кожна певним чином характеризує аудіо, практика показує, що найбільш вдало, повно та якісно для обробки аудіо нейронними мережами працюють саме MFCC (Mel-frequency cepstral coefficients). Вони мають достатньо інформації для досить точного визначення, у нашому випадку, емоційного стану.

Відомим та зручним інструментом для отримання цих характеристик є бібліотека на мові програмування Python – librosa [7]. Вона дозволяє легко маніпулювати аудіо сигналами, обробляти їх та вилучати характеристики для їх подальшого використання у навчання нейронній мережі.

У машинному навчанні широко використовується така техніка обробки даних як аугментація. Ця техніка передбачає створення нових прикладів даних шляхом застосування різних операцій трансформації до існуючих зразків. Ця техніка застосовується з метою розширення обсягу тренувального набору та покращення загальної здатності моделі до узагальнення на нові, реальні дані.

Серед цілей аугментації можна виділити наступні:

- дозволяє моделі бачити більше різноманітності в тренувальному наборі, що може допомогти уникнути перенавчання та поліпшити узагальнювальні властивості моделі;
- додавання різних варіацій до даних допомагає моделі впоратися з різними умовами та вхідними даними;
- застосування аугментації може допомогти зробити модель менш чутливою до змін в умовах зйомки або в реальних сценаріях.

Для аугментації були обрані методи додавання шуму до аудіо, розтягування у часі (довше та коротше) та зміна тону (вище та нижче). Тобто аугментована вибірка має у 4 рази більше наборів у порівнянні з оригінальною.

Для дослідження було обрано 4 емоції: anger, happiness, sadness, neutral.

Першою нейронною мережею був різновид рекурентних мереж – GRU (Gated Recurrent Units). Він має більш складну структуру у порівнянні звичайної RNN, що дозволяє такій мережі запам'ятовувати більше зв'язків між даними, розподіленими у часі.

Власне графічна репрезентація архітектури мережі із GRU представлена на рисунку 1.

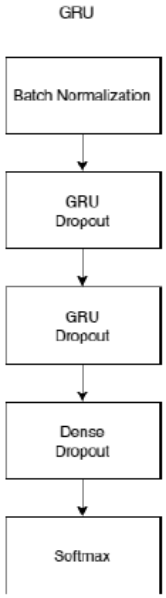


Рисунок 1 – Архітектура мережі із GRU

Модель складається з шару batch normalization (нормалізує дані та дозволяє пришвидшити навчання мережі), два шари GRU (256 та 512 рекурентних блоків відповідно із функцією активації tanh) із шарами Dropout, які дозволяють відключати певні блоки для запобігання перенавчанню моделі. Далі іде звичайний повнозв'язний шар із 128 нейронами та шар із 4 нейронами (за кількістю класів) та функцією активації softmax для безпосередньо активації.

Результати навчання моделі зображені у таблицях 1 та 2. Метрика точності (accuracy) показує, що аугментація має досить серйозний вплив на об'єктивність моделі та її здатність у розпізнавання. 63% та 96,4% без та з аугментаціями відповідно. Можна говорити, що GRU з аугментаціями є досить гарною моделлю з високою точністю і здатністю у подальшому застосуванні.

Рисунок В.4 – Друга сторінка тез

Таблиця 1

Результати навчання моделі GRU без аугментації

	precision	recall	f1-score	support
anger	0,766	0,752	0,757	254
happiness	0,621	0,603	0,608	254
neutral	0,480	0,522	0,498	217
sadness	0,655	0,625	0,638	254
accuracy			0,630	
macro avg	0,630	0,626	0,625	980
weighted avg	0,636	0,630	0,630	980

Таблиця 2

Результати навчання моделі GRU з аугментацією

	precision	recall	f1-score	support
anger	0,982	0,977	0,980	1 525
happiness	0,971	0,964	0,967	1 525
neutral	0,937	0,957	0,947	1 304
sadness	0,963	0,957	0,960	1 525
accuracy			0,964	
macro avg	0,963	0,964	0,963	5 880
weighted avg	0,964	0,964	0,964	5 880

Архітектура згорткової мережі дещо більш складана , зображена на рисунку 2.

Першим шаром іде також шар з нормалізацією. Далі розташовані 6 так званих згорткових блокі, кожен з яких містить:

- conv 2D (власне шар згортання);
- batch normalization;
- max pooling 2D (шар для виокремлення домінуючих ознак з даних);
- dropout.

Далі іде шар Flatten для перетворення багатовимірного масиву даних у вектор значень. Останніми є повнозв'язний шар та шар із softmax, так саме, як і для GRU.

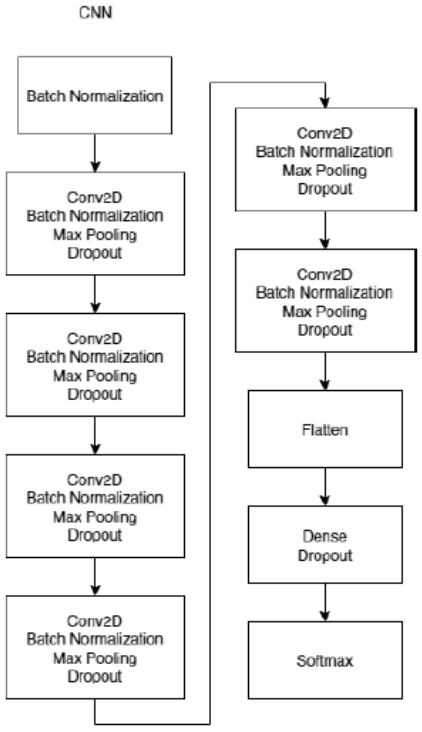


Рисунок 2 – Архітектура згорткової мережі

Рисунок В.5 – Третя сторінка тез

Результати навчання зображені у таблицях 3 та 4 без та з аугментаціями відповідно.

