

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ І. І. МЕЧНИКОВА

(повне найменування закладу вищої освіти)

Факультет математики, фізики та інформаційних технологій

(повне найменування факультету)

Кафедра інформаційних технологій

(повна назва кафедри)

Кваліфікаційна робота

на здобуття ступеня вищої освіти «Магістр»

«Веб-система для розпізнавання негативних емоцій із
використанням нейромережових технологій»

(тема кваліфікаційної роботи українською мовою)

«Web-Based System for Recognizing Negative Emotions
Using Neural Network Technologies»

(тема кваліфікаційної роботи англійською мовою)

Виконала: здобувачка заочної форми навчання
спеціальності 122 Комп'ютерні науки

(код, назва спеціальності)

Освітня програма Комп'ютерні науки

(назва)

Муравська Наталія Миколаївна

(прізвище, ім'я, по-батькові здобувача)

Керівник к.ф.-м.н, Ткач Т.Б.

(науковий ступінь, вчене звання, прізвище, ініціали)

(підпис)

Рецензент к.т.н., доцент кафедри інженерії ПЗ національного університету

“Одеська політехніка”, Зіноватна Світлана Леонідівна

(науковий ступінь, вчене звання, прізвище, ініціали)

Рекомендовано до захисту:

Протокол засідання кафедри

Інформаційних технологій

№ від 2024 р.

Завідувачка кафедри

(підпис)

(прізвище, ім'я)

Захищено на засіданні ЕК №

протокол № від 2024 р.

Оцінка / / .

(за національною шкалою/шкалою ECTS/ бали)

Голова ЕК

(підпис)

(прізвище, ім'я)

Одеса 2024

АНОТАЦІЯ

У магістерській дисертації розглядається проблема розпізнавання людських емоцій за допомогою технології нейронних мереж. Актуальність цієї теми визначається зростаючим впливом штучного інтелекту на різні сфери життя, особливо в медицині, освіті та бізнесі.

Мета кваліфікаційної роботи – підвищення точності розпізнавання негативних емоційних станів на основі аудіоданих.

Результати дослідження в роботі продемонстрували високий потенціал нейронних мереж для підвищення точності розпізнавання емоцій.

В результаті виконання кваліфікаційної роботи було розроблено модель нейронної мережі для розпізнавання емоцій, проведено оцінку точності розробленої системи, розроблено прототип програмного забезпечення.

Ключові слова: нейронні мережі, розпізнавання, емоції, негативні емоції, розпізнавання негативу.

ABSTRACT

The master's thesis considers the problem of recognizing human emotions using neural network technology. The relevance of this topic is determined by the growing influence of artificial intelligence on various spheres of life, especially in medicine, education and business.

The purpose of the qualification work is to increase the accuracy of recognizing negative emotional states based on audio and video data.

The results of the study in the work demonstrated the high potential of neural networks for increasing the accuracy of recognizing emotions.

As a result of the qualification work, a neural network model for recognizing emotions was developed, the accuracy of the developed system was assessed, and a software prototype was developed.

Keywords: neural networks, recognition, emotions, negative emotions, recognition of negativity.

ЗМІСТ

ВСТУП	6
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ	8
1.1 Характеристика емоцій та їх класифікація.....	8
1.2 Аналіз існуючих підходів та алгоритмів для розпізнавання емоцій	
12	
1.3 Аналіз існуючих систем для розпізнавання емоцій	18
1.4 Постановка задачі	21
2 ПРОЄКТУВАННЯ СИСТЕМИ	24
2.1 Функціональні вимоги.....	25
2.2 Проєктування діаграми послідовності.....	29
2.3 Проєктування бази даних.....	31
2.4 Опис початкового набору даних	33
2.5 Попередня обробка файлів.....	38
2.6 Опис набору даних.....	45
3 РЕАЛІЗАЦІЯ СЕРВІСУ	47
3 РЕАЛІЗАЦІЯ СЕРВІСУ	47
3.1 Засоби реалізації та оточення	47
3.2 Топологія згорткової нейронної мережі	48
3.3 Реалізація та навчання нейронної мережі	50
3.4 Реалізація системи класифікації емоцій	55
4 ТЕСТУВАННЯ СЕРВІСУ АНАЛІЗУ ТА РОЗПІЗНАВАННЯ ЕМОЦІЙ	
.....	59
4.1 Тестування точності моделі	59

4.2 Функціональне тестування сервісу	62
ВИСНОВКИ.....	66
СПИСОК ЛІТЕРАТУРИ.....	67

ВСТУП

В нинішній комунікації люди вживають різні форми спілкування, подібні як мову, жести і емоції. Дієздатність дізнаватися людські почуття і відчуття, закодовані в міміці обличчя або тональності мови, підсобляє точніше розуміти співрозмовника. Оскільки є безліч різних модальностей, за допомогою яких можуть бути проявлені почуття і емоції, їх розпізнавання представляється одним з найбільш непростих завдань машинного навчання. З розвитком вивчень в області робототехники, штучного розуму і нейронних мереж, зокрема, стає захоплюючою можливістю інтеграції визначення почуттів в машини, щоб зробити спілкування більше різноманітним і природним. Окрім взаємодії людини і комп'ютера, інші області на додаток зможуть отримати вигоду з цієї технології, наприклад, системи спостереження за безпекою водія, програми для розбору міри залученості учасників відеоконференцій, системи онлайн-прокторинга під час проведення іспитів, олімпіад і так далі.

Нині в системах штучного інтелекту активно використовуються штучні нейронні мережі. Нейронна мережа – це обчислювана структура, що полягає з безлічі елементів одного типу. Ці елементи виконують прості функції, а усі процеси, що відбуваються в штучній нейронній мережі, можливо сполучати з процесами, що відбуваються в нервовій системі живих організмів.

Нейронні мережі нелінійні по власній природі, у них відсутня очевидна залежність, що дозволяє сходу вживати розроблену технологію (інформативну модифікацію нейромережі). Впродовж численних років лінійне моделювання представлялося головним методом моделювання, оскільки для нього добре розроблені процедури оптимізації.

У магістерській роботі розглядається можливість використання нейронної мережі при рішенні завдань розпізнавання емоцій людини. Штучні нейронні мережі дають багатообіцяючої перспективи в розвитку, а програмне

забезпечення має величезні переваги від їх використаннями. Окрім цього, кожне завдання, що реалізовується, володіє повний і незвичайний набір методів рішення.

Метою цієї магістерської роботи є розробка сервісу для аналізу та розпізнавання негативних емоційних станів людини на основі нейромережевих технологій. У процесі виконання роботи передбачено розробити методологію збору та аналізу даних, побудувати нейромережеву модель для класифікації емоційних станів та провести її експериментальне тестування.

Об'єктом даного дослідження є розпізнавання негативних емоцій за допомогою голосу із застосуванням нейромережевих технологій.

Предметом дослідження є особливості реалізації системи розпізнавання емоцій із використанням нейромережевих технологій.

Нейронна мережа зобов'язана бути оптимальною по внутрішній структурі і методу управління інформаційними потоками між нейронами.

Головні завдання дисертаційної роботи :

- дослідження наявних варіантів штучних інтелектуальних систем, на додаток способів їх функціонування;
- дослідження основних видів інформативних моделей штучних нейронних мереж. Підбір оптимальної інформаційної моделі нейронної мережі для вирішення завдань розпізнавання емоцій;
- дослідження наявних способів розпізнавання міміки і виділення універсальних можливостей серед їх.

Для реалізації нейронної мережі, яка стане розпізнавати емоції, треба виділити ключові ознаки особи. Секрет здійсненні полягає в навчанні нейронної мережі.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Характеристика емоцій та їх класифікація

Сучасні технології, такі як штучний інтелект та машинне навчання, роблять значний прогрес в аналізі та інтерпретації людської поведінки. Однією з найскладніших і найперспективніших завдань у цій галузі є розпізнавання емоцій. Емоції займають важливе місце у людському спілкуванні, формуючи основу для взаєморозуміння, взаємодії та прийняття рішень.

Розробка систем, здатних інтерпретувати емоції, відкриває безліч нових можливостей, особливо у контексті цифрової взаємодії. Такі системи можуть як реагувати на команди користувачів, а й враховувати їх емоційний стан. Це робить взаємодію більш інтуїтивною і персоналізованою, створюючи унікальні враження користувача.

Емоції відіграють ключову роль в різних сферах життя. Наприклад, у медицині аналіз емоційного стану пацієнта може допомогти у діагностиці депресії, тривожних розладів чи інших психічних захворювань. У сфері освіти системи, що розпізнають емоції, можуть адаптувати навчання під настрій учня, підвищуючи його ефективність (табл. 1.1). У бізнесі аналіз емоцій клієнтів дозволяє краще розуміти їхні потреби, що сприяє створенню більш персоналізованих пропозицій [1].

Сучасні дослідження підтверджують, що точність розпізнавання емоцій може значно покращити взаємодію між людиною та машиною. Системи, здатні аналізувати емоції на основі аудіофайлів, особливо актуальні в умовах великого обсягу даних, таких як дзвінки в кол-центри, подкасти або інтерв'ю. Такі технології знаходять застосування у підтримці користувачів, аналізі настроїв у соціальних медіа, покращенні обслуговування та багатьох інших галузях.

Таблиця 1.1 – Приклади застосування розпізнавання емоцій

Сфера	Приклад використання	Очікуваний результат
Медицина	Діагностика депресії	Поліпшення психічного здоров'я пацієнтів
Освіта	Адаптація уроків під настрій учнів	Підвищення мотивації та успішності
Маркетинг	Аналіз реакції клієнтів на продукти	Оптимізація пропозицій
Бізнес	Оцінка якості обслуговування у кол-центрі	Поліпшення клієнтського досвіду

Існує безліч теорій, що пояснюють природу та функції емоцій. Однією з найвідоміших є теорія Пола Екмана, який виділив сім базових емоцій: радість, смуток, подив, огиду, гнів, страх і зневагу. Ці емоції, як вважає Екман, універсальні всім культур і виявляються однаково, незалежно від етнічної чи соціальної приналежності людини (табл. 1.2).

Таблиця 1.2 – Порівняння моделей емоцій

Модель	Основні емоції	Особливості
Пол Екман	Радість, смуток, гнів, страх тощо.	Універсальні для всіх культур
Роберт Плутчик	8 базових емоцій, змішані стани	Колесо емоцій: інтенсивність та поєднання
Когнітивна теорія	Залежить від інтерпретації подій	Враховує особистий досвід та контекст

Наприклад, радість супроводжується посмішкою, а страх виражається через розширені зіниці та напруження м'язів обличчя.

Інша відома модель – це колесо емоцій Роберта Плутчика (рис.1.1). У цій моделі емоції представлені як спектр, де базові стани можуть поєднуватися, утворюючи складніші емоції. Наприклад, гнів і огида можуть призвести до зневаги, а радість та здивування – до захоплення.

Когнітивна теорія емоцій передбачає, що емоції формуються через інтерпретацію людиною подій, що відбуваються навколо нього. Таким чином, та сама ситуація може викликати різні емоції в різних людей, залежно від їхнього особистого досвіду, переконань і сприйняття.



Рисунок 1.1 – «Колесо емоцій» Р. Плутчика [1]

Різноманіття емоцій можна класифікувати за кількома підходами.

1. Базові емоції.

Концепція базових емоцій була запропонована Полом Екманом, який визначив універсальні емоції, що притаманні всім людям незалежно від культури. До базових емоцій належать:

- радість;
- смуток;
- гнів;

- страх;
- відраза;
- здивування.

Ці емоції вважаються універсальними через їхню біологічну природу, що підтверджується подібністю у їхньому вираженні через міміку, голосові характеристики та фізіологічні реакції.

2. Похідні або складні емоції.

Це емоції, які виникають внаслідок комбінації базових емоцій та контексту:

- захоплення (поєднання радості та здивування);
- провина (поєднання смутку та страху);
- ненависть (поєднання гніву та відрази).

Для дослідження та аналізу емоцій у рамках автоматизованих систем використовують такі підходи:

1. Аналіз фізіологічних сигналів (електрокардіограма, електроенцефалограма, зміни температури тіла).
2. Аналіз міміки (розпізнавання виразу обличчя за допомогою комп'ютерного зору).
3. Аналіз голосу та мовлення (визначення емоцій через тональність, темп, гучність, вибір слів).
4. Опитувальники та суб'єктивна оцінка (самозвіти людей).

