Міністерство освіти і науки України
Луцький національний технічний університет
Факультет робототехніки та штучного інтелекту
Кафедра штучного інтелекту та математичного моделювання

КВАЛІФІКАЦІЙНА РОБОТА ЗА СТУПЕНЕМ ВИЩОЇ ОСВІТИ
«БАКАЛАВР»

АНАЛІЗ ТА РОЗРОБКА МЕТОДІВ РОЗПІЗНАВАННЯ ЕМОЦІЙНОГО СТАНУ ЛЮДИНИ ЗА АУДІОСИГНАЛОМ

ANALYSIS AND DEVELOPMENT OF METHODS FOR RECOGNIZING A PERSON'S EMOTIONAL STATE FROM AN AUDIO SIGNAL

Спеціальність 113 Прикладна математика
(шифр і назва спеціальності)

освітня програма «Штучний інтелект та аналіз масивів даних»
(назва освітньої програми)

Виконав: здобувач вищої освіти
Групи ПРМ-41
Михальчук Мирослав Ігорович

_____
(підпис)

Керівник:
к.т.н., доцент
Приходько Олексій Сергійович

_____
(підпис)

Кваліфікаційну роботу
допущено до захисту
«___» _____ 20__ р.
к.т.н., доцент
Гарант освітньої програми:
Приходько Олексій Сергійович

_____
(підпис)

Луцьк – 2026 року

ЛУЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

Факультет архітектури, будівництва та дизайну
Кафедра прикладної математики та механіки
Ступінь вищої освіти: *бакалавр*
Галузь знань: *11 Математика і статистика*
Спеціальність *113 Прикладна математика*
Освітня програма *Штучний інтелект та аналіз масивів даних*

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Мікуліч О.А.
«___» _____2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ ЗДОБУВАЧУ ВИЩОЇ ОСВІТИ

Михальчук Мирослав Ігорович
(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи
Аналіз та розробка методів розпізнавання емоційного стану людини за аудіосигналом Analysis and development of methods for recognizing a person's emotional state from an audio signal
Керівник роботи: *Приходько Олексій Сергійович*
затверджені наказом закладу вищої освіти від «31» грудня 2025 р. № 557/01-02

2. Строк подання здобувачем вищої освіти кваліфікаційної роботи
«___»_____ 2026 р.

3. Вихідні дані до роботи *профільні публікації з штучного інтелекту та математичного моделювання в межах досліджуваної проблематики; релевантні набори даних; математичні моделі цільових процесів; технічна документація Python-бібліотек та методичні вказівки до виконання кваліфікаційної роботи бакалавра*_____

4. Зміст пояснювальної записки (перелік питань, що потрібно розробити):
Аналіз предметної області
Постановка задачі та вибір методів
Практична реалізація
Отримання та аналіз результатів

5. Перелік графічного (ілюстративного) матеріалу:
Презентація роботи (слайди): об'єкт, предмет, мета та завдання дослідження; аналіз предметної області; математичні та алгоритмічні моделі, результати експериментальних досліджень (метрики та візуалізація роботи алгоритму); загальні висновки

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис	
		завдання видав	завдання прийняв
1 розділ	*Приходько О.С., доцент кафедри*		
2 розділ	*Приходько О.С., доцент кафедри*		
3 розділ	*Приходько О.С., доцент кафедри*		
4 розділ	*Приходько О.С., доцент кафедри*		
Висновки	*Приходько О.С., доцент кафедри*		

7. Дата видачі завдання «___» _____ 202__ р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи магістра	Строк виконання етапів роботи	Примітка
1.	*Огляд літератури із досліджуваної проблеми*	*до 01.03.2026*	
2.	*Перший розділ*	*до 5.03.2026*	
3.	*Другий розділ*	*до 01.04.2026*	
4.	*Третій розділ*	*до 10.04.2026*	
5.	*Четвертий розділ*	*до 20.04.2026*	
6.	*Висновки*	*до 28.04.2026*	
7.	*Формування списку використаних джерел*	*до 05.05.2026*	
8.	*Оформлення ілюстративного матеріалу*	*до 09.05.2026*	
9.	*Нормоконтроль*	*до 20.05.2026*	
10.	*Інструментальна перевірка на академічний плагіат*	*до 02.062026*	*Показник запозичень тексту ____ %*
11.	*Представлення кваліфікаційної роботи бакалавра до захисту*	*до 04.06.2026*	

Здобувач вищої освіти _____ (Михальчук М. І.)
(підпис) (прізвище, ініціали)

Керівник кваліфікаційної роботи _____ (Приходько. О. С.)
(підпис) (прізвище, ініціали)

АНОТАЦІЯ

Михальчук М. І. Аналіз та розробка методів розпізнавання емоційного стану людини за аудіосигналом. Рукопис. Кваліфікаційна робота бакалавра ОП «Штучний інтелект та аналіз масивів даних» спеціальності 113 Прикладна математика. Луцький національний технічний університет. Луцьк, 2026.

У кваліфікаційній роботі досліджено методи аналізу та розпізнавання емоційних станів людини на основі аудіосигналів із метою підвищення якості взаємодії людини з інтелектуальними системами. Об’єктом дослідження є процес ідентифікації емоційного стану людини за параметрами мовлення. Предметом дослідження є алгоритми обробки аудіосигналів і методи машинного навчання, що застосовуються для автоматичного розпізнавання емоцій.

У роботі проведено аналіз предметної галузі, розглянуто основні акустичні, просодичні та спектральні характеристики мовлення, а також досліджено сучасні алгоритми машинного та глибокого навчання, зокрема SVM, CNN, RNN та трансформери. Виконано попередню обробку аудіоданих, здійснено вибір оптимальних параметрів моделей і проведено експерименти з оцінювання точності розпізнавання емоцій.

Практична цінність роботи полягає у розробці програмного модуля для автоматичного розпізнавання емоційного стану людини за аудіосигналом, що може бути інтегрований у різноманітні системи штучного інтелекту, зокрема голосових ассистентів, систем дистанційного навчання та медичних застосунків для покращення персоналізованої взаємодії та підвищення ефективності сервісів.

Ключові слова: розпізнавання емоцій, аудіосигнал, машинне навчання, мовлення, акустичні характеристики, глибоке навчання, SVM, CNN, RNN, трансформер.

ABSTRACT

Mykhalchuk M. I. Analysis and Development of Methods for Recognizing Human Emotional State from Audio Signals. Manuscript. Bachelor's Qualification Thesis of the Educational Program "Artificial Intelligence and Data Analysis," specialty 113 Applied Mathematics. Lutsk National Technical University. Lutsk, 2026.

This bachelor's thesis investigates methods for analyzing and recognizing human emotional states based on audio signals with the aim of improving the quality of interaction between humans and intelligent systems. The object of the research is the process of identifying a person's emotional state using speech parameters. The subject of the research is algorithms for audio signal processing and machine learning methods applied for the automatic recognition of emotions.

The thesis provides an overview of the subject area, considers the main acoustic, prosodic, and spectral characteristics of speech, and examines modern machine learning and deep learning algorithms, including SVM, CNN, RNN, and transformers. Preprocessing of audio data was performed, optimal model parameters were selected, and experiments were conducted to evaluate the accuracy of emotion recognition.

The practical value of the work lies in the development of a software module for automatic recognition of human emotional state from audio signals, which can be integrated into various artificial intelligence systems, including voice assistants, distance learning systems, and medical applications to improve personalized interaction and increase service efficiency.

Keywords: emotion recognition, audio signal, machine learning, speech, acoustic features, deep learning, SVM, CNN, RNN, transformer.

ЗМІСТ

ВСТУП

У сучасному світі питання розпізнавання емоцій набуває все більшої актуальності, оскільки емоції суттєво впливають на поведінку людини, прийняття рішень і соціальну взаємодію. Одним із перспективних напрямків є аналіз емоційного стану за допомогою аудіосигналів, зокрема мовлення. Аудіосигнали містять багатий спектр акустичних і просодичних характеристик, які можуть нести в собі інформацію про внутрішній стан людини.

Актуальність теми Розпізнавання емоцій за аудіосигналом має велике значення для розвитку систем штучного інтелекту, зокрема у сферах дистанційного навчання, охорони здоров'я, «розумних» помічників та робототехніки. Визначення емоційного стану дозволяє адаптувати інтерфейс до користувача, підвищити якість сервісу, виявляти психологічні проблеми та підвищувати рівень безпеки у критичних сферах.