Таблиця 3

Результати навчання моделі CNN без аугментації

	precision	recall	f1-score	support
anger	0,828	0,789	0,803	254
happiness	0,727	0,710	0,718	254
neutral	0,678	0,732	0,698	217
sadness	0,799	0,762	0,771	254
accuracy			0,749	
macro avg	0,758	0,748	0,747	980
weighted avg	0,761	0,749	0,749	980

Одразу видно, що для обох випадків значення точності моделі згорткової вище за значенням моделі із GRU блоками. Також прослідковується значне збільшення точності із використанням аугментацій для набору даних. 75% та 98,8% точність для згорткової мережі без та з аугментаціями відповідно.

Таблиця 4

Результати навчання моделі CNN з аугментаціями

	precision	recall	f1-score	support
anger	0,998	0,992	0,995	1 525
happiness	0,992	0,987	0,989	1 525
neutral	0,977	0,986	0,981	1 304
sadness	0,985	0,987	0,986	1 525
accuracy			0,988	
macro avg	0,988	0,988	0,988	5 880
weighted avg	0,988	0,988	0,988	5 880

В ході низки експерементальних досліджень на наборі даних CREMA-D із аугментацією та без на двох моделях нейромереж (GRU та CNN) було виявлено, що найбільш прийнятною для розпізнавання емоції за аудіо у використаному підході є модель, що використовую згорткові шари. Така модель видає точність більше 98%. Також було визначено, що аугментація даних вкрай необхідна для отримання якісної моделі.

Список літератури

1. Speech Emotion Recognition Using Deep Learning. Blog - Dataiku. URL: https://blog.dataiku.com/speech-emotion-recognition-deep-learning (дата звернення: 30.12.2023).

2. Speech Emotion Recognition with deep learning. ScienceDirect. URL: https://www.sciencedirect.com/science/article/pii/S1877050920318512 (дата звернення: 30.12.2023).

3. A Deep Learning Approach for Speech Emotion Recognition Optimization Using Meta-Learning. MDPI. URL: https://www.mdpi.com/2079-9292/12/23/4859 (дата звернення: 15.04.2024).

4. Introduction to Convolutional Neural Networks (CNN). Analytics Vidhya. URL: https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/ (дата звернення: 30.12.2023).

5. What are Recurrent Neural Networks? | IBM. IBM in Deutschland, Österreich und der Schweiz | IBM. URL: https://www.ibm.com/topics/recurrent-neural-networks (дата звернення: 30.12.2023).

6. GitHub - CheyneyComputerScience/CREMA-D: Crowd Sourced Emotional Multimodal Actors Dataset (CREMA-D). GitHub. URL: https://github.com/CheyneyComputerScience/CREMA-D?tab=readme-ov-file (дата звернення: 30.12.2023).

7. librosa – librosa 0.10.1 documentation. Librosa. URL: https://librosa.org/doc/latest/index.html (дата звернення: 30.12.2023).

Рисунок В.6 – Остання сторінка тез

VOL.1

№89/2024

Znanstvena misel journal

The journal is registered and published in Slovenia.

ISSN 3124-1123

The frequency of publication – 12 times per year.

Journal is published in Slovenian, English, Polish, Russian, Ukrainian.

The format of the journal is A4, coated paper, matte laminated cover.

All articles are reviewed

Edition of journal does not carry responsibility for the materials published in a journal.

Sending the article to the editorial the author confirms it's uniqueness and takes full responsibility for possible consequences for breaking copyright laws

Free access to the electronic version of journal

Znanstvena misel journal

Slovenska cesta 8, 1000 Ljubljana, Slovenia

Email: info@znanstvena-journal.com

Website: www.znanstvena-journal.com

Рисунок В.7 – Остання сторінка журналу

ДОДАТОК Г

Повідомлення про прийняття статті на конференцію MoMLeT 2024

MoMLeT 2024, submission 21 Внешняя переписка

MoMLeT&DS Workshop
пт, 17 мая, 22:10 (11 дней назад)
кому: мне, iryna.afanasieva

Dear Authors

Daniil Suvorov, Iryna Afanasieva, Oleksii Turuta, Nataliia Golian, Vira Golian and Kostiantyn Onyshchenko

"Audio Processing Methods for Speech Emotion Recognition using Machine Learning"

We are pleased to inform you that your submissions have been accepted for MoMLeT 2024.

To further promote your materials, please follow these steps:

1) By May 25, 2024, send the following documents to momlet.in.ua@gmail.com:

- Author's consent (the template is attached). The consent must be printed, filled out by hand, and scanned in .pdf format;
- Presentation in .pdf format;
- A video report that will be posted on the workshop's YouTube channel.

2) Unfortunately, the CEUR publishing house has changed the templates again. Not only was the information at the bottom of the first page replaced, but the entire template was updated (the template is attached). Please revise your materials accordingly and submit the updated versions through the EasyChair system by May 25, 2024.

