

## АНАЛІЗ ЕМОЦІЙ З ВИКОРИСТАННЯМ ВИРАЗІВ ОБЛИЧЧЯ ТА ГОЛОСОВИХ ОЗНАК

*Анотація. У роботі представлені методи для розпізнавання людських емоцій. Розпізнавання людських емоцій - це галузь штучного інтелекту, що швидко розвивається і має важливе значення для поліпшення взаємодії людини з комп'ютером. Однак більшість сучасних систем розпізнавання емоцій обмежені використанням тільки одного джерела даних голосових характеристик, або виразів обличчя, що знижує точність і стійкість розпізнавання в складних умовах. В цій статті розглядаються методи розпізнавання емоцій з використанням двох основних каналів - голосового та візуального. У процесі дослідження аналізуються наявні підходи щодо розпізнавання емоцій. Для аналізу емоцій розглянуто технології ResNet (для зображень), SER (для звуку). Особливу увагу приділено аналізу й обробці акустичних характеристик, таких як інтонація, гучність, темп мовлення і тривалість пауз, а також використанню методів комп'ютерного зору для детекції виразів обличчя, таких як посмішка, затиснуті губи або насуплені брови. Це дослідження просуває галузь вивчення емоцій, забезпечуючи краще розуміння емоційних станів людини.*

*Ключові слова: розпізнавання емоцій, розпізнавання емоцій через звук, розпізнавання емоцій за обличчям, машинне навчання, SVM, комп'ютерний зір, глибоке навчання, CNN.*

**Постановка проблеми.** Проблема розпізнавання емоцій полягає в тому, що емоції людини проявляються багатогранно та змінно. Сучасні системи розпізнавання емоцій на основі аналізу голосу або виразів обличчя мають певні обмеження, які впливають на їх точність і спроможність у реальних умовах.

**Мета і завдання досліджень.** Вивчити моделі для розпізнавання емоцій з використанням акустичних і візуальних даних для покращення точності та стабільності системи в умовах реальної взаємодії, що дасть змогу точніше інтерпретувати емоційні стани людини.

Емоційне розпізнавання з використанням штучного інтелекту (ШІ) охоплює кілька різних підходів, які дають змогу системам аналізувати емоції людини.

Розглянемо технологію розпізнавання емоцій за виразами обличчя. Це один із найпоширеніших методів, який використовує комп'ютерний зір для аналізу міміки людини. ШІ може знаходити і визначати вирази облич (усмішка, насуплені брови тощо) і на основі цього визначати емоції, такі як щастя, сум, здивування, злість та інші.

Для розв'язання цієї задачі використовують різні методи машинного навчання та глибокого навчання, а одним із найефективніших підходів є використання глибоких нейронних мереж (зокрема, архітектур, таких як ResNet, SVM, CNN та інші)[1].

Обробка та аналіз зображень облич є ключовим етапом під час розпізнавання емоцій за мімікою обличчя. Цей процес включає кілька важливих кроків, які допомагають підготувати зображення для подальшої класифікації емоцій.

У першому етапі, потрібно спочатку ідентифікувати обличчя на зображенні. Це називається детекцією обличчя. На зображенні обличчя може бути нечітким або частково закритим, а також можуть бути кілька облич. Використовуються різні методи для виявлення обличчя у зображенні.

На другому етапі після детекції обличчя ми отримуємо координати прямокутника, що обмежує обличчя, що дає змогу сфокусуватися тільки на одній ділянці зображення для подальшого опрацювання (див. рис 1).



Рисунок 1 – Детекція обличчя

На третьому етапі після того, як обличчя виявлено, важливо виконати кілька кроків до обробки та аналізу. Для початку потрібно змінити розміри для подачі зображення в нейронну мережу, розмір обличчя має бути стандартизованим, зазвичай, використовується розмір 224x224 пікселів. Для того щоб модель могла працювати з однаковими розмірами зображень, а також для прискорення обчислень. Іноді, зображення обличчя конвертуються у відтінки сірого, для спрощення обробки. Але сучасні моделі часто працюють із кольоровими зображеннями (RGB), щоб враховувати більш тонкі деталі. Нормалізація, після зміни розміру і можливо перетворення кольору, зазвичай нормалізують піксельні значення, щоб вони перебували в одному діапазоні. Наприклад, у нейронних мережах часто застосовується нормалізація, де значення пікселів від 0 до 255 приводяться до діапазону [0, 1] або [-1, 1]. Для поліпшення загальних здібностей моделі використовують аугментацію даних, наприклад, повороти, відображення, зміни яскравості, контрасту тощо. Це допомагає уникнути перенавчання і поліпшити здатність моделі справлятися з різними умовами.