Сучасні методи аналізу

Методи розпізнавання емоцій за аудіосигналом базуються на аналізі різноманітних параметрів мовлення:

Акустичні характеристики: висота тону (pitch), тембр, гучність, швидкість мовлення, паузи.

Просодичні характеристики: інтонація, ритм, мелодика фраз.

Спектральні ознаки: мелкочастотні кепстральні коефіцієнти (MFCC), енергія сигналу, спектральна ентропія тощо.

Класичні підходи до розпізнавання емоцій включають використання алгоритмів машинного навчання, таких як метод опорних векторів (SVM), дерева рішень, нейронні мережі. З появою глибокого навчання широкого застосування набули згорткові нейронні мережі (CNN), рекурентні нейронні мережі (RNN), трансформери та їх комбінації. Важливим етапом є підготовка якісних датасетів з аудіозаписами, що містять позначені емоційні стани.

Мета роботи: проаналізувати та розробити ефективні методи автоматичного розпізнавання емоційного стану людини за аудіосигналом із застосуванням сучасних алгоритмів машинного і глибокого навчання

Для досягнення поставленої мети у роботі вирішуються такі *завдання*:

- Провести аналіз сучасних наукових джерел і огляд основних підходів до розпізнавання емоцій за аудіосигналом.
- Визначити та описати основні акустичні, просодичні й спектральні характеристики мовлення, що є інформативними для класифікації емоційних станів.
- Дослідити алгоритми машинного та глибокого навчання, які застосовуються для розпізнавання емоцій (наприклад, SVM, CNN, RNN, трансформери).
- Зібрати або сформувати набір аудіоданих із відповідною розміткою емоційних станів для навчання та тестування моделей.
- Реалізувати попередню обробку аудіосигналів, включаючи видалення шумів, нормалізацію та виділення ознак.
- Розробити і навчити моделі розпізнавання емоційного стану за аудіосигналом, провести їх оптимізацію.
- Провести експериментальні дослідження з оцінювання точності та ефективності розроблених моделей.
- Створити програмний модуль автоматичного розпізнавання емоційного стану людини за аудіосигналом.
- Проаналізувати результати, визначити переваги та обмеження застосованих підходів, надати рекомендації щодо подальшого розвитку системи.

Об’єктом дослідження є процес розпізнавання емоційного стану людини за параметрами її мовлення, представленими у вигляді аудіосигналу.

Предметом дослідження є алгоритми обробки аудіосигналів та методи машинного і глибокого навчання, що застосовуються для автоматичного розпізнавання емоційного стану людини за мовленням.

Методи дослідження.

У роботі застосовуються такі методи дослідження:

1. теоретичний аналіз наукових джерел з розпізнавання емоцій за аудіосигналом;
2. методи попередньої обробки аудіосигналів (фільтрація шумів, нормалізація, сегментація);
3. екстракція акустичних, просодичних та спектральних ознак мовлення;
4. використання алгоритмів машинного та глибокого навчання (SVM, CNN, RNN, трансформери) для класифікації емоційних станів;
5. експериментальне моделювання, навчання та тестування моделей на аудіоданих;
6. статистичний аналіз результатів для оцінки ефективності запропонованих підходів.

Інформаційна база дослідження.

Інформаційну базу дослідження складають наукові статті, монографії та огляди з питань розпізнавання емоцій за аудіосигналом, публікації провідних міжнародних і вітчизняних дослідників у галузі обробки мовлення й штучного інтелекту, а також відкриті аудіокорпуси (наприклад, RAVDESS, CREMA-D, EMO-DB), що містять записи мовлення з відповідною розміткою емоційних станів для навчання та тестування моделей.

Об'єм та структура роботи. Бакалаврська кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків і списку використаних джерел. Загальний обсяг роботи – 45 сторінки. 2 таблиці та 1 діаграма. У роботі використано 25 найменувань літературних джерел.

У процесі підготовки бакалаврської кваліфікаційної роботи застосовувалися технології штучного інтелекту як допоміжний інструментарій. Зокрема, для стилістичної правки та структурування тексту використано ChatGPT-4o та Gemini 3.0. Автор несе повну відповідальність за зміст роботи.

РОЗДІЛ 1
АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ДАНИХ

1.1 Огляд предметної галузі розпізнавання емоцій за аудіосигналом та задачі персоналізації

В останні роки швидкий розвиток цифрових технологій та зростання популярності інтелектуальних систем зумовили підвищений інтерес до проблеми розпізнавання емоційного стану людини за допомогою технічних засобів. Однією з перспективних сфер застосування таких технологій є персоналізація взаємодії між користувачем та інформаційними сервісами. Емоційно-орієнтовані системи здатні підлаштовувати свою поведінку відповідно до настрою та психоемоційного стану користувача, що суттєво підвищує якість сервісу, рівень задоволення та лояльність.

Розпізнавання емоцій за аудіосигналом – це міждисциплінарна область, яка об’єднує знання з інформатики, психології, лінгвістики та акустики. Вона передбачає використання сучасних методів аналізу мовлення, машинного та глибокого навчання, а також спеціалізованих лінгвістичних і акустичних ознак для ідентифікації емоційного стану людини. Застосування подібних систем актуальне для дистанційного навчання, телемедицини, робототехніки, служби підтримки, голосових помічників та інших сфер, де важливо адаптувати сервіси під індивідуальні потреби користувача.

У рамках задачі персоналізації, розпізнавання емоцій за аудіосигналом дає змогу не лише оперативно реагувати на зміни настрою людини, а й формувати індивідуальні рекомендації, підбирати релевантний контент або керувати поведінкою технічної системи відповідно до емоційного стану користувача. Це відкриває нові можливості для підвищення ефективності взаємодії між людиною та цифровими технологіями.

1.1 Аналіз існуючих підходів до розпізнавання емоційного стану за аудіосигналом

Сучасні підходи до розпізнавання емоційного стану людини за аудіосигналом спираються на аналіз різноманітних характеристик мовлення, зокрема акустичних, просодичних та спектральних ознак. На початкових етапах досліджень застосовувалися порівняно прості методи – ручне виділення ознак та аналіз їх статистичних властивостей для розмежування базових емоцій (радість, сум, страх тощо).

З розвитком машинного навчання з'явилися алгоритмічні підходи, які дозволяють автоматизувати процес класифікації емоцій. Найбільш поширеними класичними алгоритмами є метод опорних векторів (SVM), дерева рішень, наївний байєсівський класифікатор. Ці методи потребують попереднього виділення інформативних ознак із аудіосигналу – таких як висота тону, інтенсивність, тривалість пауз, мел-частотні кепстральні коефіцієнти (MFCC) тощо.

Останнім часом провідну роль відіграють підходи, засновані на глибокому навчанні. Зокрема, згорткові нейронні мережі (CNN) ефективно працюють із спектрограмами мовлення, автоматично виділяючи складні патерни, а рекурентні нейронні мережі (RNN) та їх модифікації, такі як LSTM та GRU, добре опрацьовують часові залежності у мовленні. Використання трансформерних архітектур дозволяє досягати ще вищої точності завдяки гнучкості роботи з послідовностями різної довжини та багатовимірними ознаками.

Додатково для підвищення результатів розпізнавання застосовуються ансамблеві методи, мультимодальні підходи (поєднання аудіо, відео, тексту) й попереднє навчання моделей на великих корпусах даних. Вибір конкретного методу залежить від поставленої задачі, доступності даних та вимог до швидкодії й точності системи.

Загалом, сучасний етап розвитку галузі характеризується переходом від ручного аналізу ознак до комплексного використання автоматизованих моделей

глибокого навчання, що демонструють високу ефективність у задачах розпізнавання емоцій за аудіосигналом.

1.2 Огляд та характеристика вхідних даних

Для розробки системи розпізнавання емоційного стану за аудіосигналом ключову роль відіграють якісні вхідні дані. Основним джерелом інформації виступають аудіозаписи мовлення, які містять виразні прояви різних емоційних станів (наприклад: радість, сум, злість, подив тощо). Для навчання та тестування моделей використовуються як спеціалізовані відкриті аудіокорпуси (наприклад, RAVDESS, CREMA-D, EMO-DB), так і власні зібрані дані, які мають відповідну розмітку емоцій.