3) To expedite the release of the collection of materials, please adhere to the following author listing rules: The first author listed should be the one with the maximum number of works in DBLP. Authors should be listed in the order of their number of works in DBLP. Do not include articles published by CEUR-WS in the general list of DBLP works.

Thank you for your cooperation.

Best regards,
The MoMLeT 2024 Organizing Committee

Рисунок Г.1 – Повідомлення про прийняття статті

ДОДАТОК Д

Фрагменти коду

diploma - iemocap_feature_extraction.ipynb

```python
def get_iemocap_files_path():
    columns = ["Path", "Class"]
    df = pd.DataFrame(columns=columns)


    for root, dirs, files in os.walk("dataset_iemocap\IEMOCAP_full_release"):
        for dir in dirs:
            if dir == "EmoEvaluation":
                evaluation_path = os.path.join(root, dir)
                for file in os.listdir(evaluation_path):
                    file_path = os.path.join(evaluation_path, file)
                    if file_path.endswith(".txt"):
                        with open(file_path, 'r') as f:
                            for line in f:
                                match = re.match(r'\[(.*?) - (.*?)\]\s+(\S+)\s+(\S+)\s+(\[.*?\])', line)
                                if match:
                                    start_time, end_time, turn_name, emotion, vad = match.groups()
                                    match = re.search(r'.*Session\d+', evaluation_path)
                                    session_path = match.group()
                                    wav_dir = file.replace(".txt", "")
                                    wav_path = os.path.join(session_path, "sentences", "wav", wav_dir, f"{turn_name}.wav")

                                    df.loc[len(df)] = [wav_path, get_iemocap_emotion(emotion)]
    return df
```

Рисунок Д.1 – Формування набору даних зі шляхами до файлів та їх емоціями

diploma - models.ipynb

```python
def build_bilstm_model(input_shape, n_classes):
    model = Sequential(name="BiLSTM")

    # input layer
    model.add(Input(shape=input_shape))

    # recurrent layers
    model.add(BatchNormalization())
    model.add(Bidirectional(LSTM(128, return_sequences=True)))
    model.add(Dropout(0.2))
    model.add(Bidirectional(LSTM(256)))
    model.add(Dropout(0.2))

    # dense layers
    model.add(Dense(128, activation="relu"))
    model.add(Dropout(0.2))

    # output layer
    model.add(Dense(n_classes, activation="softmax"))

    optimizer = Adam(learning_rate=0.0001)
    model.compile(loss="categorical_crossentropy",
                  optimizer=optimizer,
                  metrics=["accuracy"])

    return model
```

Рисунок Д.2 – Код побудови моделі із використанням BiLSTM

diploma - models.ipynb

```python
def build_gru_model(input_shape, n_classes):
    model = Sequential(name="GRU")

    # input layer
    model.add(Input(shape=input_shape))

    # recurrent layers
    model.add(BatchNormalization())
    model.add(GRU(256, return_sequences=True))
    model.add(Dropout(0.2))
    model.add(GRU(512))
    model.add(Dropout(0.2))

    # dense layers
    model.add(Dense(128, activation="relu"))
    model.add(Dropout(0.2))

    # output layer
    model.add(Dense(n_classes, activation="softmax"))

    optimizer = Adam(learning_rate=0.0001)
    model.compile(loss="categorical_crossentropy",
                  optimizer=optimizer,
                  metrics=["accuracy"])

    return model
```

Рисунок Д.3 - Код побудови моделі із використанням GRU

diploma - models.ipynb

```python
def build_2d_cnn_model(input_shape, n_classes):
    model = Sequential(name="CNN2D")

    # input layer
    model.add(Input(shape=input_shape))
    model.add(BatchNormalization())

    # conv layers
    for _ in range(6):
        model.add(Conv2D(128, 5, padding="same", activation="relu"))
        model.add(BatchNormalization())
        model.add(MaxPooling2D(pool_size=4, padding="same"))
        model.add(Dropout(0.2))

    model.add(Flatten())

    # dense layer
    model.add(Dense(128, activation="relu"))
    model.add(Dropout(0.2))

    # output layer
    model.add(Dense(n_classes, activation="softmax"))

    optimizer = Adam(learning_rate=0.001)
    model.compile(loss="categorical_crossentropy",
                  optimizer=optimizer,
                  metrics=["accuracy"])

    return model
```

Рисунок Д.4 – Код побудови моделі із використанням CNN

diploma - models.ipynb

```python
def build_2d_crnn_model(input_shape, n_classes):
    model = Sequential(name="CRNN2D")

    # input layer
    model.add(Input(shape=input_shape))
    model.add(BatchNormalization())

    # conv layers
    for _ in range(6):
        model.add(Conv2D(128, 5, padding="same", activation="relu"))
        model.add(BatchNormalization())
        model.add(MaxPooling2D(pool_size=4, padding="same"))
        model.add(Dropout(0.2))

    model.add(Reshape((1, 128)))

    # recurrent layers
    model.add(LSTM(256, return_sequences=True))
    model.add(LSTM(512))
    model.add(Dropout(0.2))

    model.add(Flatten())

    # dense layer
    model.add(Dense(128, activation="relu"))
    model.add(Dropout(0.2))

    # output layer
    model.add(Dense(n_classes, activation="softmax"))

    optimizer = Adam(learning_rate=0.001)
    model.compile(loss="categorical_crossentropy",
                  optimizer=optimizer,
                  metrics=["accuracy"])

    return model
```