На четвертому етапі модель або алгоритм витягує ключові характеристики (ознаки) із зображення, які можуть допомогти в класифікації емоцій. Для цього використо-

вують сучасні методи машинного навчання, згорткові нейронні мережі, для вилучення ознак. Ці мережі вчать вилучати просторові ознаки, такі як контури, текстури та інші особливості.

На п'ятому етапі визначається геометрія обличчя, використовуються традиційні методи для визначення точок на обличчі та їхнього аналізу (відстань між очима, кут нахилу рота тощо). Для цього етапу можна використовувати “згорткові нейронні мережі” (CNN). Ці мережі використовують фільтри, щоб виявляти важливі візуальні ознаки: (форма обличчя, вирази очей і т.д.).

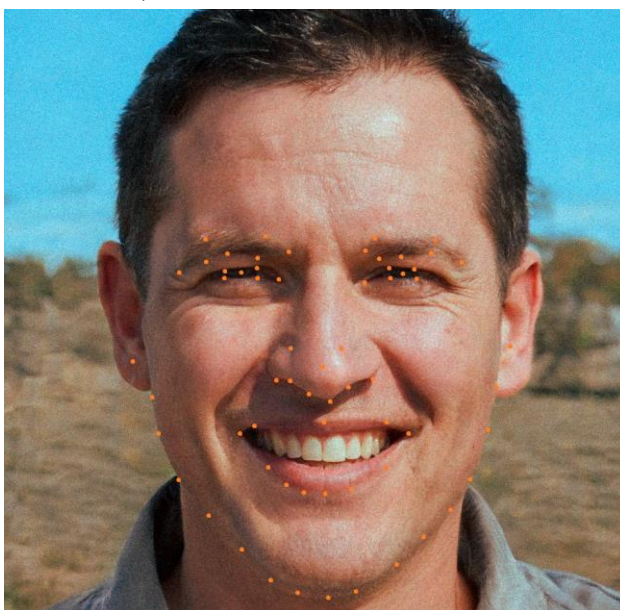


Рисунок 2 - Визначення точок на обличчі

Після опрацювання та аналізу зображення обличчя на наступному етапі відбувається використання класифікатора, який передбачає емоцію обличчя за зображенням, ґрунтуючись на даних ознаках. Для класифікатора емоції використовують різні алгоритми, як-от ResNet, або Residual Network, що слугує основою для багатьох завдань комп'ютерного зору. ResNet ефективно дає змогу навчати глибокі нейронні мережі з більш ніж 150 шарами [2]. За витягнутими раніше ознаками можна донавчити модель для класифікації емоцій. Після того як модель навчена і протестована, ви можете використовувати її для передбачення емоцій на нових зображеннях. А також можна використовувати Support Vector Machine (SVM) - алгоритм, який будує гіперплощини для поділу різних класів.

Далі розглянемо технологію розпізнавання емоцій за голосом, або паралінгвістику, яка вивчає, як люди передають емоційні стани за допомогою особливостей мови, таких як інтонація, темп, гучність, висота голосу та інші голосові характеристики. Цей процес має важливе значення в галузі психолінгвістики, когнітивних наук і штучного інтелекту. Коли ми слухаємо когось, ми інтуїтивно розпізнаємо не тільки зміст слів, а й емоції, які вони висловлюють. Наприклад, якщо людина говорить із підвищеним темпом і високим тоном, це може свідчити про радість або хвилювання. Якщо її голос ни-

жчий і повільніший, це може вказувати на смуток або розчарування. Приклади технологій: Emotion Recognition (SER), Neural Networks (CNN)[3]

До основних аспектів паралінгвістики, що допомагають розпізнавати емоції за голосом, належать такі параметри мовлення:

1. Тональність голосу - це найважливіший елемент у визначенні емоційного стану людини, тому що вона безпосередньо пов'язана з її емоційною реакцією на ситуацію. Тон може передавати навіть точніші емоції, ніж самі слова. Високий тон часто асоціюється з позитивними емоціями, такими як радість або збудження. Низький тон може сигналізувати про сум, розчарування або втому. Коливання в тоні (монотонність), можуть вказувати на хвилювання, здивування або інтерес.

2. Гучність голосу може ефективно відображати емоційний стан, оскільки люди схильні підвищувати гучність у моменти сильних емоцій і знижувати її в більш спокійних або сумних станах. Високий рівень гучності часто вказує на гнів, лютю, агресію, збудження або тривогу. Низька гучність, коли людина говорить тихо, це може бути пов'язано зі смутком, нудьгою, депресією або невпевненістю. Раптові зміни гучності, різкі підвищення і зниження гучності можуть свідчити про те, що людина переживає сильні емоції, такі як здивування або шок.