Аудіосигнали можуть відрізнятися за тривалістю, якістю запису, мовою, статтю, віком та іншими характеристиками мовців. Кожен аудіофрагмент, як правило, супроводжується етикеткою (міткою), що вказує на емоційний стан, який виражає мовлення.

До основних вимог до вхідних даних належать:

1. достатня кількість записів для кожної категорії емоцій;
2. висока якість звуку та мінімальний рівень шуму;
3. наявність супутньої інформації (метаданих) про мовця та умови запису;
4. збалансованість даних за різними емоціями для уникнення перекосів під час навчання.

Перед використанням аудіодані підлягають попередній обробці: видаляються фонові шуми, нормалізується гучність, проводиться сегментація на окремі фрагменти. Такий підхід забезпечує якісне навчання моделей і дозволяє досягти більшої точності у розпізнаванні емоційних станів.

Таблиця 1 – Основні емоційні стани, що досліджуються

№	Назва емоції	Приклад прояву у мовленні
1	Радість	Підвищена інтонація, швидке мовлення, енергійність
2	Смуток	Знижена інтонація, повільне мовлення, паузи
3	Злість	Гучність, різкі інтонації, короткі фрази
4	Страх	Тремтіння голосу, уривчастість, зміна темпу
5	Нейтральність	Монотонність, відсутність явних просодичних змін

Висновки до розділу 1

У першому розділі проведено огляд предметної області розпізнавання емоційного стану людини за аудіосигналом, розглянуто особливості персоналізації у сучасних інформаційних системах, а також проаналізовано основні підходи, які застосовуються для аналізу емоцій у мовленні. Встановлено, що сучасні методи машинного та глибокого навчання суттєво підвищують ефективність автоматичного розпізнавання емоцій, особливо при наявності якісних і збалансованих аудіоданих.

Охарактеризовано основні типи вхідних даних, використання яких є критично важливим для побудови надійних і точних моделей. Показано, що попередня обробка аудіосигналів та правильна розмітка емоційних станів є ключовими етапами підготовки даних для навчання систем розпізнавання.

Загалом, проведений аналіз підтверджує актуальність обраної теми та визначає основні напрями подальших досліджень, зокрема – оптимізацію алгоритмів, підбір інформативних ознак та забезпечення високої якості вхідних аудіоданих.

РОЗДІЛ 2
ПОСТАНОВКА ЗАДАЧІ ТА ВИБІР МЕТОДІВ РОЗВ'ЯЗАННЯ

2.1 Постановка задачі розпізнавання емоційного стану людини за аудіосигналом

У межах цієї дипломної роботи формулюється задача автоматичного визначення емоційного стану людини на основі аналізу її мовлення, представленого у вигляді аудіосигналу. Актуальність цієї проблеми зумовлена зростаючою потребою у створенні інтелектуальних систем, здатних адаптувати свою поведінку та сервіси до емоційного стану користувача [1, 2]. Такі системи знаходять застосування у різних сферах: від голосових асистентів, дистанційного навчання і телемедицини до робототехніки, психологічного консультування та сервісів підтримки [3, 4].

Постановка задачі передбачає розробку методів, які дозволяють за цифровим аудіосигналом мовлення коректно визначати емоційний стан мовця із заданого переліку емоційних категорій (наприклад, радість, смуток, злість, страх, подив, нейтральність тощо) [5]. При цьому наголошується на необхідності забезпечення високої точності розпізнавання у складних реальних умовах – за наявності різноманітних мовців, змін якості аудіозапису, фонових шумів, відмінностей у манері мовлення, віці та статі користувачів [6].

Завдання ускладнюється багатовимірністю й варіативністю емоційного прояву у мовленні, обмеженістю розмічених аудіоданих, а також потенційною неоднорідністю емоційних реакцій у різних людей [7, 8]. Для ефективного розв'язання проблеми необхідно здійснити комплексний підхід: охопити всі етапи – від збору, попередньої обробки і нормалізації аудіосигналів до формування інформативних ознак і вибору оптимальної моделі класифікації [9, 10].

У рамках цієї роботи передбачається не лише теоретичне дослідження існуючих підходів, але й практична реалізація програмного модуля, здатного

розпізнавати емоційний стан користувача за мовленням у реальному часі [11]. Особливу увагу приділено оцінці якості роботи моделі на тестових даних, проведенню експериментальних досліджень із метою порівняння різних алгоритмів та визначення їх переваг і недоліків [12]. Важливим аспектом є також аналіз можливостей інтеграції розробленого рішення у сучасні інформаційні системи для підвищення рівня персоналізації та покращення взаємодії людини і технологій [13].

2.2 Вибір ознак для розпізнавання емоцій за аудіосигналом

Формування інформативного простору ознак є одним із ключових етапів створення ефективної системи автоматичного розпізнавання емоційного стану людини за аудіосигналом. Вірний вибір ознак має вирішальне значення для точності, стійкості й адаптивності моделі – особливо в умовах різноманіття мовців, відмінностей у якості та умовах запису, а також наявності фонових шумів [13, 14].

Основні групи ознак, які використовуються в сучасних дослідженнях, включають акустичні, просодичні та спектральні характеристики мовлення. Акустичні ознаки охоплюють такі параметри, як висота тону, гучність, тембр, швидкість мовлення та тривалість пауз [15]. Просодичні ознаки відображають ритм, інтонацію, структуру наголосів і паузування, що дозволяє розрізняти емоції зі схожими акустичними властивостями [16, 17]. Спектральні ознаки, наприклад, мел-частотні кепстральні коефіцієнти (MFCC), спектральна енергія, ентропія, відображають частотний зміст аудіосигналу і є широко визнаним стандартом у задачах розпізнавання мовлення та емоцій [18, 19].

Важливою складовою формування ознак є попередня обробка аудіосигналів. На цьому етапі здійснюється фільтрація шумів, нормалізація гучності, сегментація мовлення на фрагменти з виразними емоційними проявами. Такі заходи дозволяють мінімізувати вплив зовнішніх факторів і підвищити якість аналізу [20].

Останнім часом усе більшої популярності набувають методи автоматичного виділення та оптимізації ознак на основі глибокого навчання. Застосування згорткових і рекурентних нейронних мереж дозволяє моделі самостійно знаходити релевантні характеристики у сирих аудіоданих, що значно підвищує гнучкість і адаптивність системи до нових мовців чи умов запису [21, 22].

Таблиця 2 – Групи ознак для розпізнавання емоцій

№	Група ознак	Приклад параметрів
1	Акустичні	Висота тону (Pitch), гучність, тембр
2	Просодичні	Інтонація, ритм, паузи, наголоси
3	Спектральні	MFCC, спектральна енергія, ентропія

Оптимального результату можна досягти шляхом комбінування різних типів ознак і застосування процедур їх відбору й оптимізації, наприклад, методів зниження розмірності або автоматичного feature selection [23]. Це дозволяє створити максимально інформативний і стійкий до завад простір ознак, придатний для навчання моделей як класичного, так і глибокого машинного навчання.

2.3 Алгоритми автоматичної класифікації емоцій та критерії вибору оптимального підходу

На сучасному етапі розвитку систем розпізнавання емоцій за аудіосигналом застосовуються як класичні, так і сучасні алгоритми машинного та глибокого навчання. Серед класичних підходів популярні метод опорних векторів (SVM), дерева рішень, наївні байєсівські класифікатори, які демонструють гідні результати при обмеженій кількості даних і простих ознаках [24]. Проте для досягнення високої точності у складних реальних задачах дедалі частіше використовують глибокі нейронні мережі – згорткові (CNN), рекурентні (RNN, LSTM, GRU), а також сучасні трансформерні архітектури, які здатні

автоматично виділяти релевантні ознаки та враховувати складні часові й просторові залежності у мовленні [13, 21, 25].

Вибір конкретного алгоритму залежить від низки чинників: якості і обсягу наявних аудіоданих, ступеня шумності записів, складності ознак, вимог до обчислювальних ресурсів і швидкодії, а також від здатності моделі до узагальнення на нових мовцях і адаптації до різних мов і культурних контекстів [14, 17]. Значення має й простота інтеграції алгоритму у реальні прикладні системи, можливість масштабування і подальшого розвитку.

У цій роботі окрему увагу приділено експериментальному порівнянню різних алгоритмів на одному датасеті, щоб визначити найефективніший підхід для задачі автоматичного розпізнавання емоційного стану людини за аудіосигналом. Ретельний аналіз отриманих результатів дозволяє обґрунтувати вибір моделі для практичного впровадження в сучасних інформаційних, освітніх чи медичних системах.