Розпізнавання емоцій у голосі. Голос є одним із найінформативніших каналів для виявлення емоцій, оскільки він поєднує у собі як лінгвістичну (зміст сказаного), так і паралінгвістичну (тон голосу, швидкість мовлення, паузи) інформацію. Використання нейронних мереж дозволяє аналізувати ці характеристики та будувати точні моделі для класифікації емоцій [2].

Класифікація емоцій у голосі зазвичай спирається на такі характеристики:

- акустичні особливості: висота тону (pitch), амплітуда (гучність), тембр;
- параметри ритму: швидкість мовлення, довжина пауз;
- семантичний аналіз: аналіз змісту сказаного.

Емоції є складними психофізіологічними реакціями, які залежать від безлічі факторів, включаючи індивідуальні особливості, культурні та соціальні контексти, а також конкретні обставини.

Вони є невід'ємною частиною людської природи, впливаючи на поведінку, сприйняття та взаємодію з навколишнім світом.

1.2 Аналіз існуючих підходів та алгоритмів для розпізнавання емоцій

Технології розпізнавання емоцій являють собою область, що швидко розвивається, яка знаходить застосування в різних сферах життя. Ці технології дозволяють виявляти емоційний стан людини через аналіз її міміки, інтонації голосу, жестів та текстових повідомлень. Емоції можуть бути класифіковані на безліч типів, включаючи радість, смуток, подив, гнів та інші.

Сучасні методи, засновані на глибоких нейронних мережах, таких як згорткові нейронні мережі (CNN), значно збільшили точність розпізнавання. Наприклад, використання CNN дозволяє вилучати ключові ознаки безпосередньо з зображень, що усуває необхідність попередньої обробки даних, що спрощує і прискорює процес аналізу.

Емоції грають найважливішу роль невербальної комунікації. Вони є невід'ємною частиною нашої взаємодії з оточуючими. Люди використовують різні способи вираження своїх емоцій – від слів до жестів та міміки. Успішне розпізнавання емоцій співрозмовника дозволяє більш точно інтерпретувати його наміри, настрої та відносини, що, своєю чергою, сприяє зміцненню соціальних зв'язків та підвищенню рівня довіри.

Розуміння емоцій критично важливе, особливо у контексті взаємодії людини з комп'ютером та розробки людиноподібних роботів. Ці технології роблять спілкування більш природним та різноманітним, що має велике значення як для повсякденного життя, так і для спеціалізованих програм.

Технології розпізнавання емоцій знаходять застосування в різних областях. У маркетингу, наприклад, емоції відіграють ключову роль у сприйнятті реклами та сприйнятті споживчих рішень. Системи, що аналізують емоції, допомагають відстежувати реакції клієнтів на рекламу та пропозиції, а також дозволяють персоналізувати контент.

У сфері медицини розпізнавання емоцій можна використовувати для діагностики різних емоційних розладів. Такі технології дозволяють відстежувати емоційний стан пацієнтів у процесі терапії, що особливо важливо при роботі з дітьми та людьми з обмеженими можливостями. Це може значно покращити якість медичного обслуговування та взаємодію з пацієнтами.

У сфері безпеки технології розпізнавання емоцій застосовуються виявлення підозрілих реакцій, які можуть означати агресію чи потенційну небезпеку. Такі системи корисні як у громадській безпеці, так і в системах безпеки для водіїв, дозволяючи оцінювати рівень уваги та підвищуючи загальну безпеку на дорозі.

Нейронні мережі є складні архітектури, натхненні структурою та функціями біологічного мозку. Вони складаються з багатьох взаємопов'язаних «нейронів», які обробляють інформацію та навчаються на основі великих обсягів даних. Нейронні мережі знаходять широке застосування у таких галузях, як комп'ютерний зір, обробка природної мови та розпізнавання емоцій. Ці системи здатні виявляти складні взаємозв'язки між даними та адаптуватися до різних завдань, що робить їх особливо цінними для аналізу емоційного стану людини.

Повно зв'язкові нейронні мережі (FCNN).

Повно зв'язкові нейронні мережі – це базовий тип нейронних мереж, де кожен нейрон одного шару пов'язаний з усіма нейронами наступного шару. Такі мережі добре підходять для завдань, де дані не мають просторової структури, наприклад, завдання класифікації простих текстів або зображень з невеликими розмірами (рис.1.2).

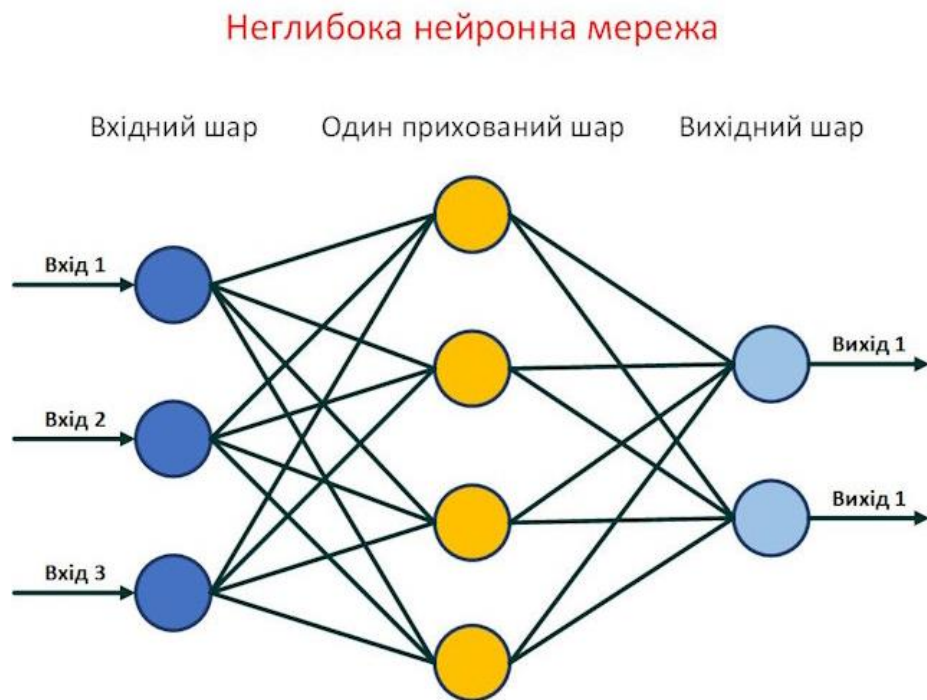


Рисунок 1.2 – Архітектура повнозв'язкові нейронні мережі

Опис та застосування: FCNN можуть бути використані в різних додатках, включаючи розпізнавання образів та аналіз даних. Однак їх ефективність знижується під час роботи з великими обсягами даних або складними візуальними завданнями.

Приклади використання: ці мережі часто використовуються для класифікації простих зображень та аналізу даних невеликих обсягів, таких як фінансові звіти або текстові дані.

Згорткові нейронні мережі (CNN).[3]

Згорткові нейронні мережі є основним інструментом для аналізу зображень та відео. Вони включають спеціальні згортки, які допомагають ефективно витягувати просторові ознаки з даних.

Структура та ключові компоненти : CNN складаються з декількох шарів, включаючи пакунки, pooling -шари та функції активації, такі як ReLU . Ці елементи дозволяють мережі навчатися на великих обсягах даних та виявляти патерни, які можуть бути використані для розпізнавання об'єктів і емоцій.

Приклади використання : CNN широко застосовуються у розпізнаванні об'єктів на зображеннях, аналізі відео та розпізнаванні емоцій по обличчях, демонструючи високу точність та ефективність.

Багатошарова мережа з прямими зв'язками.

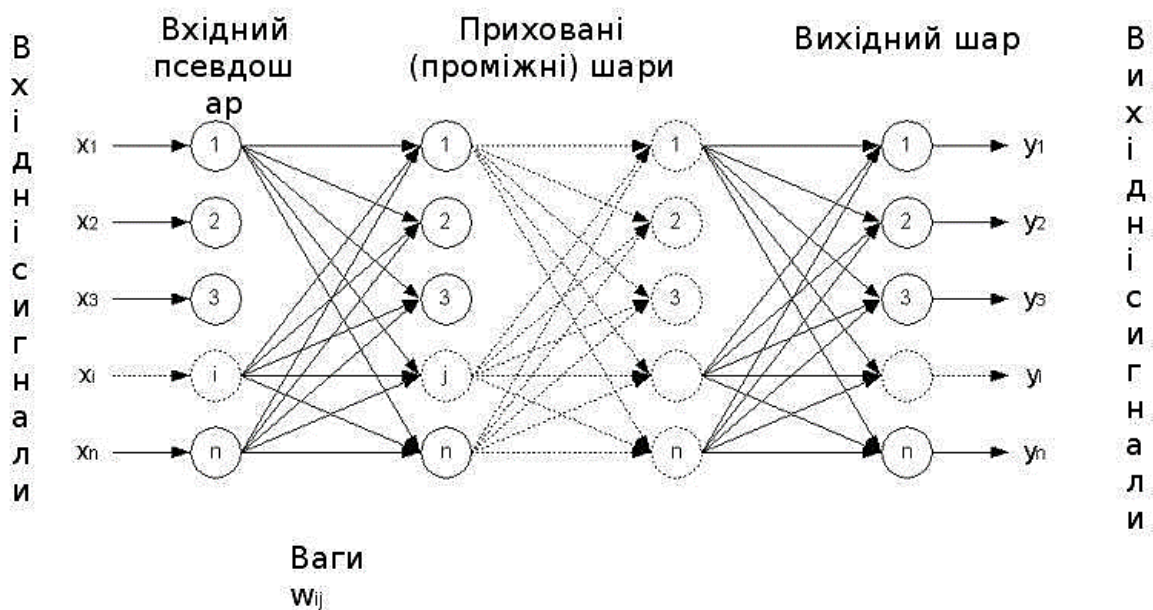


Рисунок 1.3 – Архітектура згорткової нейронної мережі

Рекурентні нейронні мережі (RNN).

Рекурентні нейронні мережі є іншим типом нейронних мереж, які мають унікальну здатність зберігати інформацію про попередні входи. Це робить їх особливо корисними для обробки послідовних даних (рис.1.4).

Особливості роботи з послідовними даними : RNN мають "пам'ять", що дозволяє їм враховувати контекст попередніх даних при аналізі поточного стану. Ця якість робить їх ідеальними для роботи з тимчасовими рядами та послідовностями, такими як текстові дані або аудіозаписи.

Приклади застосування в розпізнаванні емоцій: RNN можуть відстежувати зміни емоцій у відеопотоку, враховуючи контекст та динаміку, що дозволяє більш точно інтерпретувати емоційні стани людини.

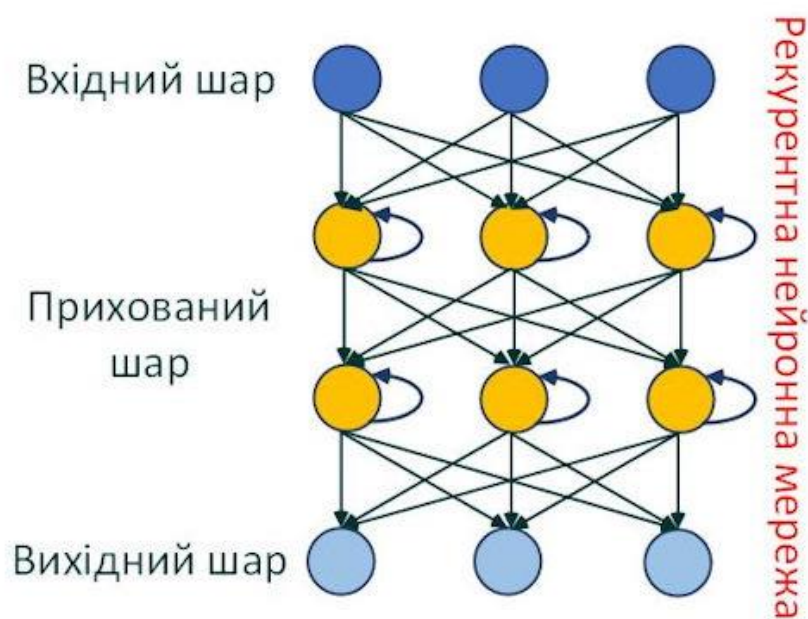


Рисунок 1.4 – Архітектура рекуррентної нейронної мережі

Автоенкодери.

Автоенкодери – це особливий клас нейронних мереж, які навчаються для стиснення даних та подальшого їх відновлення. Вони можуть бути корисними для попередньої обробки даних перед їх аналізом.