3. Темп мовлення це швидкість, з якою людина вимовляє слова. Він відіграє важливу роль у розпізнаванні емоцій, оскільки на емоційному тлі темп мовлення часто змінюється. Швидка мова часто вказує на збудження, тривогу, стрес або хвилювання. Повільне мовлення може сигналізувати про смуток, втому, депресію або роздуми. Перерви і паузи в мовленні, це може означати, що людина нервує, сумнівається, намагається підібрати правильні слова або переживає.

4. Інтонація мови це зміна висоти голосу та інших його характеристик упродовж мовлення, яка допомагає передавати емоційні відтінки. Підйом інтонації, це може вказувати на здивування, хвилювання або інтерес. Спад інтонації може сигналізувати про завершення думки, спокій або сум. Змінна інтонація, це коли інтонація варіюється протягом фрази, це може бути ознакою збудження, радості чи інтересу.

Для аналізу та розпізнавання емоцій на основі голосу часто використовують методи машинного навчання. Моделі класифікації можуть працювати з алгоритмами: логістична регресія, метод опорних векторів (SVM), нейронні мережі. Ці моделі можуть бути навчені на наборі даних із заздалегідь розміченими емоціями, наприклад, з використанням аудіофайлів, які містять різні емоції (гнів, радість, смуток тощо)[4].

Акустичні ознаки та їх ваги визначаються наступним чином:

$$E = w_1 \cdot F0 + w_2 \cdot V + w_3 \cdot T + w_4 \cdot P ,$$

де E - емоційний показник,

F0 - основна частота,

V - гучність,

T - темп мовлення,

P - тривалість пауз,

$w_1, w_2, w_3, w_4$  - ваги, які визначаються в процесі навчання моделі.

Кожен із цих параметрів має різний внесок у загальну оцінку емоції, їхні ваги можуть бути налаштовані на основі навчальних даних, представлена блок схема (див. рис 3).

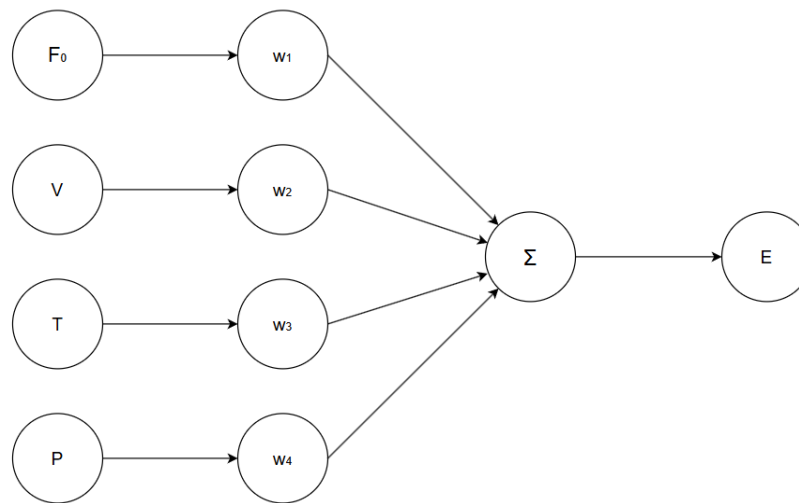


Рисунок 3 - Акустичні ознаки та їх ваги

Для більш наочного уявлення розпізнавання емоцій за голосом можна уявити, як можуть змінюватися акустичні характеристики при різних емоціях (див. табл. 1)

Таблица 1

Акустичні характеристики при різних емоціях

| Емоція         | Основна частота (F0) | Гучність (V)   | Темп мовлення (T) | Паузи (P) |
|----------------|----------------------|----------------|-------------------|-----------|
| <i>Радість</i> | Висока               | Середня/висока | Швидка            | Короткі   |
| <i>Сум</i>     | Низька               | Середня/низька | Повільна          | Тривалі   |
| <i>Гнів</i>    | Дуже висока          | Висока         | Швидка            | Короткі   |
| <i>Страх</i>   | Висока               | Висока         | Швидка            | Тривалі   |

Для класифікації емоції можуть бути використані певні порогові значення цих характеристик.

**Висновок.** Розглянуті в роботі методи розпізнавання емоцій за виразами обличчя і розпізнавання емоцій за голосом - це дві різні технології, кожен з яких використовує різні види даних для аналізу та інтерпретації емоцій.

Розпізнавання емоцій за виразами обличчя - це складне завдання в галузі комп'ютерного зору і глибокого навчання. Використання моделей, таких як ResNet, дає змогу ефективно розв'язувати поставлені задачі, завдяки глибокому навчанню і здатності витягувати складні ознаки із зображень облич.

Розпізнавання емоцій за голосом ґрунтується на складному аналізі безлічі акустичних ознак, таких як частота, гучність, темп мовлення, інтонація та інші. Ці ознаки можуть бути піддані аналізу з використанням різних математичних і статистичних моделей, таких як машинне навчання і нейронні мережі, для точної класифікації емоцій.