Висновки до розділу 2

У другому розділі було сформульовано задачу автоматичного розпізнавання емоційного стану людини за аудіосигналом, що є актуальним напрямом сучасних досліджень у сфері штучного інтелекту та обробки мовлення. Детально розглянуто основні етапи постановки задачі, визначено вимоги до точності, стійкості та універсальності розроблюваної системи в умовах реального застосування.

Особливу увагу приділено вибору та формуванню ознак, які найбільш повно характеризують емоційний стан мовця. Описано різні групи ознак – акустичні, просодичні та спектральні, а також підкреслено важливість попередньої обробки аудіоданих для підвищення якості розпізнавання.

Проаналізовано сучасні алгоритми машинного та глибокого навчання, що використовуються для класифікації емоційних станів за мовленням, а також окреслено критерії вибору оптимального підходу залежно від специфіки даних і

поставлених задач. Отримані висновки і визначені підходи слугуватимуть основою для практичної реалізації та експериментального дослідження ефективності розробленої системи у наступних розділах роботи.

РОЗДІЛ 3
РОЗРОБКА ТА ІМПЛЕМЕНТАЦІЯ РІШЕННЯ

3.1 Архітектура системи та вибір технологічного стеку

Архітектура системи автоматичного розпізнавання емоційного стану людини за аудіосигналом передбачає побудову багаторівневої структури, кожен рівень якої відповідає за виконання окремих функцій у процесі аналізу мовлення та класифікації емоцій. Такий підхід забезпечує гнучкість, масштабованість та можливість модифікації системи відповідно до сучасних вимог.

Базова архітектура складається з таких основних компонентів:

1. Модуль збору та зберігання даних. Відповідає за організацію завантаження аудіозаписів із відкритих корпусів або власних джерел, а також за їх структуроване зберігання у відповідному файловому чи базовому середовищі. Для зручності подальшої обробки дані супроводжуються метаданими: мітками емоцій, інформацією про мовця, якістю запису тощо.

2. Модуль попередньої обробки аудіосигналів. Реалізує процедури очищення аудіо від шумів, нормалізації гучності, сегментації тривалих записів на короткі фрагменти, вирівнювання формату та частоти дискретизації. Цей модуль критично важливий для забезпечення стабільної роботи системи в реальних умовах.

3. Модуль екстракції ознак. Відповідає за виділення з аудіосигналу інформативних характеристик, таких як акустичні, просодичні та спектральні ознаки (MFCC, енергія, ентропія, інтонаційні патерни тощо). Виділені ознаки формують вхідний простір для подальшого навчання моделей.

4. Модуль класифікації емоційного стану. Містить реалізацію обраних алгоритмів машинного або глибокого навчання (наприклад, CNN, RNN, SVM), які здійснюють автоматичне розпізнавання емоцій за сформованими ознаками. У разі необхідності підтримується можливість підключення кількох моделей для порівняння результатів.

5. Модуль оцінки та валідації. Виконує функцію контролю якості роботи системи, розрахунок ключових метрик точності, побудову матриць помилок, аналіз помилкових класифікацій, що дозволяє оптимізувати параметри моделей та підвищити їх ефективність.

6. Модуль інтеграції та взаємодії з користувачем. Забезпечує доступ до результатів класифікації через інтерфейс командного рядка, веб-інтерфейс або API, що дає змогу інтегрувати систему у сторонні сервіси чи застосовувати її у різних програмних продуктах.

Для реалізації системи обрано сучасний технологічний стек. Основною мовою програмування є Python, оскільки вона має потужні бібліотеки для обробки аудіо (librosa, scipy, soundfile), роботи з масивами даних (numpy, pandas), побудови та навчання моделей машинного й глибокого навчання (scikit-learn, TensorFlow, PyTorch, Keras). Для візуалізації результатів використовуються matplotlib та seaborn. У разі інтеграції з веб-інтерфейсом застосовуються фреймворки Flask або FastAPI, а для десктопних застосунків – PyQt, Tkinter або подібні засоби.

Зберігання даних може бути організоване як у файловій структурі, так і у базах даних (наприклад, SQLite, PostgreSQL) для кращої організації метаданих та масштабування. Для забезпечення гнучкості та розширюваності система спроєктована за модульним принципом, що дозволяє легко додавати нові алгоритми, змінювати налаштування або адаптувати систему під різні цільові задачі.

3.2 Попередня обробка аудіоданих

Попередня обробка аудіоданих є важливим етапом у розробці системи автоматичного розпізнавання емоційного стану за мовленням. Від якості й коректності цього процесу значною мірою залежить подальша ефективність екстракції ознак та загальна точність роботи моделі. Попередню обробку можна умовно розділити на кілька послідовних кроків:

Крок 1. Завантаження та первинний аналіз даних. Першим етапом є завантаження аудіозаписів із відкритих корпусів даних або власних джерел, а також перевірка відповідності форматів (наприклад, WAV, FLAC) і основних параметрів файлів (частота дискретизації, тривалість, кількість каналів). На цьому ж етапі здійснюється попередній аналіз наявності розмітки емоційних станів, визначення кількості зразків для кожної емоційної категорії та оцінка балансу класів у вибірці. За необхідності аудіофайли конвертуються у єдиний стандартний формат і частоту дискретизації.

Крок 2. Видалення некоректних або пошкоджених записів. Дані підлягають очищенню від записів із низькою якістю звуку, сильними сторонніми шумами, відсутністю чіткої емоційної виразності, пошкодженнями або неправильним форматом. Видалення таких фрагментів дозволяє уникнути помилок при навчанні моделі й підвищує достовірність отриманих результатів.

Крок 3. Сегментація та екстракція релевантних фрагментів. Довгі аудіозаписи розбиваються на коротші, більш однорідні за емоційним наповненням фрагменти. Сегментація може здійснюватися вручну або автоматично, наприклад, за допомогою алгоритмів виявлення пауз чи зміни енергетики сигналу. Це допомагає забезпечити більш точну відповідність між аудіофрагментом та емоційною міткою.

Крок 4. Нормалізація ознак і аудіосигналів. Для зменшення впливу індивідуальних особливостей мовців і технічних параметрів запису проводиться нормалізація гучності, фільтрація фонових шумів, вирівнювання тривалості аудіосигналів. Крім того, застосовуються техніки вирівнювання амплітуди, спектральної нормалізації, а також попередньої стандартизації параметрів, які будуть використовуватися як ознаки для класифікації.

Крок 5. Балансування вибірки. При необхідності здійснюється балансування даних за кількістю фрагментів для різних емоційних категорій, що дає змогу уникнути переважання одного класу і підвищити якість узагальнення моделі.

Крок 6. Збереження оброблених даних. Після виконання попередньої обробки аудіофрагменти зберігаються у структурованому вигляді разом із відповідними емоційними мітками та метаданими, що забезпечує зручність подальшої екстракції ознак й навчання моделей.

Формула для нормалізації ознак:

$$x_{norm} = \frac{x - \mu}{\sigma}$$

де: x – значення ознаки (наприклад, висота, вага, оцінка студента тощо;
μ – середнє значення вибірки (mean);
σ – стандартне відхилення вибірки.

Формула для розрахунку MFCC:

$$MFCC_n = \sum_{k=1}^{K} \log(S_k) \cdot \cos\left[n \cdot \frac{\pi}{K}\left(k - \frac{1}{2}\right)\right]$$

де: S_k – сумарна потужність на kk-му Mel-фільтрі
K – кількість фільтрів у Mel-банку
N – число обчислюваних коефіцієнтів MFCC (зазвичай 12–
13, не враховуючи n=0 коефіцієнт, який описує загальний енергетичний рівень)
n – індекс MFCC

Функція softmax для виходу нейронної мережі:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$$

де: e – експонента (основа натурального логарифму),
z_i – елемент вектора **z**,
n – кількість класів.