Принцип роботи та застосування в попередній обробці даних: автоенкодери стискаються в «код», який зберігає важливі характеристики вихідних даних. Цей код може бути використаний для відновлення даних, що допомагає зменшити шум і виділити ключові ознаки.

Автоенкодера складаються з двох основних частин: енодера, який стискає вхідні дані в компактне представлення (код), та декодера, який намагається відновити оригінальні дані з цього стислого представлення. Завдяки цьому автоенкодера можуть ефективно виділяти приховані патерни та важливі ознаки з великих обсягів даних. Вони широко використовуються в задачах, де потрібно зменшити розмірність або очистити дані від зайвого шуму, як у випадку з аудіо-, зображеннями чи текстами. У контексті попередньої обробки даних, автоенкодера можуть бути застосовані для покращення якості вхідних даних, що дозволяє підвищити ефективність подальших етапів аналізу або моделювання.

Таблиця 1.3 – Порівняння різних типів нейронних мереж

Тип мережі	Застосування	Переваги	Обмеження
FCNN	Класифікація простих даних	Простота, швидка обробка	Неефективні для просторових даних
CNN	Аналіз зображень та відео	Висока точність, стійкість	Високі обчислювальні витрати

Продовження таблиці 1.3

Тип мережі	Застосування	Переваги	Обмеження
RNN	Обробка тимчасових послідовностей	Облік контексту та тимчасових залежностей	Проблеми із довгостроковою пам'яттю
Автоенкодери	Стиснення та фільтрація даних	Зменшення шуму, збереження ознак	Вимагають налаштування для конкретних даних

1.3 Аналіз існуючих систем для розпізнавання емоцій

Існують різні системи, розроблені для розпізнавання емоцій, кожна з яких використовує різні підходи та технології. Розглянемо деякі з них:

DeerFace: ця система розпізнає емоції по лицьовому зображенню і працює на основі мереж згортання. Хоча вона демонструє високу точність при аналізі осіб, вона обмежена набором емоцій, що розпізнаються, і може демонструвати зниження точності в умовах поганого освітлення. DeerFace включає попередньо навчені моделі для детектування осіб та класифікації емоцій, але не враховує аналізу голосу.

EmoVoice: ця система призначена для аналізу аудіоданих та визначення емоцій за голосом. Однак її ефективність може значно знизитися за наявності фонових шумів або низької якості запису. EmoVoice надає інструменти для навчання та використання в реальному часі, але попередньо навчені моделі відсутні, що обмежує кількість емоцій, що розпізнаються.

Multimodal-Emotion-Recognition: ця система пропонує інструменти для класифікації емоцій на основі текстових, аудіо та відео даних. Для детектування осіб використовуються каскади Хаара, а класифікації емоцій застосовуються згорткові нейронні мережі. Ця система дозволяє комбінувати аналіз різних видів даних, що підвищує точність, проте потребує значних обчислювальних ресурсів.[4]

Для детального аналізу існуючих підходів до розпізнавання емоцій було проведено порівняння найпоширеніших рішень, яке представлено у табл.1.4.

Також важливим аспектом систем розпізнавання емоцій є їх здатність адаптуватися до нових даних, що дозволяє підвищити точність у реальних умовах. Системи, які використовують глибокі нейронні мережі, мають потенціал для самонавчання, але потребують значних обчислювальних потужностей. Окрім цього, важливо враховувати етичні питання, пов'язані з використанням таких систем, особливо в контексті конфіденційності та безпеки даних. Подальший розвиток таких технологій спрямований на інтеграцію з іншими інтелектуальними системами, наприклад, віртуальними асистентами та системами охорони здоров'я. Таким чином, сучасні рішення відкривають нові можливості для аналізу емоцій, але водночас вимагають вирішення низки технологічних і соціальних викликів.

Таблиця 1.4 – Порівняння існуючих рішень

Система	Тип даних	Переваги	Обмеження	Точність
DeepFace	Лицьові зображення	Висока точність на обличчях	Залежність від освітлення	Середня

Продовження таблиці 1.4

Система	Тип даних	Переваги	Обмеження	Точність
EmoVoice	Аудіо	Аналіз емоцій за голосом	Чутливість до фонових шумів	Середня
Multimodal-Emotion-Recognition	Відео та аудіо	Вища точність завдяки мультимодальності	Потрібні значні ресурси	Висока
Happy and Routray	Зображення облич	Висока точність для 6 основних емоцій	Обмеження на кількість емоцій, що розпізнаються	94%
Byeon та Kwak	Відео	Аналіз емоцій на основі послідовності кадрів	Залежність від якісних даних	95%
Song та співавтори	Зображення облич	Висока точність при використанні CNN	Ймовірність перенавчання	99,2%
Kotsia та Pitas	Зображення облич	Моделювання обличчя за допомогою сітки	Вимагає точного налаштування	99,7%

Проблеми та недоліки поточних систем.

Незважаючи на досягнення в галузі розпізнавання емоцій, більшість систем стикаються з кількома проблемами. Часто вони залежать від якості даних, що може обмежувати їхню ефективність. Крім того, багато систем мають обмежений набір емоцій, що розпізнаються, і вимагають значних обчислювальних ресурсів для обробки мультимодальних даних. Відсутність універсальних навчальних вибірок також ускладнює розробку еталонних рішень.

Більшість поточних систем демонструють задовільні результати лише за обмежених умов. Мультимодальні системи, які комбінують різні види даних, показують більш високу точність, але потребують покращення обчислювальних можливостей та оптимізації для роботи у реальному часі.

1.4 Постановка задачі

Основною метою дослідження є розробка веб-системи для розпізнавання емоцій з високою точністю. Ця система буде спрямована на покращення взаємодії між людиною та комп'ютером, а також на застосування технологій розпізнавання емоцій у різних сферах, таких як маркетинг, медицина та безпека.

Емоції відіграють ключову роль у нашому житті, і їхнє розпізнавання може значно змінити якість взаємодії з технологіями. Зокрема, успішне розпізнавання емоцій може поліпшити досвід користувача і підвищити ефективність різних додатків.

Для досягнення цієї мети необхідно вирішити такі завдання:

Аналіз існуючих систем : провести детальний аналіз поточних рішень у галузі розпізнавання емоцій, виявити їх сильні та слабкі сторони, а також визначити галузі, які потребують покращення. Це дозволить зрозуміти, які аспекти можна адаптувати чи покращити у запропонованій системі.

Наприклад, деякі системи, такі як AffectNet , використовують величезні набори даних для покращення точності, але мають недоліки в мультимодальному аналізі. Аналіз таких систем дозволить визначити, які технології можуть бути інтегровані у нову веб-систему.

Збір даних та їх попередня обробка: сформувані набір даних, що включає зображення, аудіо та текстові дані, які будуть використовуватися для навчання моделі. Попередня обробка даних включає нормалізацію, видалення шуму та аугментацію, що допоможе покращити якість аналізу. Дослідження показують, що використання аугментації даних може покращити точність розпізнавання на 20-30%, що робить цей етап важливим.

Вибір відповідної архітектури нейронної мережі: дослідити різні архітектури нейронних мереж, такі як CNN, RNN і автоенкодерів , для визначення найбільш підходящої для вирішення поставленого завдання.

Наприклад, CNN добре підходять для аналізу зображень, тоді як RNN можуть бути ефективнішими для обробки послідовних даних, таких як аудіо. Необхідно враховувати специфіку даних та цільові вимоги системи, а також можливості кожної архітектури.

Розробка веб-додатка з впровадженням навченої моделі: створення інтерфейсу для системи, який буде зручним і інтуїтивно зрозумілим. Веб-програма повинна забезпечувати можливість завантаження мультимедійних даних та відображення результатів аналізу в реальному часі. Зручність інтерфейсу може вплинути на сприйняття системи користувачами, тому важливо провести тестування інтерфейсу з реальними користувачами на ранніх етапах розробки.

Тестування системи на реальних даних для оцінки точності: Проведення тестування системи на реальних даних для перевірки її ефективності. Це дозволить виявити можливі недоліки та провести додаткові доопрацювання, якщо це буде потрібно. Статистичні методи, такі як перевірка точності та

повноти, допоможуть оцінити результати тестування та зробити висновки про продуктивність системи.

2 ПРОЄКТУВАННЯ СИСТЕМИ

Процес розробки веб-системи можна розбити на кілька ключових етапів:

Проведення аналізу аналогічних систем : дослідження існуючих рішень допоможе виявити найкращі практики та помилки, яких слід уникнути. Наприклад, системи, які не враховують контекст та динаміку змін емоційного стану, можуть виявитися менш ефективними. Це наголошує на важливості мультимодального підходу (рис.2.1).

Збір даних та їх попередня обробка: необхідно зібрати різноманітні дані для навчання моделі, що забезпечить більш високу точність розпізнавання.

Вибір відповідної архітектури нейронної мережі: це ключовий крок, який визначить ефективність усієї системи. Слід враховувати специфіку завдань та тип даних. Наприклад, використання комбінації CNN для аналізу зображень та RNN для обробки аудіо може значно підвищити точність.

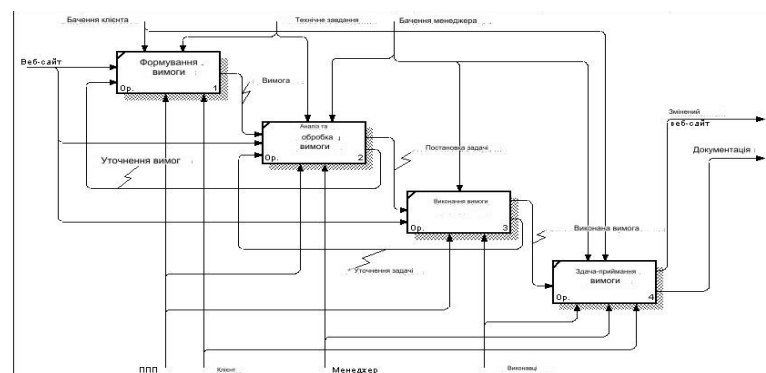


Рисунок 2.1 – Діаграма процесу розробки веб-системи

Розробка веб-застосунку з впровадженням навченої моделі : важливо створити зручний інтерфейс, який дозволить користувачам легко взаємодіяти з системою. Для створення веб-програми можна розглянути використання фреймворків, таких як Flask або Django .

Тестування системи на реальних даних : перевірка системи на реальних даних дозволить оцінити її ефективність та виявити можливі недоліки. Використання тестування користувача допоможе зрозуміти, як різні групи користувачів сприймають систему [5].

Очікується, що розроблена система досягне високої точності розпізнавання емоцій, що дозволить її використовувати у різних галузях.

Зокрема, в маркетингу система зможе аналізувати реакції клієнтів та адаптувати рекламні стратегії, в медицині – відстежувати емоційний стан пацієнтів та допомагати у діагностиці, а в галузі безпеки – підвищувати рівень захисту та запобігати потенційним загрозам.

Таким чином, реалізація даної веб-системи не тільки покращить якість аналізу інформації, а й зробить значний внесок у розвиток технологій розпізнавання емоцій. У майбутньому така система може бути адаптована для використання в освітніх технологіях, психологічної допомоги і навіть у розвагах, що відкриває нові горизонти для застосування.

2.1 Функціональні вимоги

Функціональні вимоги описують, що саме має робити система виявлення негативних емоцій. Представимо їх у вигляді діаграми використання (рис.2.2).

Діаграма варіантів використання є ключовим етапом проектування сервісу, оскільки вона дозволяє зрозуміти взаємодію користувача із системою та визначити основні функціональні можливості. Така діаграма допомагає чітко окреслити ролі, які бере на себе користувач, а також описує всі дії, які може виконати система у відповідь на запити. Сценарії використання, у свою чергу, деталізують процеси, що відбуваються під час виконання кожної з цих дій, описують передумови, етапи роботи та очікувані результати. У цьому розділі представлено діаграму варіантів використання для сервісу аналізу та

розпізнавання емоцій, а також описано основні й альтернативні сценарії для кожного з можливих варіантів.