Рисунок Д.5 – Код побудови моделі із використанням CNN та LSTM

ДОДАТОК Е

Приклади порівняння оригінального аудіо з аугментованим

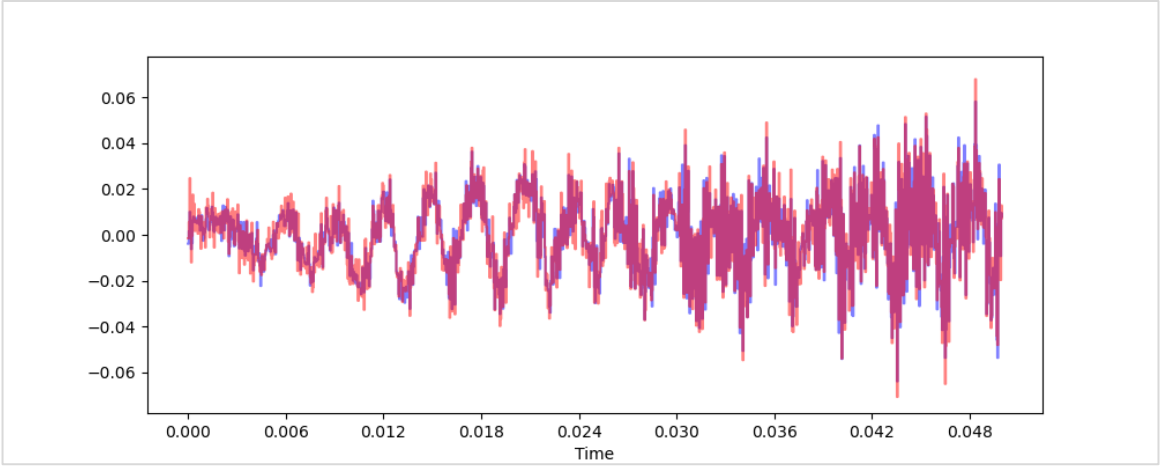


Рисунок Е.1 – Оригінальне аудіо та аудіо із доданим шумом

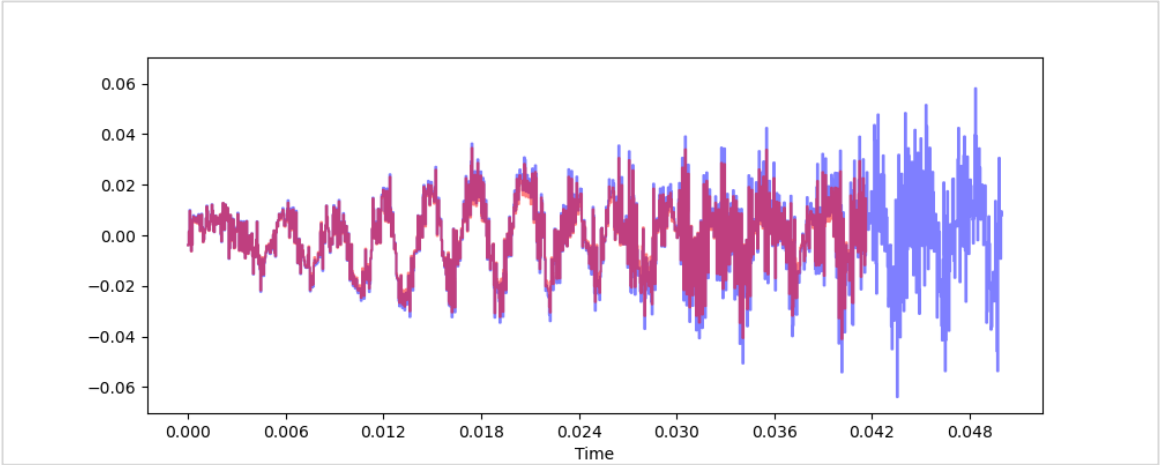


Рисунок Е.2 – Оригінальне аудіо та розтягнуте аудіо

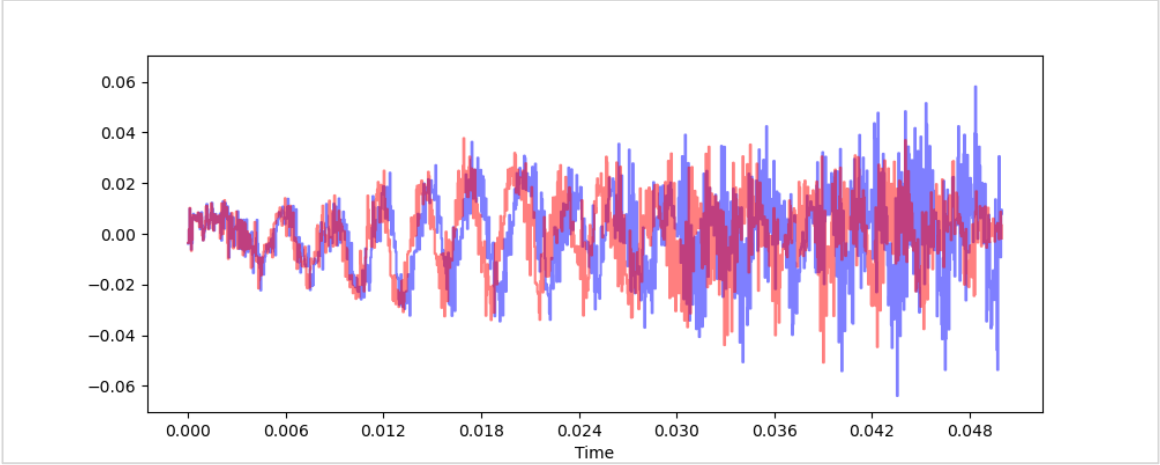


Рисунок Е.3 – Оригінальне аудіо та аудіо із підвищенням тону

ДОДАТОК Ж

Опис архітектури моделей нейронних мереж

```
Model: "BiLSTM"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 batch_normalization (BatchN  (None, 128, 40)          160
 ormalization)

 bidirectional (Bidirectiona  (None, 128, 256)         173056
 l)

 dropout (Dropout)           (None, 128, 256)          0

 bidirectional_1 (Bidirectio  (None, 512)              1050624
 nal)

 dropout_1 (Dropout)         (None, 512)               0

 dense (Dense)               (None, 128)               65664

 dropout_2 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 4)                 516

=================================================================
Total params: 1,290,020
Trainable params: 1,289,940
Non-trainable params: 80
_________________________________________________________________
```

Рисунок Ж.1 – Опис архітектури моделі з BiLSTM

```
Model: "GRU"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 batch_normalization (BatchN  (None, 128, 40)          160
 ormalization)

 gru (GRU)                   (None, 128, 256)          228864

 dropout (Dropout)           (None, 128, 256)          0

 gru_1 (GRU)                 (None, 512)               1182720

 dropout_1 (Dropout)         (None, 512)               0

 dense (Dense)               (None, 128)               65664

 dropout_2 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 4)                 516

=================================================================
Total params: 1,477,924
Trainable params: 1,477,844
Non-trainable params: 80
_________________________________________________________________
```

Рисунок Ж.2 – Опис архітектури моделі з GRU

```