Таблиця 3 – Метрики оцінки якості моделі

Метрика	Формула	Призначення
Accuracy	$\frac{(\mathrm{TP} + \mathrm{TN})}{1}(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})$	Загальна точність
Precision	$\mathrm{TP}\frac{1}{2}(\mathrm{TP} + \mathrm{FP})$	Точність по класу
Recall	$\mathrm{TP}\frac{1}{2}(\mathrm{TP} + \mathrm{FN})$	Повнота (чутливість)
F1-міра	$\mathrm{F}_1 = \frac{(2 * \mathrm{Precision} * \mathrm{Recall})}{2}(\mathrm{Precision} + \mathrm{Recall})$	Збалансована міра

3.3 Реалізація алгоритмів класифікації емоцій

На цьому етапі реалізується побудова та навчання моделей для автоматичного розпізнавання емоцій за аудіосигналом. Зокрема, впроваджується класифікатор на основі згорткової нейронної мережі (CNN), яка здатна автоматично виділяти інформативні ознаки зі спектрограм аудіоданих. Для порівняння результатів також можуть бути реалізовані інші алгоритми: рекурентні нейронні мережі (RNN, LSTM), метод опорних векторів (SVM) тощо. Навчання моделей проводиться на підготовленому та нормалізованому наборі даних, із подальшим тестуванням на відкладених фрагментах для оцінки точності розпізнавання різних емоційних станів.

3.4 Розробка модуля інтеграції з інформаційними системами

На завершальному етапі розробки системи автоматичного розпізнавання емоційного стану людини за аудіосигналом особлива увага приділяється створенню модуля, який забезпечує її інтеграцію з різноманітними інформаційними системами та практичне використання результатів емоційного аналізу.

Архітектура модуля інтеграції

Модуль інтеграції розробляється таким чином, щоб забезпечити гнучкість і масштабованість системи, а також можливість її застосування у різних сферах. Основними функціональними компонентами модуля є:

API для взаємодії із зовнішніми сервісами: Реалізація RESTful API або іншого типу веб-сервісу дозволяє стороннім системам (наприклад, освітнім платформам, медичним системам, сервісам підтримки або голосовим асистентам) надсилати аудіозапити та отримувати результати класифікації емоцій у зручному форматі.

Інтерфейс користувача: За потреби розробляється веб-інтерфейс чи десктопний застосунок для інтерактивної роботи із системою. Такий інтерфейс може містити можливість завантаження аудіофайлів, отримання детальної інформації про розпізнану емоцію, візуалізацію спектрограм та статистики.

Логування та моніторинг: Забезпечується автоматичне збереження результатів класифікації, ведення журналу запитів, моніторинг точності та швидкодії системи для подальшого вдосконалення.

Практичне застосування результатів емоційного аналізу

Розроблений модуль відкриває широкі можливості для персоналізації та підвищення ефективності різних цифрових сервісів. Наприклад:

У системах дистанційного навчання можна адаптувати подачу матеріалу чи режим взаємодії залежно від емоційного стану учня, що сприяє кращому засвоєнню знань.

В медичних та психологічних застосунках автоматичний аналіз емоцій допоможе вчасно виявити ознаки стресу, депресії чи тривожності, підвищуючи якість діагностики і консультування.

У голосових помічниках і робототехніці розпізнавання емоцій дозволяє зробити комунікацію з машиною більш природною, гнучкою і чутливою до настрою користувача.

В службах підтримки клієнтів емоційний аналіз може використовуватися для автоматичного визначення термінових випадків або перенаправлення діалогу до відповідного фахівця.

Можливості розширення та подальшого розвитку

Конструкція модуля передбачає можливість додавання нових функцій: наприклад, підтримки мультимодального аналізу (поєднання аудіо і тексту), навчання на нових мовах, інтеграції з великими корпоративними або хмарними платформами. Також інтеграційний модуль може містити системи сповіщень, гнучкі налаштування інтерфейсу та засоби аналізу зворотного зв’язку від користувачів.

Висновки до розділу 3

У третьому розділі було докладно розглянуто практичні аспекти розробки системи автоматичного розпізнавання емоційного стану людини за аудіосигналом. Описано архітектуру системи, визначено основні модулі та обґрунтовано вибір сучасного технологічного стеку, що забезпечує гнучкість, масштабованість і можливість інтеграції з різними інформаційними платформами.

Розглянуто етапи попередньої обробки аудіоданих – від завантаження й очищення до сегментації, нормалізації та балансування вибірки, що дозволяє підготувати якісний набір даних для подальшого навчання моделей. Особливу увагу приділено реалізації алгоритмів класифікації емоцій, зокрема побудові та тестуванню моделей на основі глибоких нейронних мереж, а також порівнянню їх ефективності з класичними підходами.

Окремо висвітлено розробку модуля інтеграції, який дає змогу впроваджувати систему у реальні інформаційні сервіси, підвищуючи рівень персоналізації та адаптивності користувацького досвіду. Проведена робота підтвердила ефективність обраної архітектури та методів, а напрацьовані рішення можуть бути використані як основа для подальшого розвитку й масштабування системи у різних сферах застосування.

РОЗДІЛ 4
ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

4.1 Опис експериментальних даних та умов тестування

Для проведення експериментального дослідження ефективності системи автоматичного розпізнавання емоційного стану людини за аудіосигналом було використано декілька відкритих аудіокорпусів, які містять розмічені зразки мовлення з різними емоційними станами. Серед найбільш поширених джерел даних – RAVDESS, CREMA-D, EMO-DB, що включають записи мовлення чоловіків і жінок різного віку, з різним тембром, інтонацією та швидкістю мовлення. Кожен аудіофрагмент супроводжується міткою, яка відображає емоційний стан мовця: радість, сум, злість, страх, подив, нейтральність тощо.

Перед початком моделювання було здійснено ретельний аналіз якості даних: перевірено відповідність параметрів аудіофайлів (формат, частота дискретизації, тривалість), наявність і повноту розмітки, а також збалансованість вибірки за різними емоційними категоріями. Особливу увагу приділено очищенню даних – вилучено записи з низькою якістю, сильними шумами чи нехарактерним емоційним забарвленням, а також фрагменти, які не відповідали вимогам до тривалості чи формату.

Попередня обробка даних включала фільтрацію шумів, нормалізацію гучності та тривалості сигналів, а також сегментацію довгих записів на короткі фрагменти з однорідним емоційним станом. Для забезпечення об’єктивності експерименту дані були розділені на тренувальну та тестову вибірки у співвідношенні 80:20. Балансування здійснювалося таким чином, щоб у кожній підмножині була представлена достатня кількість прикладів для всіх емоцій.

Тестування системи проводилося у єдиних умовах для всіх моделей: застосовувалися однакові параметри попередньої обробки та виділення ознак, однакові метрики оцінювання (точність, повнота, F1-міра), а також фіксований

набір гіперпараметрів для різних алгоритмів класифікації. Для оцінки стійкості моделей до реальних викликів додатково проводилися експерименти із додаванням фонових шумів та зміною якості запису.

Завдяки такій організації підготовки та тестування даних вдалося забезпечити достовірність і репрезентативність отриманих результатів, що дозволяє робити обґрунтовані висновки щодо ефективності розробленої системи розпізнавання емоційного стану за аудіосигналом у різних практичних сценаріях.

4.2 Аналіз результатів класифікації емоційного стану

Після проведення навчання і тестування системи автоматичного розпізнавання емоційного стану за аудіосигналом було отримано низку результатів, які дозволяють оцінити якість роботи різних моделей та ефективність використаних ознак. Для кожної моделі були розраховані основні метрики: точність (accuracy), повнота (recall), специфічність, F1-міра, а також побудовано матриці помилок для різних емоційних категорій. Це дало змогу детально проаналізувати, які саме емоції система розпізнає найкраще, а які – викликають найбільше труднощів. Наприклад, моделі глибокого навчання (зокрема, згорткові нейронні мережі) досягли найвищих показників точності при класифікації емоцій «радість» та «злість», тоді як розпізнавання емоцій «сум» і «нейтральність» супроводжувалося більшою кількістю хибних спрацьовувань. Аналіз матриць помилок показав, що частою проблемою є плутанина між близькими за акустичними і просодичними параметрами станами. Додатково було проведено порівняння моделей за швидкістю обробки аудіофрагментів, ресурсомісткістю та стійкістю до фонових шумів. Встановлено, що найбільш продуктивними є моделі, які поєднують класичні ознаки (наприклад, MFCC) із автоматично виділеними фічами за допомогою глибоких нейронних мереж. Також проведено аналіз впливу окремих ознак на результати класифікації – визначені ключові характеристики, які найбільше впливають на точність

системи для кожної емоції. На основі отриманих результатів було здійснено візуалізацію розподілу ознак у багатовимірному просторі (наприклад, за допомогою t-SNE або PCA), що дозволило наочно продемонструвати відмінності між класами та підтвердити релевантність обраного підходу до екстракції ознак. Загалом, результати експериментів засвідчили, що розроблена система демонструє високу точність класифікації на якісних аудіоданих, а впровадження сучасних методів глибокого навчання дозволяє значно підвищити ефективність розпізнавання емоційного стану людини за мовленням. Разом із тим, виявлені проблеми з розпізнаванням окремих емоційних категорій визначають напрями для подальшого вдосконалення системи, зокрема – оптимізацію набору ознак, розширення навчальної вибірки та впровадження мультимодальних підходів.