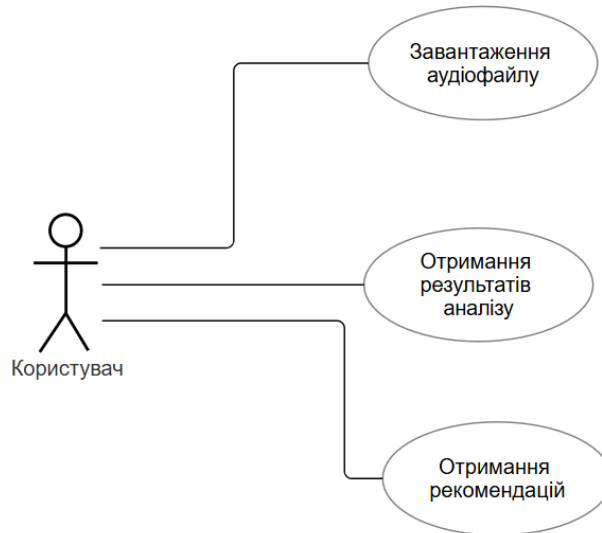


Рисунок 2.2 – Діаграма варіантів використання

Варіант використання «Завантаження аудіо файлу».

Актори: користувач.

Передумови:

- користувач має доступ до сервісу;
- користувач має аудіофайл у підтримуваному форматі (WAV, MP3);
- система працює стабільно.

Основний сценарій:

1. Користувач входить у сервіс і вибирає функцію завантаження файлу.
2. Користувач натискає кнопку «Завантажити аудіо файл» і обирає файл зі свого пристрою.
3. Система перевіряє формат і розмір файлу.
4. Файл завантажується та зберігається тимчасово на сервері.
5. Система відображає повідомлення про успішне завантаження.

Постумови:

– файл збережений у базі даних або тимчасово у файловій системі сервера;

– користувач може перейти до аналізу файлу.

Альтернативний сценарій №1: Непідтримуваний формат.

1. Користувач завантажує файл у форматі, який не підтримується.

2. Система показує повідомлення: «Файл має невідомий формат.»

3. Користувач завантажує інший файл.

Альтернативний сценарій №2: Файл занадто великий.

1. Користувач завантажує файл, розмір якого перевищує ліміт.

2. Система показує повідомлення: «Файл перевищує максимально допустимий розмір.»

3. Користувач обирає інший файл.

Варіант використання «Отримання результатів аналізу».

Актори: користувач.

Передумови:

– користувач успішно завантажив файл;

– система виконала попередню обробку файлу.

Основний сценарій:

1. Користувач натискає кнопку «Розпочати аналіз».

2. Система проводить аналіз емоцій, використовуючи нейромережеву модель.

3. Система відображає результати: визначені емоції, рівень впевненості для кожної емоції.

4. Користувач переглядає результати.

Постумови:

– результати збережені в базі даних;

– користувач ознайомився із результатами.

Альтернативний сценарій №1: Помилка аналізу.

1. Під час аналізу виникає помилка (наприклад, через пошкоджений файл).

2. Система показує повідомлення: «Помилка під час аналізу. Спробуйте ще раз».

3. Користувач завантажує файл повторно.

Варіант використання «Отримання рекомендацій».

Актори: користувач.

Передумови:

- аналіз емоцій був виконаний успішно;
- рекомендаційна система налаштована.

Основний сценарій:

1. Користувач натискає кнопку «Отримати рекомендації».

2. Система обробляє результати аналізу та генерує персональні рекомендації.

3. Рекомендації відображаються у вигляді списку (наприклад, техніки релаксації, медитації, музика тощо).

4. Користувач переглядає рекомендації.

Постумови:

- користувач отримав рекомендації для покращення емоційного стану;
- рекомендації збережені в базі даних.

Альтернативний сценарій №1: Відсутність достатніх даних.

1. Система виявляє недостатню кількість даних для генерації рекомендацій.

2. Система показує повідомлення: «Неможливо згенерувати рекомендації через недостатність даних».

3. Користувач звертається до служби підтримки або завантажує інший файл.

2.2 Проектування діаграми послідовності

Діаграма послідовності є важливим інструментом моделювання, що дозволяє відобразити процес взаємодії між користувачем і системою у вигляді послідовності повідомлень та відповідей. Вона демонструє, як дані проходять між різними компонентами сервісу для виконання певного функціоналу. Нижче наведено опис діаграми послідовності для сервісу аналізу та розпізнавання емоцій (рис.2.3).

Елементи діаграми.

Актор: Користувач.

Об'єкти: інтерфейс користувача, модуль обробки, база даних, модуль аналізу, модуль рекомендацій.

Повідомлення: взаємодія між об'єктами через запити та відповіді.

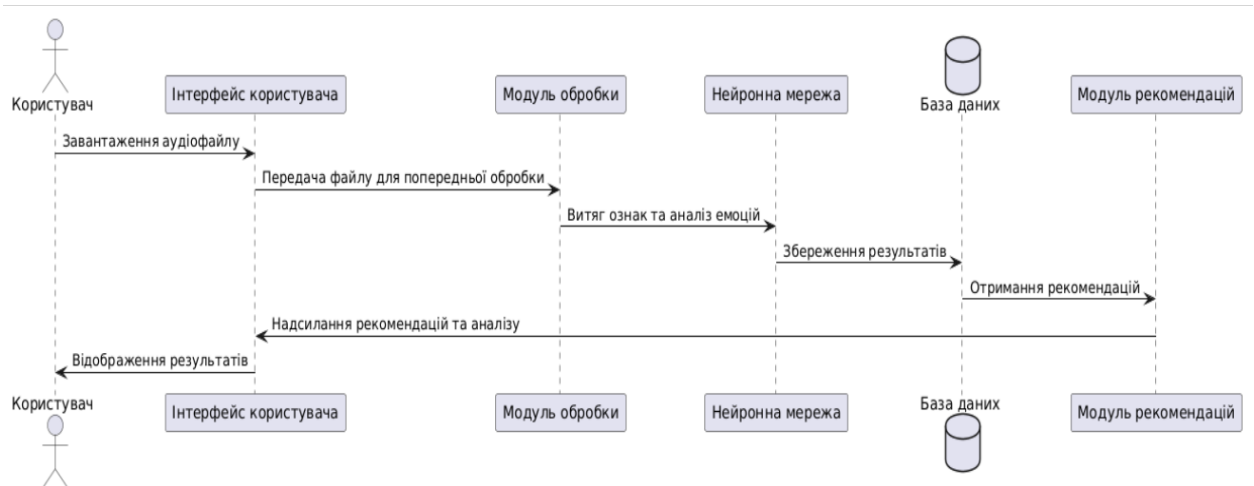


Рисунок 2.3 – Діаграма послідовності

Опишемо діаграму послідовності.

На діаграмі показано етапи обробки запиту користувача від моменту завантаження аудіофайлу до отримання рекомендацій.

1. Користувач завантажує аудіофайл. Цей запит надсилається до сервісу через інтерфейс користувача.
2. Сервіс обробляє аудіо файл.
3. Запит передається до модуля попередньої обробки.
4. Модуль видаляє шум, нормалізує дані, виділяє ключові характеристики голосу (пітч, тон, гучність).
5. Виконується аналіз емоційного стану.
6. Попередньо оброблені дані передаються до нейронної мережі для аналізу.
7. Мережа визначає емоційний стан, використовуючи мовні й акустичні характеристики.
8. Виконується збереження результатів.
9. Результати розпізнавання емоцій зберігаються в базі даних для подальшого використання.
10. Формуються рекомендації.
11. На основі розпізнаних емоційних станів генеруються персоналізовані рекомендації.
12. Цей процес включає доступ до таблиці рекомендацій у базі даних.
13. Відображуються результати користувачу.
14. Сервіс передає результати аналізу й рекомендації на інтерфейс користувача.

2.3 Проектування бази даних

Було спроектовано базу даних для системи розпізнавання емоцій. Основною метою цієї бази є зберігання та обробка даних, необхідних для роботи системи. База даних повинна містити таблиці, які зберігають інформацію про користувачів, а також результати аналізу емоційних станів. Усі дані, що надходять на сервер, зберігаються в окремих таблицях, що дозволяє швидко здійснювати доступ та обробку інформації [6].

В табл. 2.1 зберігається інформація про користувачем системи.

Таблиця 2.1 – Користувачі

Поле	Тип даних	Опис
id	INTEGER PRIMARY KEY AUTOINCREMENT	Унікальний ідентифікатор користувача
name	TEXT	Ім'я користувача
email	TEXT UNIQUE	Унікальний email користувача
created_at	TIMESTAMP DEFAULT CURRENT_TIMESTAMP	Час створення запису

В табл. 2.2 зберігається інформація про файли, що завантажені до систему на аналіз емоцій.

Таблиця 2.2 – Аудіофайли

Поле	Тип даних	Опис
id	INTEGER PRIMARY KEY AUTOINCREMENT	Унікальний ідентифікатор файлу
user_id	INTEGER NOT NULL	ID користувача, який завантажив файл (пов'язане з users.id)
file_name	TEXT	Назва файлу
file_path	TEXT	Шлях до файлу
uploaded_at	TIMESTAMP DEFAULT CURRENT_TIMESTAMP	Час завантаження файлу

В таблиці 2.3 зберігаються пропозиції при виявлених емоціях.

Таблиця 2.3 – Пропозиції

Поле	Тип даних	Опис
id	INTEGER PRIMARY KEY AUTOINCREMENT	Унікальний ідентифікатор пропозиції
emotion	TEXT	Емоція, для якої створена пропозиція
suggestion	TEXT	Текст пропозиції або рекомендації
created_at	TIMESTAMP DEFAULT CURRENT_TIMESTAMP	Час створення пропозиції

Діаграма БД на рис.2.4.

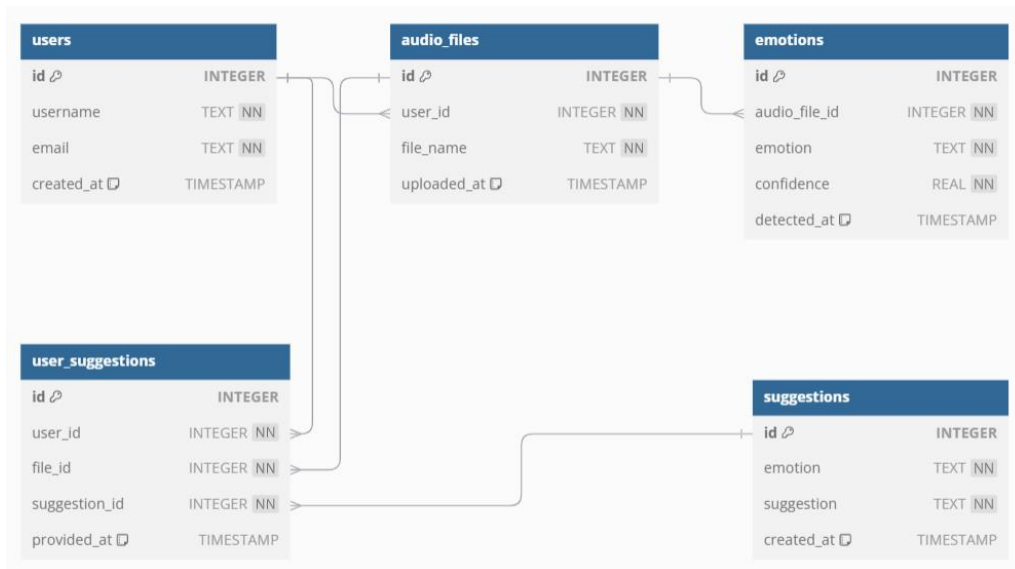


Рисунок 2.4 – Структура бази даних

2.4 Опис початкового набору даних

Під час підготовки матеріалу для навчання та тестування нейронної мережі використовувався власний набір даних. Набір даних є корпусом з емоційними висловлюваннями.

Етапи створення набору даних наведено нижче.

1. Було складено список із 40 фраз, до якого входили фрази з такими емоціями: роздратування (9 фраз), смуток (6 фраз), зневага (6 фраз), страх (3 фрази), огида (3 фрази), радість (4 фрази), здивування та нейтральність (6 фраз), єхидство.

Наведемо (табл. 2.4) декілько прикладів фраз.