Model: "CNN2D"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 batch_normalization (BatchN  (None, 128, 40, 1)       4
 ormalization)

 conv2d (Conv2D)             (None, 128, 40, 128)      3328

 batch_normalization_1 (Batc  (None, 128, 40, 128)     512
 hNormalization)

 max_pooling2d (MaxPooling2D  (None, 32, 10, 128)      0
 )

 dropout (Dropout)           (None, 32, 10, 128)       0

 conv2d_1 (Conv2D)           (None, 32, 10, 128)       409728

 batch_normalization_2 (Batc  (None, 32, 10, 128)      512
 hNormalization)

 max_pooling2d_1 (MaxPooling  (None, 8, 3, 128)        0
 2D)

 dropout_1 (Dropout)         (None, 8, 3, 128)         0

 conv2d_2 (Conv2D)           (None, 8, 3, 128)         409728

 batch_normalization_3 (Batc  (None, 8, 3, 128)        512
 hNormalization)

 max_pooling2d_2 (MaxPooling  (None, 2, 1, 128)        0
 2D)

 dropout_2 (Dropout)         (None, 2, 1, 128)         0

 conv2d_3 (Conv2D)           (None, 2, 1, 128)         409728

 batch_normalization_4 (Batc  (None, 2, 1, 128)        512
 hNormalization)

 max_pooling2d_3 (MaxPooling  (None, 1, 1, 128)        0
 2D)

 dropout_3 (Dropout)         (None, 1, 1, 128)         0

 conv2d_4 (Conv2D)           (None, 1, 1, 128)         409728

 batch_normalization_5 (Batc  (None, 1, 1, 128)        512
 hNormalization)

 max_pooling2d_4 (MaxPooling  (None, 1, 1, 128)        0
 2D)

 dropout_4 (Dropout)         (None, 1, 1, 128)         0

 conv2d_5 (Conv2D)           (None, 1, 1, 128)         409728

 batch_normalization_6 (Batc  (None, 1, 1, 128)        512
 hNormalization)

 max_pooling2d_5 (MaxPooling  (None, 1, 1, 128)        0
 2D)

 dropout_5 (Dropout)         (None, 1, 1, 128)         0

 flatten (Flatten)           (None, 128)               0

 dense (Dense)               (None, 128)               16512

 dropout_6 (Dropout)         (None, 128)               0

 dense_1 (Dense)             (None, 4)                 516

=================================================================
Total params: 2,072,072
Trainable params: 2,070,534
Non-trainable params: 1,538
_________________________________________________________________
```