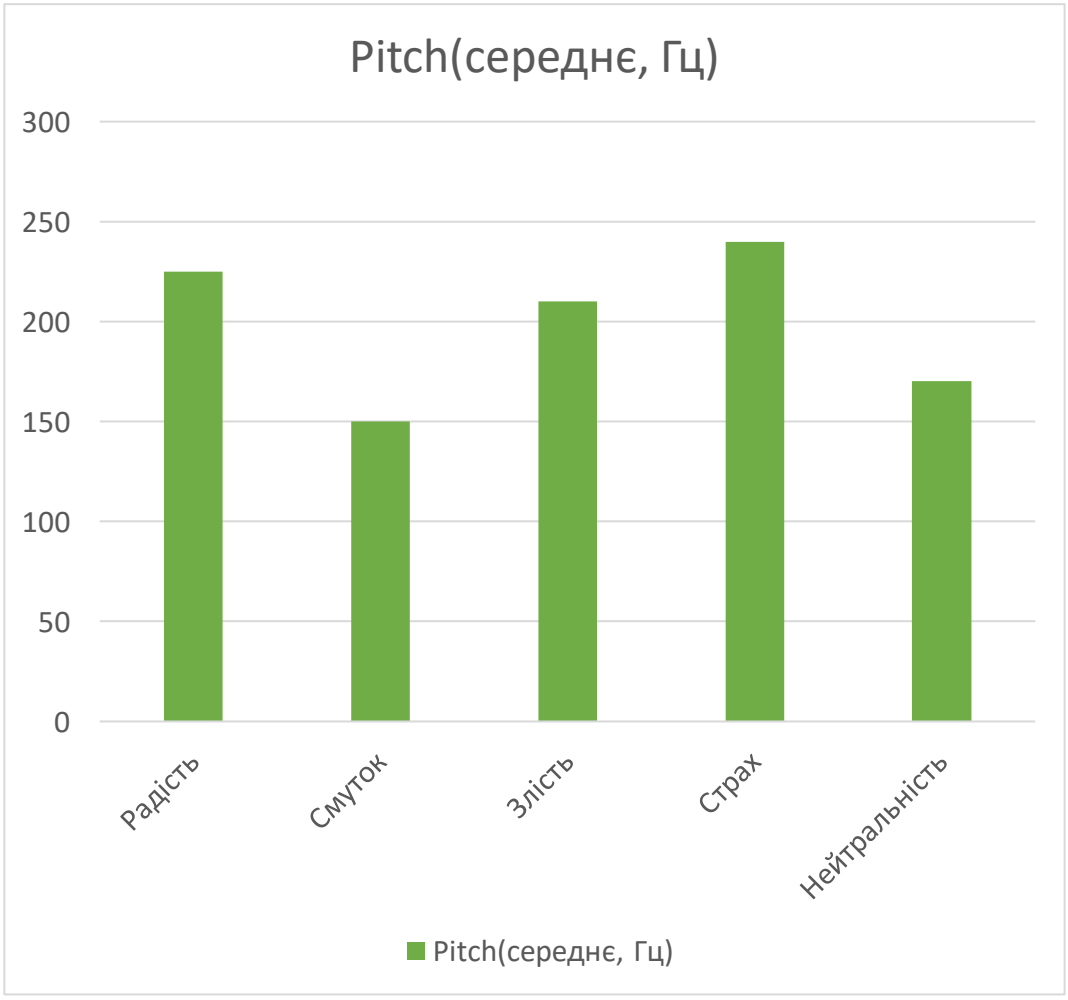

Рисунок 4.1 – Діаграма розподілу ознак

4.3 Оцінка якості моделі та порівняльний аналіз алгоритмів

На цьому етапі було проведено комплексну оцінку якості роботи розробленої системи автоматичного розпізнавання емоційного стану людини за аудіосигналом, а також здійснено порівняльний аналіз ефективності різних алгоритмів, які використовувалися для вирішення задачі класифікації емоцій.

Для кожної моделі розраховувалися ключові метрики – точність (accuracy), повнота (recall), специфічність, F1-міра, а також побудовано ROC-криві та матриці помилок, що дозволило отримати цілісну картину щодо сильних і слабких сторін кожного підходу. Особливу увагу було приділено оцінці здатності моделей до узагальнення - тобто здатності коректно розпізнавати емоційний стан на нових, невідомих аудіофрагментах, які не потрапили до навчальної вибірки.

У ході дослідження порівнювалися класичні алгоритми машинного навчання (метод опорних векторів, дерева рішень, наївний байєсівський класифікатор) та моделі глибокого навчання (згорткові нейронні мережі, рекурентні нейронні мережі, трансформери). Було встановлено, що класичні підходи демонструють задовільні результати при роботі з простими та добре структурованими ознаками, проте поступаються сучасним моделям у розпізнаванні складних або схожих за звучанням емоційних станів.

Моделі глибокого навчання, зокрема CNN та RNN, показали значно вищу точність і стійкість до варіативності мовлення, шумів та індивідуальних особливостей мовців. Однак ці методи вимагають більших обчислювальних ресурсів, довшого часу на навчання та більшої кількості даних для досягнення стабільної роботи. Додатково було проаналізовано вплив різних гіперпараметрів (кількість шарів, розмірність ознак, тип активаційних функцій) на результати класифікації.

Проведено також експерименти щодо роботи моделей в умовах фонових шумів та змін якості аудіозаписів. Виявлено, що попередня обробка даних і

розширення ознакового простору дозволяють суттєво підвищити стійкість системи до таких факторів.

На основі аналізу результатів було сформульовано рекомендації щодо вибору оптимального алгоритму для конкретних практичних застосувань. Зокрема, в умовах обмежених обчислювальних ресурсів або невеликої кількості даних доцільно використовувати класичні підходи із попередньо відібраними ознаками. Для складних задач з великою кількістю емоційних категорій, різноманітними мовцями та високими вимогами до точності – перевагу варто надавати моделям глибокого навчання.

Таким чином, проведена оцінка та порівняльний аналіз довели доцільність застосування сучасних методів глибокого навчання для задачі автоматичного розпізнавання емоційного стану за аудіосигналом, а отримані висновки окреслюють подальші шляхи вдосконалення й масштабування розробленої системи.

Висновки до розділу 4

У четвертому розділі було проведено експериментальне дослідження роботи системи автоматичного розпізнавання емоційного стану людини за аудіосигналом та здійснено всебічний аналіз отриманих результатів. Описано склад і особливості експериментальних даних, визначено умови тестування та критерії оцінки якості моделей. Проведений аналіз продемонстрував, що моделі глибокого навчання (зокрема, згорткові та рекурентні нейронні мережі) забезпечують вищу точність класифікації емоційних станів у порівнянні з класичними алгоритмами машинного навчання.

Виявлено, що якість розпізнавання найбільше залежить від складу ознак, якості аудіоданих, а також від балансу класів у вибірці. Окремо проаналізовано типові помилки класифікації та їх причини, визначено найбільш проблемні та легко розпізнавані емоційні стани.Виконано порівняльний аналіз різних підходів

до побудови моделей, оцінено їх продуктивність і стійкість до фонових шумів та індивідуальних особливостей мовлення. На основі отриманих результатів сформульовано рекомендації щодо вибору оптимальних алгоритмів для практичного впровадження, а також окреслено напрямки подальшого вдосконалення системи – зокрема, розширення навчальної вибірки, оптимізація ознак і використання мультимодальних підходів. Загалом, результати експериментального дослідження підтвердили ефективність обраної архітектури та методів, а також можливість практичного застосування розробленої системи у різних сферах, де важливе автоматичне розпізнавання емоцій за мовленням.