Таблиця 2.4 – Таблиця з прикладами фраз

Емоція	Контекст	Текст
Роздратування	Ваш колега вже вкотре, який день підходить і запитує про те, як потрібно працювати з програмою/робити розрахунки/як увімкнути принтер.	Скільки разів мені потрібно тобі повторити?
Єхидство	Ви припускали результат ситуації і говорили про це другові, але він Вас наполегливо не хотів слухати і вірив у свій успіх. У підсумку, ваші припущення виявилися дійсністю.	А я тебе попереджав.
Нейтральність	Ви зацікавлені у близькогородича як пройшов його день.	Як минув твій день?
Огида	Вас видала людина, якій Ви довіряли.	Як можна бути таким негідником?
Сум	Людина, яка Вам приваблива, має неприємний темперамент, але незважаючи на це, Ви хочете продовжити спілкування з цим людиною.	Так, характер у нього, звісно, складний.
Зневага	Ви зайшли до магазину, про який чули багато позитивних та захоплених відгуків. Але за підсумком, асортимент магазину виявився мізерним, співробітники – хамами, а розташування магазину далі, ніж ви припускали.	Магазин своєрідний.
Радість	Ваша мрія купити квартиру/машину справдилась і Ви хочете поділитися цією новиною з другом.	У мене для тебе є новини.
Страх	Ви розбили улюблену машину чоловіка/дружини/хлопця/дівчини. І дуже боїтеся, що за скоєним, підпокарання, як і формі – ви знаєте.	Я розбив твою машину.

Кожній фразі було надано контекст, у якому фраза має прозвучати.

Були знайдені та запрошені диктори, яких просили озвучити сформовані списки фраз, при цьому емоції було видалено зі списку для чистоти експерименту. Дикторам було представлено дві колонки: контекст та текст фрази.

Завдання було поставлено в такий спосіб: про себе прочитати контекст фрази, подумки поринути у нього і озвучити відповідну контексту фразу оскільки озвучив її диктор, опинившись у цій ситуації.

Кожному диктору був представлений один із 5 списків, до якого входило 8 фраз із загального списку, після прочитання списку фраз з додаванням емоційного забарвлення, дикторів просили озвучити той самий список фраз, але вже в нейтральному прочитанні.

Наведемо один список із фраз (табл. 2.5).

Таблиця 2.5 – Список фраз для дикторів

Контекст	Текст
Людина, яка Вам приваблива, має неприємний темперамент, але незважаючи на це, Ви хочете продовжити спілкування з цією людиною.	Так, характер у нього, звісно, складний.
Здійснюється запис для електронний додаток банку. Вимовте фразу не як «Бездушна» машина, а як співробітник банку, що розуміє.	На жаль, ми змушені повідомити, що банк не може надати Вам кредит.
Ви зайшли до магазину, про який чули багато позитивних та захоплених відгуків. Але за підсумком, асортимент магазину виявився мізерним, співробітники – хамами, а розташування магазину далі, ніж ви припускали.	Магазин своєрідний.

Продовження таблиці 2.5

Контекст	Текст
Ви розбили улюблену машину чоловіка/дружини/хлопця/дівчини. І дуже боїтеся, що за скоєним, піде покарання, як і формі – ви знаєте.	Я розбив твою машину.
Ваш колега вже вкотре, який день підходить і запитує про те, як потрібно працювати з програмою/робити розрахунки/як увімкнути принтер.	Скільки разів мені потрібно тобі повторити?
Ви цікавитеся близького родича як пройшов його день.	Як минув твій день?
Вас видав людина, якій Ви довіряли.	Як можна бути таким негідником?
Ви дізналися, що ваш знайомий слухає музику, яка Вам не подобається і ви не розумієте, як подібне можна не просто назвати музикою, але й тим паче слухати.	Що за музику слухаєш?

За підсумками виконаної роботи вийшли такі результати:

1. Було записано 32 диктори чоловічої статі віком від 20 до 60 років.
2. До загального корпусу записаних фраз у результаті увійшли 665 аудіо-фрагменти. Набір даних був підданий перцептивному експерименту.
3. Тип файлів – WAVE.
4. Частота дискретизації – 22050 Гц.
5. Загальний обсяг файлів становив 42 хвилини.

Результати виконаної роботи представлені в табл. 2.6. У зв'язку з тим, що виникло розбалансування даних (нерівномірне співвідношення матеріалів за класами), було прийнято рішення про скорочення кількості

одиниць матеріалу в класі «Нейтральність», це значення є медіаною інших значень емоцій представлених у колонці. Набір даних без розбалансування також наведено в таблиці 2.6.

За підсумками виконаної роботи отримано набір аудіоданих, що забезпечує репрезентативність для задач розпізнавання емоційних станів. Особливу увагу приділено перцептивному експерименту, який дозволив оцінити якість записів і відповідність їх емоційному забарвленню. Для збереження високої якості аналізу аудіофайли були обрані у форматі WAVE з частотою дискретизації 22050 Гц, що забезпечує оптимальний баланс між розміром даних і деталізацією сигналу. Загальний обсяг записаних файлів становив 42 хвилини, що дозволило сформувати достатній корпус для подальшої обробки та аналізу. В результаті вирішення проблеми розбалансування даних вдалося отримати більш однорідний набір, що підвищує точність навчання моделі розпізнавання емоцій.

Таблиця 2.6 – Початкові набори даних

Емоція	Кількість записів	
	Початковий набір даних із розбалансуванням	Початковий набір даних без розбалансування
Єхидство	12	12
Нейтральність	328	36
Огида	27	27
Смуток	55	55
Нехтування	63	63
Радість	46	46
Роздратування	74	74
Страх	32	32
Здивування	28	28
Разом:	665	373
Час:	42 хвилини	19 хвилин 13 секунд

2.5 Попередня обробка файлів

Розглянемо алгоритм перетворення аудіо-файлу на спектрограму.

Весь аудіо-матеріал був перетворений на зображення, а саме в спектрограми, на основі яких проводилося навчання та тестування реалізованої згорткової нейронної мережі [7].

Код для перетворення аудіо-фрагменту на спектрограму:

```
matplotlib.pyplot as
plot from scipy.io im-
port wavfile
path = Path('E:/Dataset').glob(' /*.wav') wavs =
wavs.sort()
number_of_files = len(wavs)
spk_ID = [wavs[i].split('/')[-1].lower() for i in
range(number_of_files)] for i in range(num-
ber_of_files):
    samplingfrequency, signaldata = wavfile.read(wavs[i])
    pxx, freq, bins, im = plot.specgram(x = signaldata, Fs = sam-
plingfrequency, noverlap = 384, NFFT = 512, cmap = 'viridis')
    plot.axis('off')
    plot.savefig("{} .png".format(spk_ID[i]), bbox_inches =
'tight', dpi = 800)
```

Приклад спектрограми, що виходить після обробки, представлений на рис. 2.5. Формат зображення png.

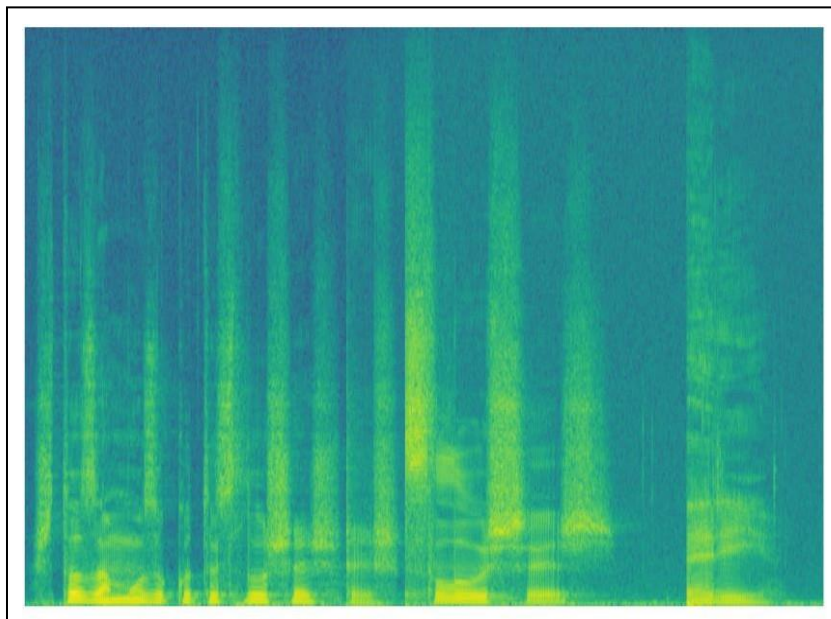


Рисунок 2.5 – Приклад спектрограми

Розглянемо алгоритм перетворення аудіо-файлу в мел-спектрограму.

Усі аудіо-фрагменти були перетворені на мел-спектрограми, на основі яких проводилося навчання та тестування реалізованої згорткової нейронної мережі.

Код для перетворення аудіо-фрагменту в мел-спектрограму:

```
import librosa
import librosa.display
import numpy as np
import matplotlib.pyplot
as plt
n_fft = 2048
hop_length = 512
n_mels = 128
filename = "..." # назва
файлу y, sr =
librosa.load(filename)
signal, _ = librosa.ef-
fects.trim(y)
D =
```

```

np.abs(librosa.stft(signal[:n_fft], n_fft =
n_fft, hop_length =
n_fft + 1)) D =
np.abs(librosa.stft(signal, n_fft = n_fft),
hop_length = hop_length
DB = librosa.amplitude_to_db(D, ref =
np.max)
mel = librosa.filters.mel (sr = sr, n_fft
= n_fft, n_mels =
n_mels)
S = librosa.feature.mel-
spectrogram (signal, sr
= sr, n_fft = n_fft,
hop_length = hop_length,
n_mels = n_mels)
S_DB =
librosa.power_to_db(S,
ref = np.max)
img = librosa.display.specshow(S_DB, sr =
sr, hop_length =
hop_length, x_axis =
'off', y_axis = 'off')
img.figure.savefig("{}{}.png".format(filename, "_mel-
spec"), dpi = 800)

```

Приклад мел-спектрограми, який виходить після обробки аудіо-фрагменту, представлений на рис.2.6

Формат зображення png.

Для обробки файлів використовувалась програма Wave Assistant. Ця програма є спеціалізованим звуковим редактором, у якій реалізовані різні режими обробки сигналу; можливість розмітки сигналу (глибиною до 12 рівнів), включаючи фонетичну розмітку та текстування.

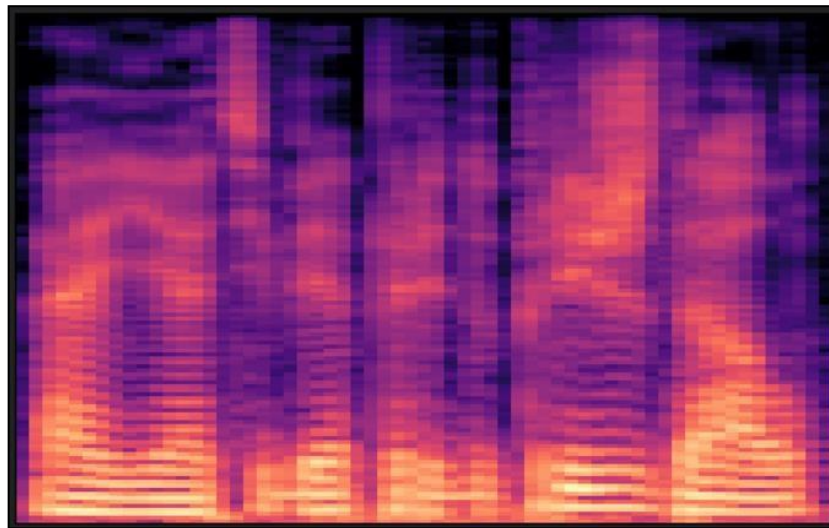


Рисунок 2.6 – Приклад мел-спектрограми

Потім потрібна передобробка аудіофайлу.

Етапи обробки аудіофайлу:

1. Відкриття запису у програмі Wave Assistant.
2. Оскільки, одні фрази диктори вимовляли тихо, інші голосно, їхнього вирівнювання по гучності перед розподілом загального файлу на фрагменти проводилася лінійна обробка файлу, амплітуда приводилася до максимального значення (рис. 2.7).