Рисунок Ж.3 – Опис архітектури згорткової моделі

```
Model: "CRNN2D"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 batch_normalization (BatchN  (None, 128, 40, 1)       4
 ormalization)

 conv2d (Conv2D)             (None, 128, 40, 128)      3328

 batch_normalization_1 (Batc  (None, 128, 40, 128)     512
 hNormalization)

 max_pooling2d (MaxPooling2D  (None, 32, 10, 128)      0
 )

 dropout (Dropout)           (None, 32, 10, 128)       0

 conv2d_1 (Conv2D)           (None, 32, 10, 128)       409728

 batch_normalization_2 (Batc  (None, 32, 10, 128)      512
 hNormalization)

 max_pooling2d_1 (MaxPooling  (None, 8, 3, 128)        0
 2D)

 dropout_1 (Dropout)         (None, 8, 3, 128)         0

 conv2d_2 (Conv2D)           (None, 8, 3, 128)         409728

 batch_normalization_3 (Batc  (None, 8, 3, 128)        512
 hNormalization)

 max_pooling2d_2 (MaxPooling  (None, 2, 1, 128)        0
 2D)

 dropout_2 (Dropout)         (None, 2, 1, 128)         0

 conv2d_3 (Conv2D)           (None, 2, 1, 128)         409728

 batch_normalization_4 (Batc  (None, 2, 1, 128)        512
 hNormalization)

 max_pooling2d_3 (MaxPooling  (None, 1, 1, 128)        0
 2D)

 dropout_3 (Dropout)         (None, 1, 1, 128)         0
```

Рисунок Ж.4 – Опис архітектури моделі CRNN (початок)

```
conv2d_4 (Conv2D)           (None, 1, 1, 128)         409728

batch_normalization_5 (Batc  (None, 1, 1, 128)        512
hNormalization)

max_pooling2d_4 (MaxPooling  (None, 1, 1, 128)        0
2D)

dropout_4 (Dropout)         (None, 1, 1, 128)         0

conv2d_5 (Conv2D)           (None, 1, 1, 128)         409728

batch_normalization_6 (Batc  (None, 1, 1, 128)        512
hNormalization)

max_pooling2d_5 (MaxPooling  (None, 1, 1, 128)        0
2D)

dropout_5 (Dropout)         (None, 1, 1, 128)         0

reshape (Reshape)           (None, 1, 128)            0

lstm (LSTM)                 (None, 1, 256)            394240

lstm_1 (LSTM)               (None, 512)               1574912

dropout_6 (Dropout)         (None, 512)               0

flatten (Flatten)           (None, 512)               0

dense (Dense)               (None, 128)               65664

dropout_7 (Dropout)         (None, 128)               0

dense_1 (Dense)             (None, 4)                 516

=================================================================
Total params: 4,090,376
Trainable params: 4,088,838
Non-trainable params: 1,538
_________________________________________________________________
```

Рисунок Ж.5 – Опис архітектури моделі CRNN (кінець)