### ЛІТЕРАТУРА

1. B. Mandal, A. Okeukwu, Y. Theis. Masked face recognition using ResNet-50. (2021), DOI:10.48550/ARXIV.2104.08997
2. Enguerrand Boitel , Alaa Mohasseb , Ella Haig. MIST: Multimodal emotion recognition using DeBERTa for text, Semi-CNN for speech, ResNet-50 for facial, and 3D-CNN for motion analysis. Expert Systems with Applications Volume 270, 25 April 2025, 126236. DOI: <https://doi.org/10.1016/j.eswa.2024.126236>
3. Z. Huang, M. Dong, Q. Mao, Y. Zhan. Speech emotion recognition using CNN. Proceedings of the 22nd ACM international conference on multimedia, Association for Computing Machinery, New York, NY, USA (2014), pp. 801-804, DOI: 10.1145/2647868.2654984
4. E. Lakomkin, C. Weber, S. Magg, S. Wermter. Reusing neural speech representations for auditory emotion recognition. (2018), DOI: 10.48550/ARXIV.1803.11508

### REFERENCES

1. B. Mandal, A. Okeukwu, Y. Theis. Masked face recognition using ResNet-50. (2021), DOI:10.48550/ARXIV.2104.08997
2. Enguerrand Boitel , Alaa Mohasseb , Ella Haig. MIST: Multimodal emotion recognition using DeBERTa for text, Semi-CNN for speech, ResNet-50 for facial, and 3D-CNN for motion analysis. Expert Systems with Applications Volume 270, 25 April 2025, 126236. DOI: <https://doi.org/10.1016/j.eswa.2024.126236>
3. Z. Huang, M. Dong, Q. Mao, Y. Zhan. Speech emotion recognition using CNN. Proceedings of the 22nd ACM international conference on multimedia, Association for Computing Machinery, New York, NY, USA (2014), pp. 801-804, DOI: 10.1145/2647868.2654984
4. E. Lakomkin, C. Weber, S. Magg, S. Wermter. Reusing neural speech representations for auditory emotion recognition. (2018), DOI: 10.48550/ARXIV.1803.11508

Received 31.03.2025.  
Accepted 01.04.2025.

### ***Analysis of emotions using facial expressions and voice features***

*This article presents methods for recognizing human emotions. Human emotion recognition is a rapidly developing field of artificial intelligence that is essential for improving human-computer interaction. However, most modern emotion recognition systems are limited to using only one data source of voice characteristics or facial expressions, which reduces the accuracy and robustness of recognition in complex environments. The problem with emotion recognition is that human emotions are multifaceted and variable. Modern emotion recognition systems based on voice or facial expression analysis have certain limitations that affect their accuracy and performance in real-world conditions. This article discusses methods of emotion recognition using two main channels - voice and visual. The study analyzes existing approaches to emotion recognition. For the analysis of emotions, we consider ResNet (for images) and SER (for sound) technologies. Particular attention is paid to the analysis and processing of acoustic characteristics such as intonation, volume, speech rate and pause duration, as well as the use of computer vision methods to detect facial*

*expressions such as a smile, pursed lips or furrowed brows. This research is advancing the field of emotion research by providing a better understanding of human emotional states. In the course of this work, we have considered methods of emotion recognition:*

*Emotion recognition from facial expressions and emotion recognition from voice are two different technologies, each of which uses different types of data to analyze and interpret emotions.*

*Emotion recognition from facial expressions is an exciting task in the field of computer vision and deep learning, with numerous applications in various industries. Emotion recognition from voice is based on a complex analysis of many acoustic features such as frequency, volume, speech rate, intonation, and others. These features can be analyzed using various mathematical and statistical models, such as machine learning and neural networks, to accurately classify emotions.*

*Keywords: emotion recognition, audio emotion recognition, facial emotion recognition, machine learning, SVM, computer vision, deep learning, CNN.*

**Дмитрієва Ірина Сергіївна** - к.т.н., доцент, доцен кафедри Інформаційних технологій і систем Дніпровський металургійний інститут, Український державний університету науки і технологій, Україна.

**Бімалов Дмитро Вікторович** – аспірант кафедри Інформаційних технологій і систем Дніпровський металургійний інститут, Український державний університету науки і технологій, Україна.

**Dmytriieva Iryna Serhiivna** - Ph.D., Associate Professor, Associate Professor of the Department of Information Technologies and Systems, Dnipro Metallurgical Institute, Ukrainian State University of Science and Technology, Ukraine.

**Bimalov Dmytro Viktorovych** - Postgraduate Student of the Department of Information Technologies and Systems, Dnipro Metallurgical Institute, Ukrainian State University of Science and Technology, Ukraine.