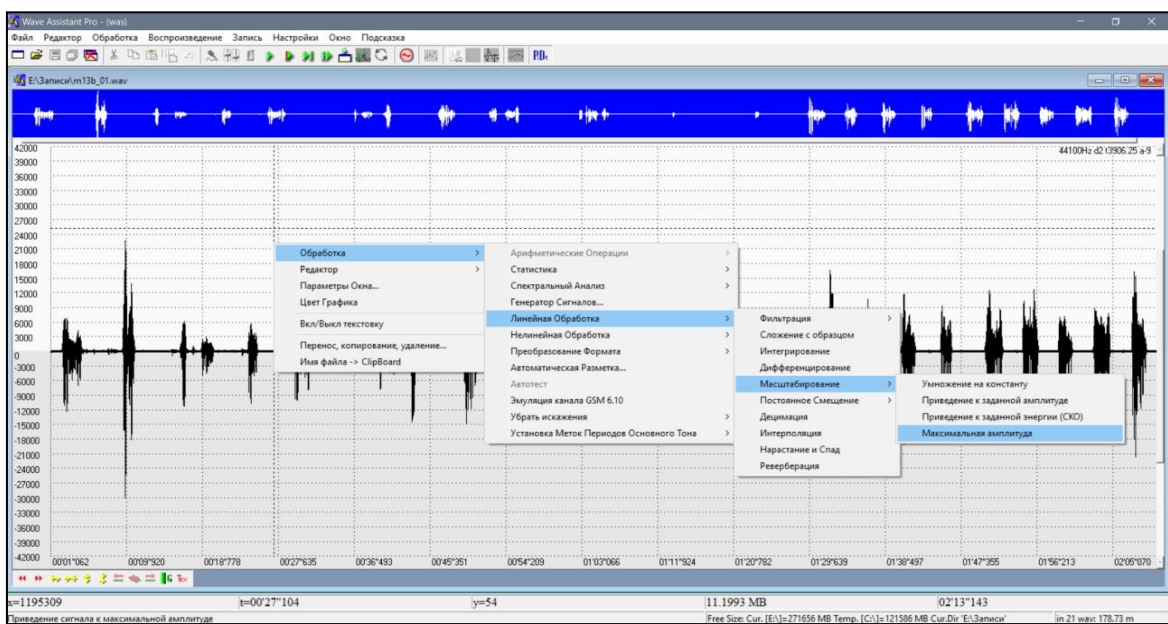


Рисунок 2.7 – Приведення амплітуди до максимального значення

3. Лівою кнопкою миші відзначається початок, правою кнопкою миші – кінець сигналу, який потрібно скопіювати. Потім виконується копіювання фрагмента файлу (фрази), виконується автоматичне встановлення міток основного тону. Ці мітки встановлюються на зеленому рівні. Виробляється візуалізація графіка (рис 2.8).

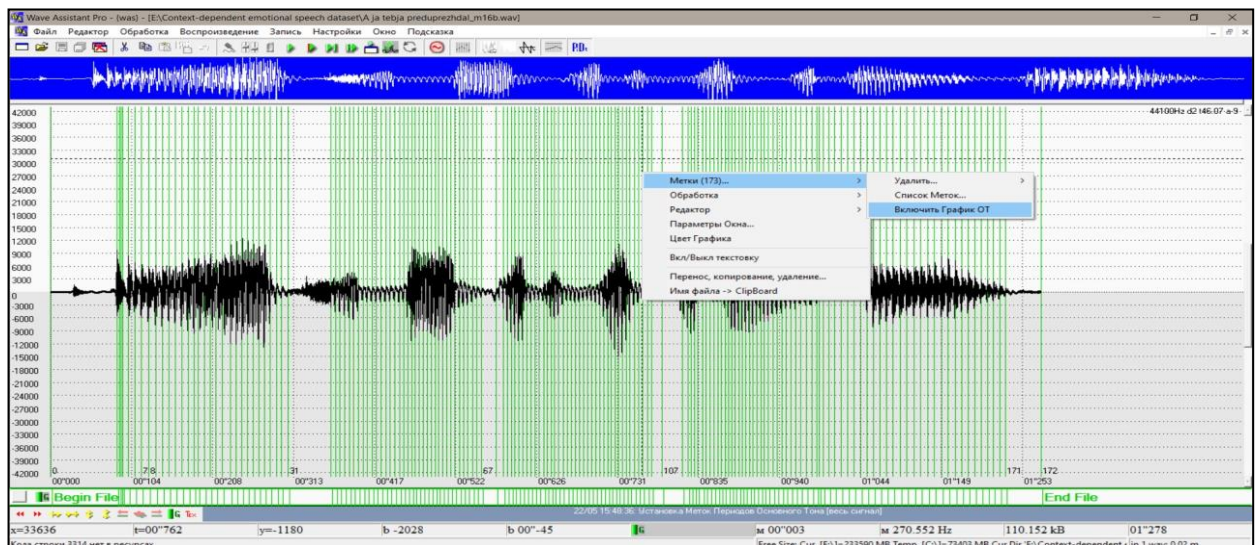


Рисунок 2.8 – Візуалізація графіка основного тону

Через збої автоматичного встановлення міток важливо оцінювати точність виявлення основного тону на кожному етапі обробки. Для цього застосовуються методи ручного коригування, які дозволяють врахувати спотворення, спричинені зміщенням міток. Отримані сегментовані рівні використовуються для подальшого аналізу, що забезпечує високу якість обробки та зменшення кількості помилок у кінцевому результаті. При автоматичному встановленні міток (зелений рівень) виникають збої у вигляді зміщення міток. В результаті чого графік від спотворюється і виникає необхідність редагування міток основного тону вручну.

Потім йде збереження підсумкового результату. В окремі файли формату `seg` зберігаються зелений рівень (SEG_G1) та синій рівень (SEG_B1).

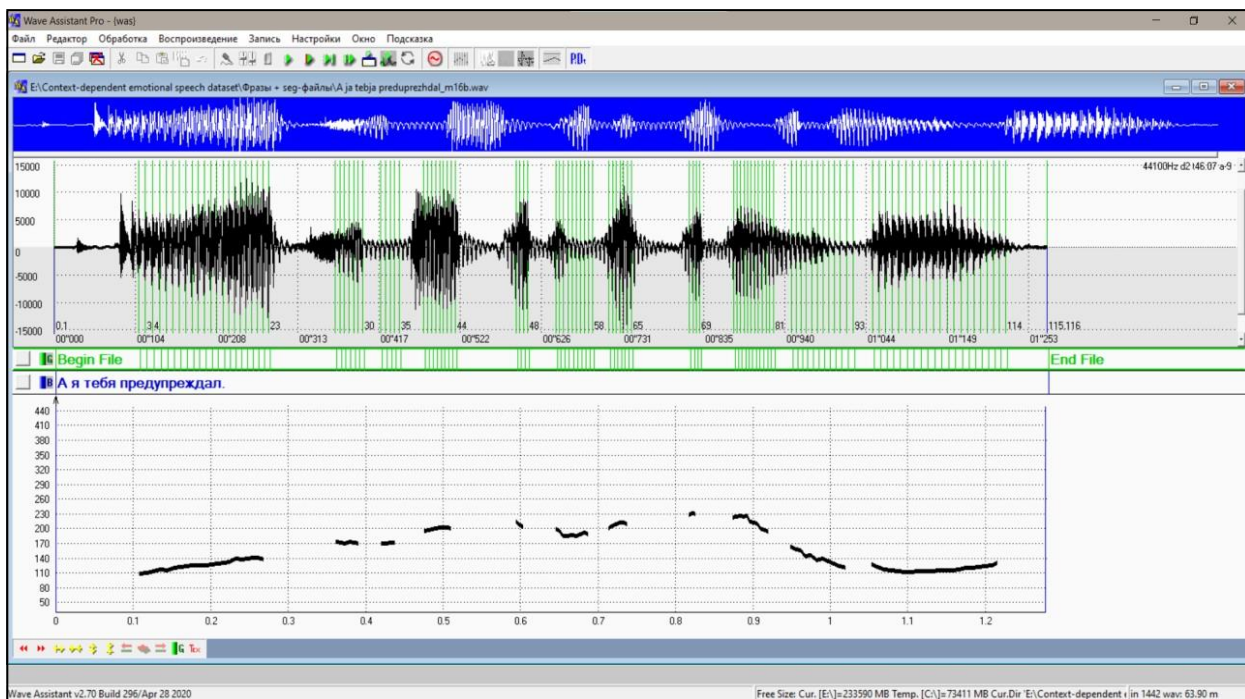


Рисунок 2.9 – Підсумковий результат обробки фрази

Як виглядають seg-файли представлено на рис.2.10 та рис.2.11.

```

*А ja tebja preduprezhdal_m16b – Блок..
Файл Правка Формат Вид Справка
[PARAMETERS]
SAMPLING_FREQ=44100
BYTE_PER_SAMPLE=2
CODE=0
N_CHANNEL=1
N_LABEL=2
[LABELS]
0,2, А я тебе попереджував
112794,2,
< >
Стр 13, стлб 1 100% Windows (CRLF) UTF-8 со специфн

```

Рисунок 2.10 – Seg-файл синій рівень

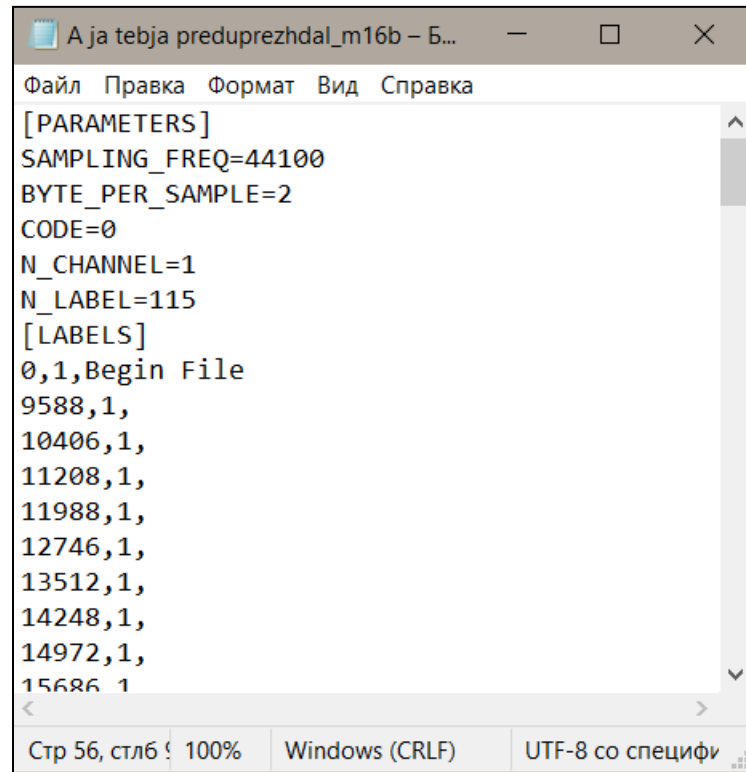


Рисунок 2.11 – Seg-файл зелений рівень

Всі файли зберігаються у спільній папці під назвою «Dataset».

У папці «Диктори» містяться записи всіх дикторів. Папка «Фрази + seg-файли» містить у собі wav- та seg-файли, оброблених аудіо-фрагментів (рис.2.12).

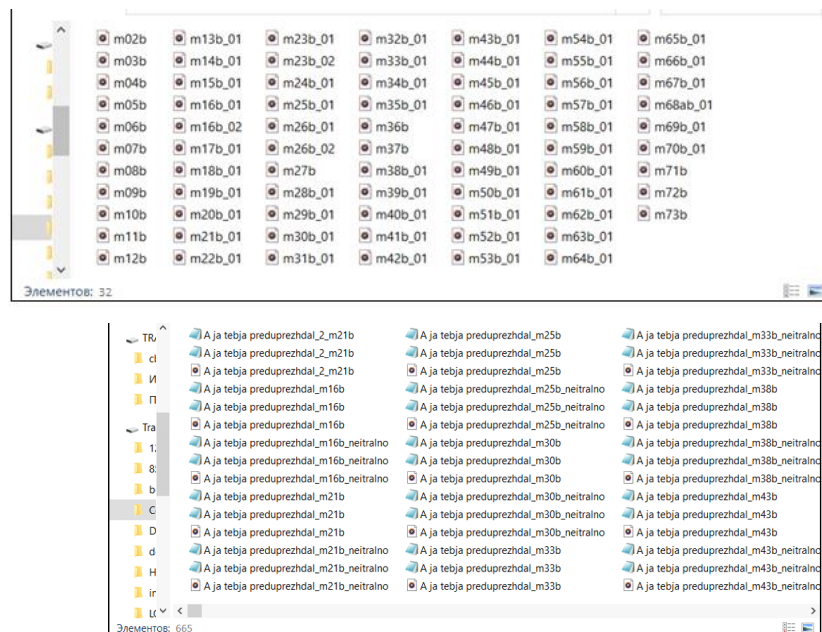


Рисунок 2.12 – Вміст папки «Диктори» та «Фрази + seg-файли»

Таким чином, загальна кількість записів, які пройшли апробацію респондентами, становила 373 аудіо-фрагментів, що становило 19 хвилин 13 секунд. Число висловлювань із негативними емоціями становило 199 записів. Число висловлювань з нейтральною реалізацією становило 174 записів.