ДОДАТОК И

Матриці помилок

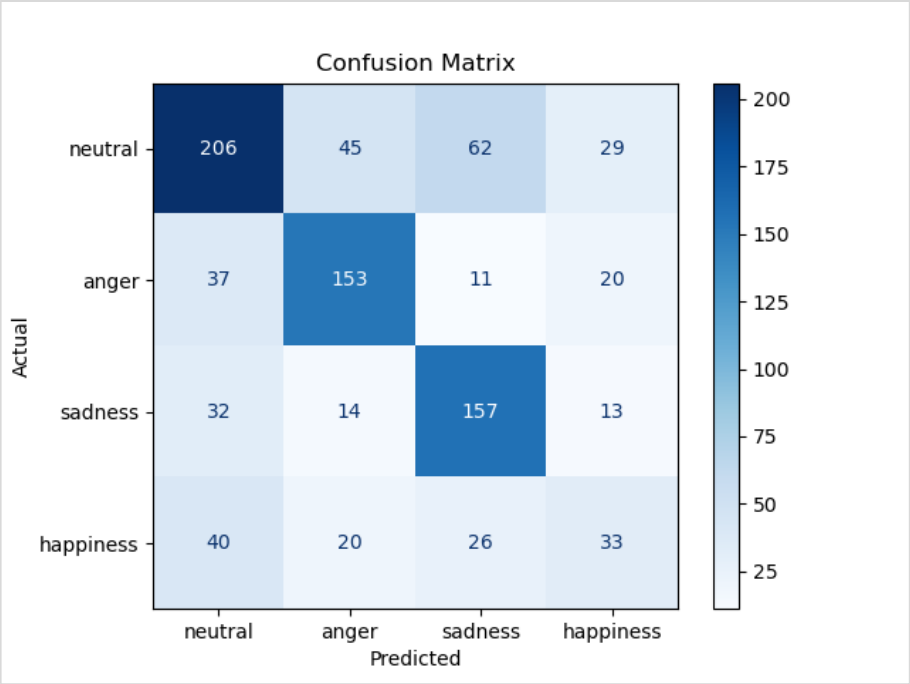


Рисунок И.1 – Матриця помилок у моделі BiLSTM на наборі IEMOCAP без аугментації

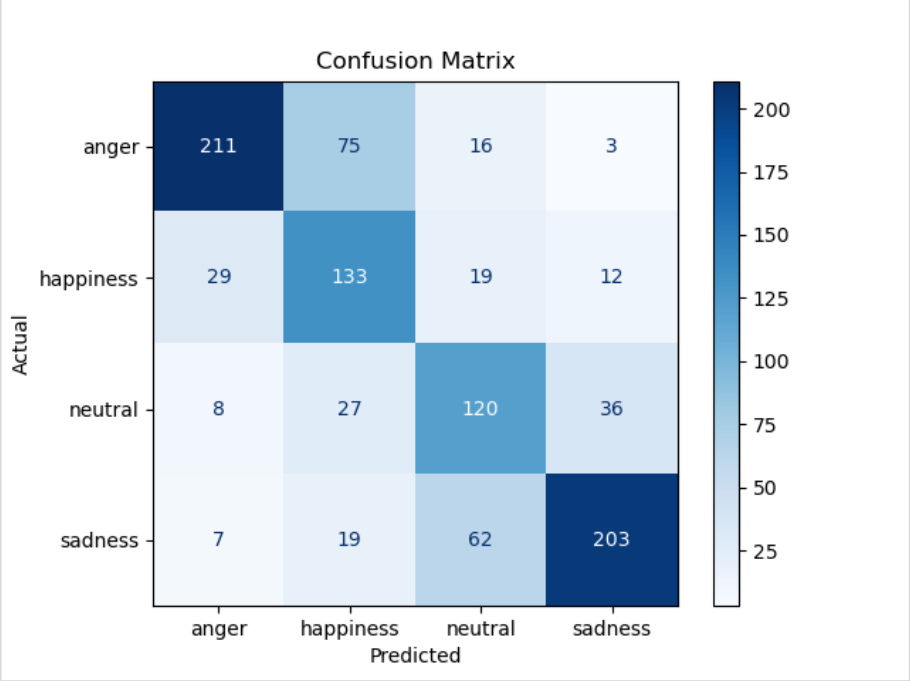


Рисунок И.2 – Матриця помилок у моделі BiLSTM на наборі CREMA-D без аугментації

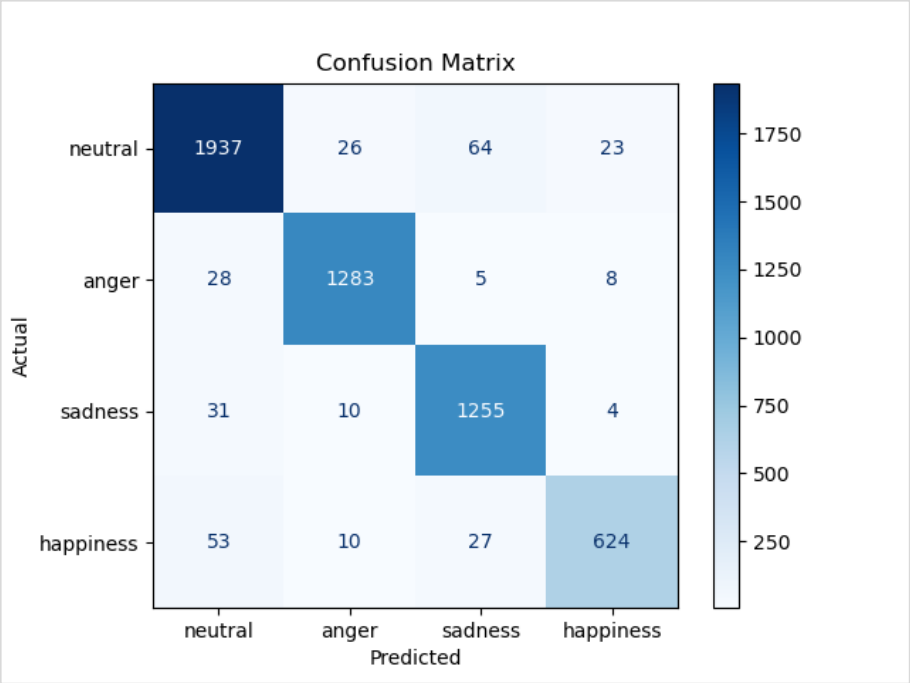


Рисунок И.3 – Матриця помилок у моделі BiLSTM на наборі IEMOCAP з аугментацією

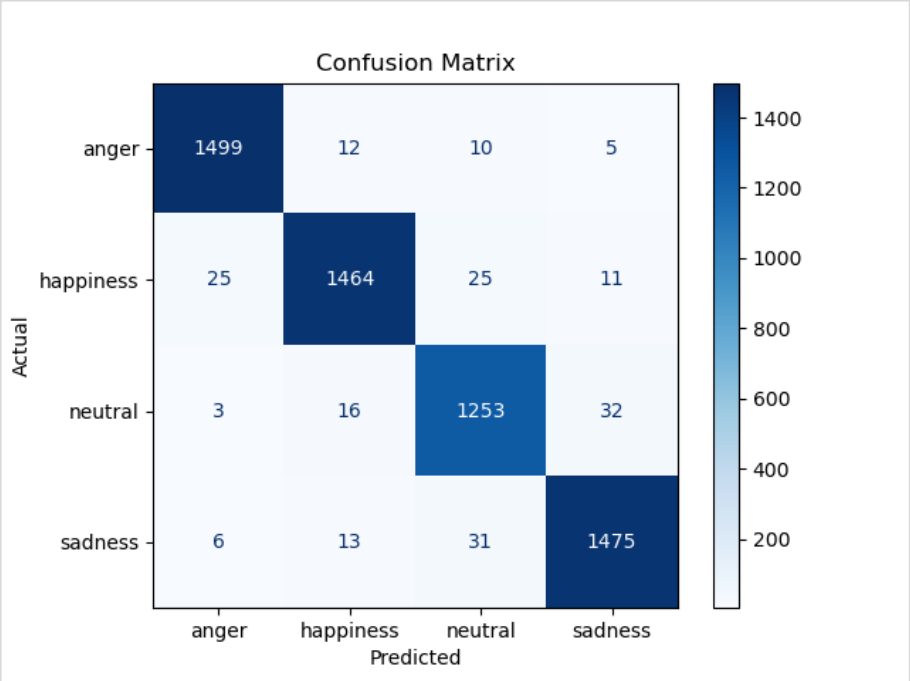


Рисунок И.4 – Матриця помилок у моделі BiLSTM на наборі CREMA-D з аугментацією

ДОДАТОК К

Результати експериментів для різних наборів даних

Таблиця К.1 – Результати досліджень для набору CREMA-D без аугментацій

	BiLSTM			**GRU**			**CNN**			**CRNN**			
	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	support
anger	0,749	0,802	0,772	0,766	0,752	0,757	0,828	0,789	0,803	0,780	0,826	0,790	254
happiness	0,686	0,592	0,633	0,621	0,603	0,608	0,727	0,710	0,718	0,724	0,666	0,689	254
neutral	0,620	0,596	0,607	0,480	0,522	0,498	0,678	0,732	0,698	0,668	0,790	0,720	217
sadness	0,700	0,762	0,729	0,655	0,625	0,638	0,799	0,762	0,771	0,856	0,691	0,758	254
	0,691	0,691	0,688	0,636	0,630	0,630	0,761	0,749	0,749	0,760	0,741	0,740	
accuracy	0,691			0,630			**0,749**			0,741			

Таблиця К.2 – Результати досліджень для набору IEMOCAP без аугментацій

	BiLSTM			**GRU**			**CNN**			**CRNN**			
	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	support
anger	0,701	0,725	0,712	0,663	0,693	0,674	0,732	0,754	0,742	0,781	0,704	0,733	221
happiness	0,346	0,227	0,272	0,275	0,180	0,215	0,414	0,190	0,259	0,402	0,276	0,317	119
neutral	0,637	0,655	0,645	0,608	0,625	0,613	0,632	0,714	0,669	0,636	0,690	0,659	342
sadness	0,632	0,693	0,661	0,606	0,650	0,627	0,686	0,704	0,684	0,660	0,713	0,681	217
	0,613	0,624	0,616	0,577	0,589	0,579	0,641	0,652	0,636	0,646	0,644	0,637	
accuracy	0,624			0,589			**0,652**			0,644			

Таблиця К.3 – Результати досліджень для набору CREMA-D з аугментаціями

	BiLSTM			**GRU**			**CNN**			**CRNN**			
	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	support
anger	0,978	0,979	0,979	0,982	0,977	0,980	0,998	0,992	0,995	0,996	0,994	0,995	1 525