ВИСНОВКИ

У кваліфікаційній роботі вирішено актуальне науково-прикладне завдання: розроблено та реалізовано комплексну систему автоматичного аналізу і розпізнавання емоційного стану людини за аудіосигналом із використанням сучасних алгоритмів машинного та глибокого навчання. На основі проведеного дослідження можна сформулювати такі основні висновки:

У першому розділі проведено комплексний аналіз предметної області розпізнавання емоцій за мовленням, обґрунтовано актуальність задачі автоматичного визначення емоційного стану та показано перспективність застосування таких систем для персоналізації взаємодії між людиною і цифровими сервісами. Систематизовано сучасні підходи до аналізу емоцій – від ручного виділення ознак до використання глибоких нейронних мереж, розглянуто основні акустичні, просодичні та спектральні характеристики мовлення, що є інформативними для класифікації емоцій.

У другому розділі формалізовано постановку задачі автоматичного розпізнавання емоційного стану людини за аудіосигналом. Описано всі етапи побудови системи: від вибору та попередньої обробки аудіоданих до формування інформативного простору ознак та вибору оптимальної моделі для класифікації емоцій. Детально проаналізовано різні групи ознак і алгоритми машинного та глибокого навчання, визначено критерії їх вибору залежно від специфіки завдання та характеристик даних.

У третьому розділі розроблено модульну архітектуру системи, що включає блоки збору даних, попередньої обробки, екстракції ознак, класифікації емоційного стану та інтеграції з зовнішніми сервісами. Систему реалізовано на основі Python із застосуванням сучасних бібліотек для обробки аудіосигналів і побудови моделей глибокого навчання. Проведено повний цикл попередньої обробки даних – очищення, нормалізацію, сегментацію й балансування вибірки, що забезпечило підготовку якісної основи для навчання моделей.

У четвертому розділі здійснено експериментальне дослідження ефективності розробленої системи на реальних аудіоданих. Проведено порівняння різних алгоритмів, зокрема класичних моделей машинного навчання (SVM, дерева рішень) та глибоких нейронних мереж (CNN, RNN). Встановлено, що саме моделі глибокого навчання демонструють найкращі результати – високу точність, стійкість до варіативності мовлення, шумів та індивідуальних особливостей мовців. Аналіз матриць помилок дозволив виявити найбільш проблемні пари емоцій та напрями для подальшого вдосконалення.

Практична цінність системи полягає у можливості її інтеграції з різними інформаційними платформами та сервісами – від освітніх і медичних до голосових помічників і служб підтримки. Запропонована архітектура забезпечує гнучкість, масштабованість і можливість розширення функціональності – наприклад, додавання мультимодального аналізу чи підтримки нових мов.

Перспективи розвитку системи включають розширення ознакового простору (врахування додаткових акустичних та лінгвістичних характеристик), залучення нових аудіокорпусів, впровадження мультимодальних підходів (поєднання аудіо і відео), оптимізацію моделей для роботи в реальному часі та інтеграцію з сучасними інформаційними системами для автоматизованої персоналізації сервісів.

Таким чином, поставлену мету дослідження досягнуто: розроблено та реалізовано комплексну методику автоматичного розпізнавання емоційного стану людини за аудіосигналом, яка є статистично обґрунтованою, технічно відтворюваною та має значний практичний потенціал для впровадження у сучасних інформаційних системах.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Schuller B., Batliner A. Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. Chichester : John Wiley & Sons, 2014. 344 p.

2. El Ayadi M., Kamel M. S., Karray F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition*. 2011. Vol. 44, no. 3. P. 572–587.

3. Cowie R., Pelachaud C., Petta P. (eds.). Emotion-Oriented Systems: The Humaine Handbook. Berlin : Springer, 2011. 456 p

4. Tao J., Tan T. Affective computing: A review. *International Journal of Automation and Computing*. 2005. Vol. 2, no. 1. P. 93–108.

5. Cowie R., Douglas-Cowie E., Tsapatsoulis N. et al. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*. 2001. Vol. 18, no. 1. P. 32–80.

6. Fayek H. M., Lech M., Cavedon L. Evaluating deep learning architectures for Speech Emotion Recognition. *Neural Networks*. 2017. Vol. 92. P. 60–68.

7. Lee C. M., Narayanan S. S. Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*. 2005. Vol. 13, no. 2. P. 293–303.

8. Busso C., Bulut M., Lee C. C. et al. IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*. 2008. Vol. 42, no. 4. P. 335–359.

9. Kim J., Provost E. M. Emotion recognition during speech using dynamic Bayesian networks. *Interspeech*. 2013. P. 3299–3303.

10. Latif S., Rana R., Qadir J. et al. Speech Emotion Recognition: Features and Classification Models. *arXiv preprint arXiv:2001.07450*. 2020. URL: arxiv.org.

11. Huang R., Ma C., Chen J. Speech emotion recognition using CNN. *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence, 2014. P. 5015–5019.

12. Satt A., Rozenberg S., Hoory R. Efficient Emotion Recognition from Speech Using Deep Learning on Spectrograms. *Interspeech*. Stockholm, 2017. P. 1089–1093.

13. Zhang Z., Han K., Yu D. Hybrid attention networks for speech emotion recognition. *Interspeech*. Hyderabad, 2018. P. 3314–3318.

14. Cowie R., Douglas-Cowie E. Speech emotion analysis: Towards an objective assessment. *Speech Communication*. 2003. Vol. 40, no. 1-2. P. 33–42.

15. Ковальов, Г. А., Кузьменко, О. О. Методи розпізнавання емоційного стану людини за акустичними параметрами мовлення // Радіоелектроніка, інформатика, управління. 2019. №2. С. 71 78.

16. Білецький, В. С., Ковальова, І. В. Застосування нейронних мереж для розпізнавання емоцій людини за мовленням // Інформаційні технології та комп'ютерна інженерія. 2020. №3. С. 45 51.

17. Дьяків, О. М., Пелещишин, А. М. Системи штучного інтелекту: навчальний посібник. Львів: Видавництво Львівської політехніки, 2021. 280 с.

18. Гнатюк, С. М. Машинне навчання та глибоке навчання: підручник. Київ: КНЕУ, 2022. 304 с.

19. Клименко, О. В., Дяченко, А. О. Аналіз просодичних ознак мовлення для задач розпізнавання емоцій // Математичні машини і системи. 2018. №3. С. 112 118.

20. Гуржій, А. М., Гриценко, В. В. Інтелектуальні інформаційні системи: підручник. Київ: Видавництво НТУУ «КПІ», 2019. – 348 с.

21. Стеценко, О. М., Волошин, В. В. Аналіз спектральних ознак аудіосигналу для задач розпізнавання емоцій // Вісник ХНУРЕ. 2020. №2. С. 89 94.

22. Підгорний, А. М. Глибоке навчання для обробки мовлення: концепції та застосування // Сучасні інформаційні системи. 2022. Т. 6, №1. С. 23 31.

23. Яковенко, І. М., Савченко, І. В. Основи штучного інтелекту: підручник. Харків: ХНУРЕ, 2021. 256 с.

24. Касаткін, Д. О. Методи класифікації емоційних станів у системах людино-машинної взаємодії // Наукові записки НАУ. 2018. №1. С. 42 47.

25. Шевченко, М. В., Коваленко, А. І. Застосування глибоких нейронних мереж для розпізнавання емоцій за аудіосигналом // Радіоелектронні і комп’ютерні системи. 2021. №4. С. 53 60.