2.6 Опис набору даних

В результаті проведеного перцептивного експерименту були отримані відповіді 14 респондентів про 665 файлів, які були записані на початку експерименту. Результати сортування записів на основі перцептивного експерименту наведено у табл. 2.7.

Таблиця 2.7 – Результати експерименту

Емоція	Кількість записів
Єхидство	13
Нейтральність	30
Огида	10
Смуток	73
Нехтування	52
Радість	60
Роздратування	66
Страх	58
Здивування	11
Разом:	373
Час:	19 хвилин 13 секунд

Таким чином, загальна кількість записів, які пройшли апробацію респондентами, становила 373 аудіо-фрагментів, що становило 19 хвилин 13 секунд. Число висловлювань із негативними емоціями становило 199 записів. Число висловлювань з нейтральною реалізацією становило 174 записів.

3 РЕАЛІЗАЦІЯ СЕРВІСУ

3.1 Засоби реалізації та оточення

Для реалізації сервісу розпізнавання негативних емоцій з використанням нейромережових технологій було використано такі засоби розробки, наведені нижче.

Середовище розробки – Jupyter Notebook. Це інтерактивне середовище для написання коду, виконання скриптів та створення документів, що містять код, текстові пояснення та графіки. Jupyter Notebook ідеально підходить для задач дослідження, аналізу даних та демонстрації результатів. Підтримує різні мови програмування (включаючи Python), є можливість інтеграції з бібліотеками машинного навчання та графічними інструментами.[8]

Мовою програмування обрано Python. Вибір Python обґрунтований тим, що це мова, яка активно використовується у сфері машинного навчання, аналізу даних та побудови нейромереж. Вона має широкий спектр бібліотек і фреймворків для цих задач, таких як NumPy, PyTorch, Pandas, Matplotlib та ін.

Бібліотеки та програмні продукти, які ще використовуються при розробці сервісу:

Matplotlib – бібліотека для побудови 2D-графіків. Використовується для візуалізації даних, що дозволяє ефективно представляти результати досліджень і тренування моделей, наприклад, для побудови графіків, метрик точності та візуалізація спектральних ознак голосових даних;

NumPy – базова бібліотека для обчислень з багатовимірними масивами та матрицями. Використовується для обробки акустичних даних (наприклад, обчислення спектральних параметрів, таких як MFCC, або для нормалізації даних перед передачею в модель). Перевагою цієї бібліотеки є висока швидкість обробки даних завдяки використанню низькорівневих алгоритмів.

PyTorch – один із провідних фреймворків для створення нейронних мереж. Його переваги:

- зручність створення складних архітектур моделей, таких як CNN, RNN, LSTM, трансформери тощо;
- автоматичне обчислення градієнтів (автодіференціювання);
- підтримка GPU для швидкого тренування моделей на великих даних. Гнучкість для реалізації кастомних функцій втрат, шарів та оптимізаторів.

PyTorch часто застосовується в задачах розпізнавання емоцій через його потужні інструменти для роботи з послідовностями та рекурентними мережами, що підходить для аналізу мовлення.

Librosa – бібліотека для обробки аудіосигналів, яка надає інструменти для екстракції таких ознак, як спектральні параметри, MFCC, Zero-Crossing Rate та інших важливих характеристик мовлення.

Flask – це мікрофреймворк для Python, який використовується для створення веб-додатків і API. Він забезпечує мінімалістичний підхід до розробки, надаючи розробникам тільки необхідні інструменти, які можна легко розширювати. Flask дозволяє створювати маршрути, які обробляють запити клієнтів, а також забезпечує інтеграцію з іншими бібліотеками.

Вибрані засоби є сучасними, ефективними та широко застосовуються в задачах аналізу даних і машинного навчання.

3.2 Топологія згорткової нейронної мережі

Поява згорткових нейронних мереж пов'язане з вивченням зорової кори головного мозку та застосовується для розпізнавання зображення з 1980-х років.

Згорткові мережі, переважно використовувалися для аналізу зображень, засновані на принципах, якими працює людський зір, тобто зображення представлене у трьох вимірах: ширина, висота, глибина.

Розглянемо архітектуру нейромережі (рис.3.1).

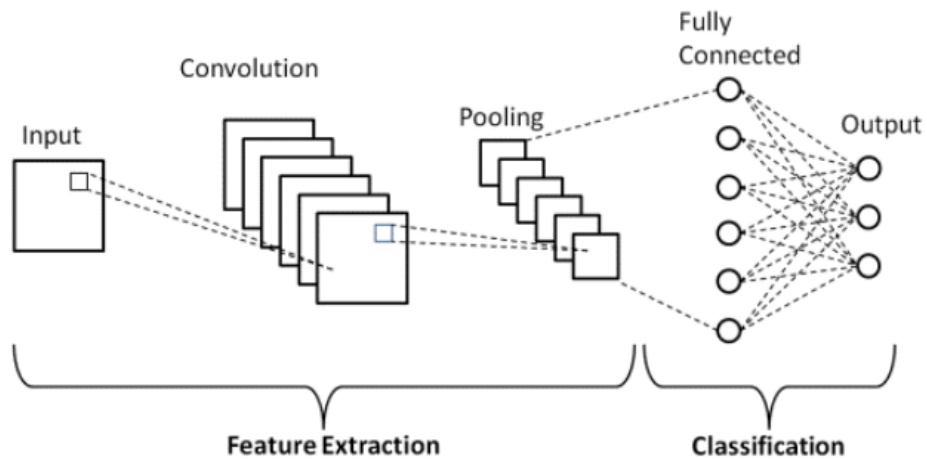


Рисунок 3.1 – Архітектура згорткової нейронної мережі

Вхідний шар (Input). Це початковий шар мережі, який отримує вхідні дані, наприклад, зображення, спектрограму або інші дані у вигляді матриць. Кожна клітинка входу представляє піксель або значення амплітуди. Вхід може мати кілька каналів (наприклад, RGB для зображень).

Виділення ознак (Feature Extraction). Виділення ознак здійснюється за допомогою двох основних операцій: згортки (Convolution) та пулінгу (Pooling).[9]

Згортка (Convolution). На цьому етапі на вхідні дані накладаються фільтри (ядра згортки). Кожен фільтр виконує операцію згортки, що полягає у зваженому підсумовуванні пікселів в області, де фільтр «ковзає» по зображенню.

Результатом є набір карт ознак, що відображають певні патерни.

Завдяки кільком фільтрам мережа може вивчати різні аспекти вхідних даних.

Пулінг (Pooling). Цей етап призначений для зменшення розміру карт ознак, зберігаючи їхню ключову інформацію.

Зменшення розмірності допомагає скоротити обчислювальні витрати та зменшити ризик перенавчання.

Етапи згортки і пулінгу можуть повторюватися кілька разів для витягування дедалі абстрактніших ознак.

Повнозв'язний шар (Fully Connected Layer). Після виділення ознак на попередніх етапах результати згортки та пулінгу вирівнюються (перетворюються в одновимірний вектор).

Цей вектор подається на вхід повнозв'язного шару, який працює як класична нейронна мережа.

Кожен нейрон у повнозв'язному шарі має зв'язки з усіма нейронами попереднього шару, що дозволяє об'єднувати виділені ознаки для прийняття рішення.

Вихідний шар (Output). Цей шар відповідає за класифікацію.

Вихідний шар має стільки нейронів, скільки є класів для класифікації (наприклад, чотири класи емоцій).

Для прийняття рішення використовується активаційна функція.

3.3 Реалізація та навчання нейронної мережі

Після тестування різних топологій була обрана модель згорткової нейронної мережі, що складається з чотирьох згорткових шарів та двох пулінгових шарів. Схема нейронної мережі представлена на рис. 3.2.

Ця архітектура згорткової нейронної мережі (CNN) складається з кількох шарів, які використовуються для обробки даних, виділення ознак та класифікації. Розглянемо алгоритм роботи мережі та пояснимо, як обробляються вхідні дані.

На вхід подається попередньо оброблений сигнал у вигляді спектрограми або мел-фільтрових коефіцієнтів. Перед подачею в модель

спектрограми або інші дані потрібно перетворити в потрібний формат: розмір вхідних даних: [batch_size, 1, height, width].

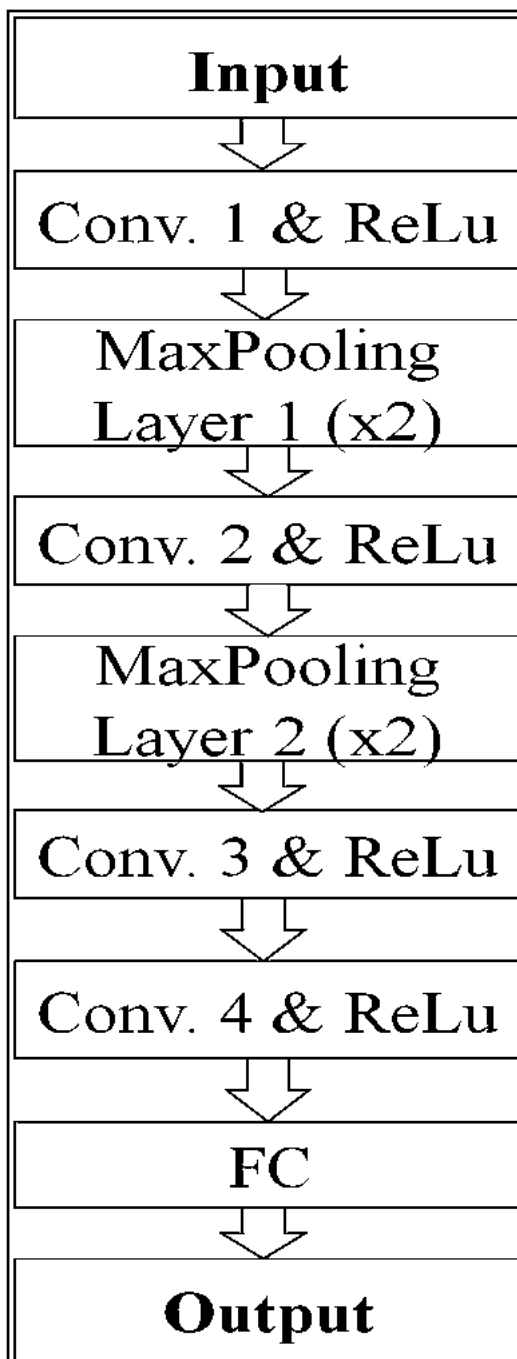


Рисунок 3.2 – Схема реалізованої згорткової нейронної мережі

Алгоритм роботи архітектури:

1. Conv. 1 & ReLU:

– шар згортки (Convolution): перший згортковий шар накладає ядра згортки на вхідний тензор, що дозволяє виявляти локальні патерни, такі як зміни інтенсивності, краї чи текстурні ознаки;

– активаційна функція ReLU (Rectified Linear Unit): використовується для додавання нелінійності. Вона зберігає позитивні значення та обнуляє від'ємні, дозволяючи моделі вивчати складні залежності.

2. MaxPooling Layer 1 (x2):

– максимальний пулінг (MaxPooling): зменшує розмір карти ознак, вибираючи максимальне значення в заданому вікні (наприклад, 2×2). Це зменшує обчислювальну складність і робить модель стійкішою до змін у даних, таких як зсув або масштаб.

3. Conv. 2 & ReLU:

– другий шар згортки із застосуванням ReLU обробляє результати після першого пулінгу. Він виділяє більш складні ознаки, наприклад, текстурні або патерни, специфічні для даного класу.

4. MaxPooling Layer 2 (x2):

– ще один шар максимального пулінгу. Цей етап ще більше зменшує розмір карт ознак, виділяючи найважливішу інформацію.

5. Conv. 3 & ReLU:

– третій згортковий шар працює із зменшеними даними та концентрується на високорівневих абстрактних ознаках, що є характерними для класифікації (наприклад, емоційних станів).

6. Conv. 4 & ReLU:

– четвертий згортковий шар завершує процес виділення ознак, готуючи карти ознак для фінального класифікатора.

7. Fully Connected Layer (FC):

– результати згорток і пулінгу вирівнюються в одномірний вектор;

– цей вектор подається на повнозв'язний шар, який комбінує отримані ознаки, створюючи представлення для класифікації.