happiness	0,969	0,960	0,965	0,971	0,964	0,967	0,992	0,987	0,989	0,993	0,981	0,987	1 525
neutral	0,944	0,963	0,953	0,937	0,957	0,947	0,977	0,986	0,981	0,981	0,984	0,983	1 304
sadness	0,969	0,958	0,963	0,963	0,957	0,960	0,985	0,987	0,986	0,982	0,992	0,987	1 525
	0,966	0,965	0,965	0,964	0,964	0,964	0,988	0,988	0,988	0,988	0,988	0,988	
accuracy	0,965			0,964			**0,988**			**0,988**			

Таблиця К.4 – Результати досліджень для набору IEMOCAP з аугментаціями

	BiLSTM			**GRU**			**CNN**			**CRNN**			
	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	support
anger	0,969	0,962	0,965	0,979	0,970	0,974	0,993	0,992	0,993	0,997	0,989	0,993	1 324
happiness	0,945	0,884	0,913	0,939	0,904	0,921	0,993	0,953	0,973	0,990	0,962	0,975	714
neutral	0,936	0,947	0,942	0,947	0,959	0,953	0,981	0,983	0,982	0,984	0,980	0,982	2 050
sadness	0,927	0,950	0,938	0,946	0,954	0,950	0,967	0,987	0,977	0,960	0,989	0,974	1 301
	0,943	0,943	0,943	0,954	0,953	0,953	0,982	0,982	0,982	0,982	0,982	0,982	
accuracy	0,943			0,953			**0,982**			**0,982**			

ДОДАТОК Л

Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ

Рисунок Л.1 – Результат перевірки на унікальність тексту

ДОДАТОК М

Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008:2015

Експертний висновок результатів перевірки кваліфікаційної роботи

студент (посада) | програмної інженерії (кафедра) | ІПЗм-22-2 (група)

Суворов Д.С
(прізвище, ім'я, по батькові)

Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні познаки	
	7.15 Додатки	
Методичні вказівки до виконання кваліфікаційної роботи магістра… **ЗАТВЕРДЖЕНО** кафедрою ІІІ протокол № 5 від 13.11.2023р. 3.2 Оформлення пояснювальної записки згідно з ДСТУ 3008:2015 Звіти у сфері науки і техніки. Структура та правила оформлювання. **Шаблон** затверджений засіданням кафедри №3 від 16.10.2023.	Назву таблиці друкують з великої літери і розміщують над таблицею з абзацного відступу та в круглих дужках вказується джерело з якого взята ця таблиця. ПРИКЛАД: шаблон, стор.15	20, далі за текстом.

Експерт _____ (підпис) Вадим НЕЧВОЛОД (прізвище, ініціали)

03.06.2024

Рисунок М.1 – Результат перевірки записки на відповідність оформлення вимогам ДСТУ 3008:2015