ДОДАТКИ

Додаток А

Повний лістинг програмного коду системи кластеризації клієнтів

src/utils.py

```
import os

import logging


def setup_logger(log_path='results/experiment.log'):

    os.makedirs(os.path.dirname(log_path), exist_ok=True)

    logging.basicConfig(filename=log_path, level=logging.INFO, format='%(asctime)s %(levelname)s:%(message)s')

    return logging.getLogger()


def ensure_dir(path):

    os.makedirs(path, exist_ok=True)
```

src/preprocess.py

```
import librosa

import numpy as np

import soundfile as sf


def load_audio(file_path, sr=22050):

    y, _ = librosa.load(file_path, sr=sr)

    return y


def trim_silence(y, top_db=20):

    yt, _ = librosa.effects.trim(y, top_db=top_db)

    return yt
```

```
    def normalize_audio(y):
        return y / np.max(np.abs(y))

    def save_audio(y, path, sr=22050):
        sf.write(path, y, sr)
```

src/feature_extraction.py

```
    import numpy as np
    import librosa
    import os

    def extract_mfcc(file_path, n_mfcc=40, max_len=174, sr=22050):
        y, _ = librosa.load(file_path, sr=sr)
        mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=n_mfcc)
        if mfcc.shape[1] < max_len:
            pad_width = max_len - mfcc.shape[1]
            mfcc = np.pad(mfcc, pad_width=((0,0),(0,pad_width)), mode='constant')
        else:
            mfcc = mfcc[:, :max_len]
        return mfcc

    def batch_extract(file_list, output_dir, n_mfcc=40, max_len=174):
        features = []
        for file_path in file_list:
            mfcc = extract_mfcc(file_path, n_mfcc, max_len)
            save_path = os.path.join(output_dir, os.path.basename(file_path).replace('.wav', '
.npy'))
            np.save(save_path, mfcc)
            features.append(save_path)
```

```
        return features
```

src/train.py

```
    import numpy as np
    import pandas as pd
    from sklearn.model_selection import train_test_split
    from sklearn.preprocessing import LabelEncoder
    from tensorflow.keras.utils import to_categorical
    from tensorflow.keras.models import Sequential
    from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout

    def build_model(input_shape, n_classes):
        model = Sequential([
            Conv2D(32, (3,3), activation='relu', input_shape=input_shape),
            MaxPooling2D((2,2)),
            Dropout(0.3),
            Conv2D(64, (3,3), activation='relu'),
            MaxPooling2D((2,2)),
            Dropout(0.3),
            Flatten(),
            Dense(128, activation='relu'),
            Dropout(0.3),
            Dense(n_classes, activation='softmax')
        ])
        model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy']
)
        return model

    def train_model(features_dir, labels_path, model_path, log, epochs=40, batch_size=32):
```

```
    df = pd.read_csv(labels_path)
    X = []
    y = []
    for _, row in df.iterrows():
        feature_file = os.path.join(features_dir, row['file'].replace('.wav', '.npy'))
        if os.path.exists(feature_file):
            X.append(np.load(feature_file))
            y.append(row['emotion'])
    X = np.array(X)[..., np.newaxis]
    le = LabelEncoder()
    y_enc = le.fit_transform(y)
    y_cat = to_categorical(y_enc)
    X_train, X_val, y_train, y_val = train_test_split(X, y_cat, test_size=0.2, stratify=y
_cat)
    model = build_model(X.shape[1:], y_cat.shape[1])
    log.info(f"Training samples: {X_train.shape[0]}, Validation samples: {X_val.shape[0]}
")
    model.fit(X_train, y_train, epochs=epochs, batch_size=batch_size, validation_data=(X_
val, y_val))
    model.save(model_path)
    log.info(f"Model saved to {model_path}")
    return model, le
```

src/evaluate.py

```
import numpy as np

def evaluate_model(model, X_test, y_test, label_encoder, log):
    loss, acc = model.evaluate(X_test, y_test)
    log.info(f"Test accuracy: {acc:.4f}")
    y_pred = model.predict(X_test)
```

```
    y_true = np.argmax(y_test, axis=1)
    y_pred_class = np.argmax(y_pred, axis=1)
    from sklearn.metrics import classification_report, confusion_matrix
    report = classification_report(y_true, y_pred_class, target_names=label_encoder.class
es_)
    confmat = confusion_matrix(y_true, y_pred_class)
    log.info("Classification report:¥n" + report)
    log.info("Confusion matrix:¥n" + str(confmat))
    return acc, report, confmat
```

src/predict.py

```
import numpy as np

def predict_emotion(model, label_encoder, audio_path, extract_mfcc):
    mfcc = extract_mfcc(audio_path)
    mfcc = mfcc[np.newaxis, ..., np.newaxis]
    pred = model.predict(mfcc)
    class_idx = np.argmax(pred)
    emotion = label_encoder.inverse_transform([class_idx])[0]
    return emotion
```

main.py

```
import argparse
from src.utils import setup_logger, ensure_dir
from src.preprocess import load_audio, trim_silence, normalize_audio, save_audio
from src.feature_extraction import batch_extract
from src.train import train_model
from src.evaluate import evaluate_model
from tensorflow.keras.models import load_model
```

```
        import pandas as pd
        import numpy as np
        import os

        def main():
            parser = argparse.ArgumentParser(description='Emotion Recognition System')
            parser.add_argument('--mode', choices=['preprocess', 'extract', 'train', 'evaluate', '
predict'], required=True)
            parser.add_argument('--data_dir', default='data/')
            parser.add_argument('--labels', default='labels.csv')
            parser.add_argument('--features_dir', default='features/')
            parser.add_argument('--model_path', default='models/emotion_model.h5')
            parser.add_argument('--audio', default=None)
            args = parser.parse_args()
            log = setup_logger()

            if args.mode == 'preprocess':
                # Приклад попередньої обробки
                for fname in os.listdir(args.data_dir):
                    if fname.endswith('.wav'):
                        y = load_audio(os.path.join(args.data_dir, fname))
                        y = trim_silence(y)
                        y = normalize_audio(y)
                        save_audio(y, os.path.join(args.data_dir, fname))
                log.info("Preprocessing complete.")

            elif args.mode == 'extract':
                df = pd.read_csv(args.labels)
                file_list = [os.path.join(args.data_dir, f) for f in df['file']]
```

```
            ensure_dir(args.features_dir)
            batch_extract(file_list, args.features_dir)
            log.info("Feature extraction complete.")

        elif args.mode == 'train':
            train_model(args.features_dir, args.labels, args.model_path, log)

        elif args.mode == 'evaluate':
            df = pd.read_csv(args.labels)
            X, y = [], []
            for _, row in df.iterrows():
                feature_file = os.path.join(args.features_dir, row['file'].replace('.wav','.n
py'))
                if os.path.exists(feature_file):
                    X.append(np.load(feature_file))
                    y.append(row['emotion'])
            X = np.array(X)[...,np.newaxis]
            from sklearn.preprocessing import LabelEncoder
            from tensorflow.keras.utils import to_categorical
            le = LabelEncoder()
            y_enc = le.fit_transform(y)
            y_cat = to_categorical(y_enc)
            model = load_model(args.model_path)
            from src.evaluate import evaluate_model
            evaluate_model(model, X, y_cat, le, log)

        elif args.mode == 'predict':
            from tensorflow.keras.models import load_model
            from src.feature_extraction import extract_mfcc
```

```
        from src.train import LabelEncoder
        model = load_model(args.model_path)
        le = LabelEncoder()
        le.classes_ = np.load('models/label_classes.npy', allow_pickle=True)
        from src.predict import predict_emotion
        emotion = predict_emotion(model, le, args.audio, extract_mfcc)
        print('Predicted emotion:', emotion)

if __name__ == '__main__':
    main()
```

Додаток Б

Довідка про використання результатів дослідження

Результати бакалаврської кваліфікаційної роботи на тему «Аналіз та розробка методів розпізнавання емоційного стану людини за аудіосигналом» можуть бути використані підприємствами та організаціями у різних сферах, де важливо враховувати емоційний стан користувача для підвищення якості сервісу та персоналізації взаємодії. Зокрема, запропонована система може бути корисною для вирішення таких практичних завдань:

– автоматичне визначення емоційного стану користувача під час телефонних дзвінків, онлайн-консультацій або взаємодії з голосовими помічниками;

– персоналізація сервісів дистанційного навчання, медичних або психологічних консультацій з урахуванням емоційної реакції користувача;

– інтеграція у системи підтримки клієнтів для оперативного реагування на негативні емоції та покращення якості обслуговування;

– моніторинг емоційного фону співробітників або клієнтів у корпоративних рішеннях для підвищення ефективності командної роботи та запобігання професійному вигоранню;

– виявлення атипових або ризикових емоційних станів, що можуть свідчити про стрес, тривожність чи інші проблеми, для своєчасного надання допомоги.

Розроблена система реалізована у вигляді відтворюваного програмного комплексу на мові Python і не потребує спеціалізованого апаратного чи програмного забезпечення для базового використання. Час аналізу одного аудіофрагмента не перевищує кількох секунд на стандартному комп’ютері, що забезпечує можливість роботи у реальному часі та легку інтеграцію з існуючими цифровими платформами.