8. Output Layer:

- вихідний шар генерує прогноз на основі оброблених ознак;
- кількість нейронів на виході відповідає кількості класів (наприклад, різні емоційні стани, такі як «злість», «сум», «радість» тощо);
- використовується функція активації для перетворення виходів у ймовірності для кожного класу.[10]

Ця архітектура дозволяє ефективно обробляти дані, такі як голосові спектрограми, виділяючи емоційно релевантні ознаки, що є важливими для розпізнавання емоційних станів. Завдяки багаторівневій обробці вона може розпізнавати як низькорівневі (зміни інтонації), так і високорівневі ознаки (емоційне забарвлення мовлення).

Основний принцип роботи:

1. Виділення ознак: шари згортки й пулінгу виділяють дедалі складніші характеристики із вхідного сигналу.
2. Розпізнавання залежностей: повнозв'язний шар навчається на цих ознаках для визначення взаємозв'язків між ними.
3. Класифікація: вихідний шар забезпечує класифікацію на основі вивчених ознак.

Для навчання моделі потрібно визначити функцію втрат (`nn.CrossEntropyLoss`) та оптимізатор (`torch.optim.Adam`):

```

criterion = nn.CrossEntropyLoss()
optimizer = torch.optim.Adam(model.parameters(), lr=0.001)

# Цикл навчання
for epoch in range(num_epochs):
    for inputs, labels in train_loader: # train_loader -
DataLoader із даними
        outputs = model(inputs)
        loss = criterion(outputs, labels)

```

```
optimizer.zero_grad() # Скидання градієнтів
loss.backward() # Обчислення градієнтів
optimizer.step() # Оновлення ваг
```

Реалізація побудови моделі мовою Python з використанням бібліотеки PyTorch мо нижче:

```
import torch
import torch.nn as nn
import torch.nn.functional as F
# Визначення архітектури моделі
class EmotionRecognitionCNN(nn.Module):
    def __init__(self, num_classes):
        super(EmotionRecognitionCNN, self).__init__()
        # Перший згортковий шар + активація ReLU + MaxPooling
        self.conv1 = nn.Conv2d(in_channels=1, out_channels=32, kernel_size=3, stride=1, padding=1) # Згортка
        self.relu1 = nn.ReLU() # Активація
        self.pool1 = nn.MaxPool2d(kernel_size=2, stride=2) # MaxPooling
        # Другий згортковий шар + активація ReLU + MaxPooling
        self.conv2 = nn.Conv2d(in_channels=32, out_channels=64, kernel_size=3, stride=1, padding=1)
        self.relu2 = nn.ReLU()
        self.pool2 = nn.MaxPool2d(kernel_size=2, stride=2)
        # Третій згортковий шар + активація ReLU
        self.conv3 = nn.Conv2d(in_channels=64, out_channels=128, kernel_size=3, stride=1, padding=1)
        self.relu3 = nn.ReLU()
        # Четвертий згортковий шар + активація ReLU
        self.conv4 = nn.Conv2d(in_channels=128, out_channels=256, kernel_size=3, stride=1, padding=1)
        self.relu4 = nn.ReLU()
        # Повнозв'язний шар
        self.fc1 = nn.Linear(256 * 8 * 8, 512) # Передбачено розмір вхідного тензора (8x8 - результат пулінгу)
        self.fc2 = nn.Linear(512, num_classes) # Вихід - кількість класів емоцій
    def forward(self, x):
        # Перший блок
        x = self.conv1(x)
        x = self.relu1(x)
        x = self.pool1(x)
        # Другий блок
        x = self.conv2(x)
        x = self.relu2(x)
        x = self.pool2(x)
        # Третій блок
        x = self.conv3(x)
        x = self.relu3(x)
```

```

# Четвертий блок
x = self.conv4(x)
x = self.relu4(x)
# Вирівнювання перед передачею в повнозв'язний шар
x = x.view(x.size(0), -1)
# Повнозв'язні шари
x = self.fc1(x)
x = F.relu(x) # ReLU активація
x = self.fc2(x)
return x

```

Параметри реалізованої згорткової нейронної мережі представлені нижче.

1. Оптимізатор – алгоритм який змінює ваги усунення під час навчання. Використовуємо Adam з параметрами $lr = 0.001$, $weight_decay = 0.0001$.

2. `loss_function`: функція втрат, тому що перед нами стоїть завдання класифікації, як параметр вказуємо `CrossEntropyLoss`, яка використовується для вирішення задач такого типу.

3. `epochs`: кількість проходів за всіма тренувальними даними. Шляхом тестування було обрано оптимальну кількість епох – 12 епох.

Співвідношення навчальної та тестової вибірки було обрано 80% та 20%, відповідно. Дані в навчальній та тестовій вибірках не перетинаються.

Для навчання та тестування нейронної мережі використовувалися раніше оброблені дані.

3.4 Реалізація системи класифікації емоцій

В результаті навчання та тестування реалізована згорткова нейронна мережа показала наступні результати.

Для тренування моделі були використані спектральні ознаки, отримані з аудіоданих. Спектрограми дозволяють візуалізувати зміну частотного спектра сигналу з часом, що є важливим для розпізнавання емоційного забарвлення голосу.

Навчання проводилося на спеціально підготовленому наборі даних, який містив різні емоційні стани. Показники, що оцінюють ефективність моделі представлені в таблиці 3.1. У ній наведено результати як для тренувальної, так і тестової вибірок, що дозволяє оцінити узагальнюючу здатність нейронної мережі.

Таблиця 3.1 – Результати навчання та тестування моделі. Спектрограми

Набір даних Результати	Початковий набір даних (розбаланс.)	Початковий набір даних (без розбаланс.)	Початковий набір даних (бін. клас.)
Training accuracy (12 епох)	0.830582	0.993377	0.936538
Test accuracy (12 епох)	0.563573	0.757961	0.746987

Мел-спектрограми є потужним інструментом для аналізу акустичних характеристик звуку. Вони використовуються для перетворення аудіо-сигналів в спектральні представлення, що дозволяють ефективно виокремлювати особливості мовлення або звукових сигналів, важливі для розпізнавання емоцій. У процесі обробки даних мел-спектрограми забезпечують високу точність у виявленні ключових ознак, таких як інтонація, тембр та частотні характеристики звуку, що є важливими для класифікації емоцій (табл. 3.2).

Таблиця 3.2. Результати навчання та тестування моделі. Мел-спектрограми

Набір даних \ Результати	Початковий набір даних (Розбаланс.)	Початковий набір даних (без розбаланс.)	Початковий набір даних (Бін. клас.)
Training accuracy (12 епох)	0.922943	0.954173	0.9875
Test accuracy (12 епох)	0.658536	0.543352	0.773076

Графік тестування та навчання моделі представлений на рис.3.3.

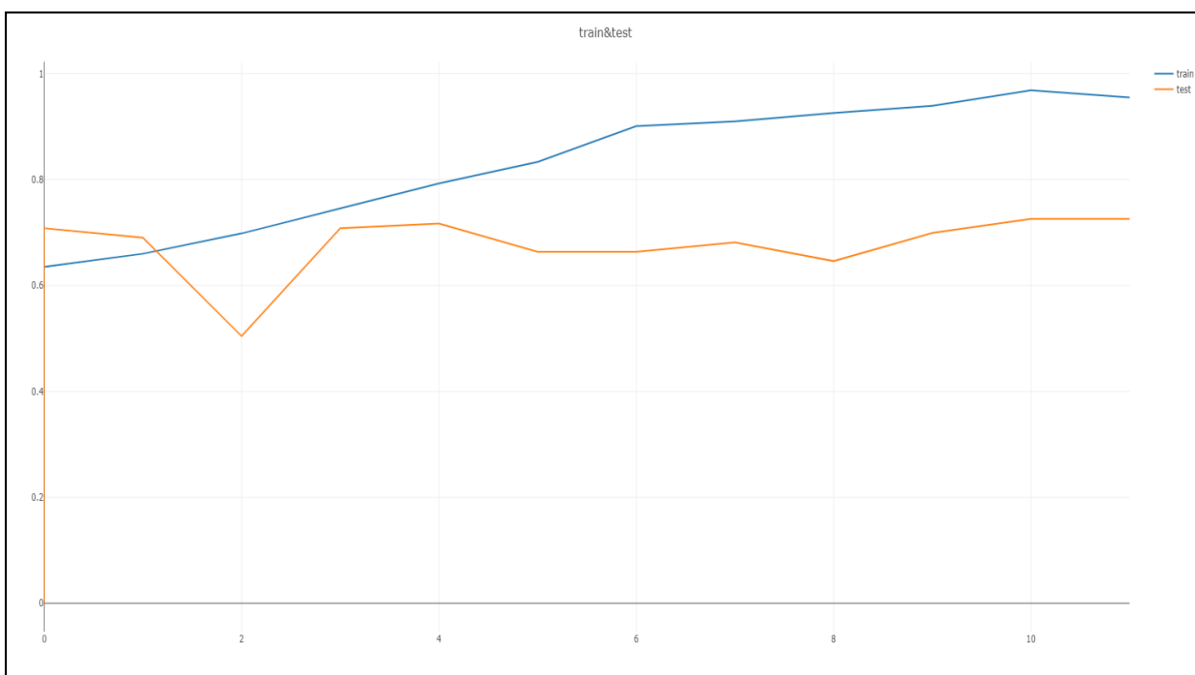


Рисунок 3.3 – Тестування та навчання моделі

Зробимо візуалізацію тестування моделі (рис 3.4).

```
Epoch: 0 Train Loss: tensor(14.6484) Train Accuracy: 0.6351351351351351 Test Accuracy: 0.7079646017699115
Epoch: 1 Train Loss: tensor(6.1682) Train Accuracy: 0.6599099099099099 Test Accuracy: 0.6902654867256637
Epoch: 2 Train Loss: tensor(2.3201) Train Accuracy: 0.6981981981981982 Test Accuracy: 0.504424778761062
Epoch: 3 Train Loss: tensor(1.0035) Train Accuracy: 0.7454954954954955 Test Accuracy: 0.7079646017699115
Epoch: 4 Train Loss: tensor(0.4908) Train Accuracy: 0.7927927927927928 Test Accuracy: 0.7168141592920354
Epoch: 5 Train Loss: tensor(0.3956) Train Accuracy: 0.8333333333333334 Test Accuracy: 0.6637168141592921
Epoch: 6 Train Loss: tensor(0.2963) Train Accuracy: 0.9009009009009009 Test Accuracy: 0.6637168141592921
Epoch: 7 Train Loss: tensor(0.2391) Train Accuracy: 0.9099099099099099 Test Accuracy: 0.6814159292035398
Epoch: 8 Train Loss: tensor(0.1711) Train Accuracy: 0.9256756756756757 Test Accuracy: 0.6460176991150443
Epoch: 9 Train Loss: tensor(0.1746) Train Accuracy: 0.9391891891891891 Test Accuracy: 0.6991150442477876
Epoch: 10 Train Loss: tensor(0.1067) Train Accuracy: 0.9684684684684685 Test Accuracy: 0.7256637168141593
Epoch: 11 Train Loss: tensor(0.1130) Train Accuracy: 0.954954954954955 Test Accuracy: 0.7256637168141593
```

Рисунок 3.4 – Результати тестування моделі. Графіки основного тону

Реалізована нейронна мережа показала середні показники на наступних наборах даних.

1. Спектрограми: початковий набір без розбалансування даних – 0.757961; початковий набір даних бінарної класифікації – 0.746987.

2. Мел-спектрограми: початковий набір даних для бінарної класифікації – 0.773076.

Оскільки експерименти були проведені на наборах даних, що пройшли перевірку перцептивним експериментом, отже, результати більшою мірою узгоджуються з тим, як сприймає та інтерпретує емоції людина.

Подібні результати можуть свідчити про те, що проведення перцептивних експериментів та апробація матеріалу за участю респондентів може підвищити успішність та якість класифікації емоцій з використанням нейромережових технологій.

4 ТЕСТУВАННЯ СЕРВІСУ АНАЛІЗУ ТА РОЗПІЗНАВАННЯ ЕМОЦІЙ

4.1 Тестування точності моделі

Для оцінки точності розробленої системи розпізнавання емоцій за допомогою нейронних мереж було проведено кілька експериментів.

Проведемо оцінку точності класифікації емоцій тестуванням системи розпізнавання емоцій на основі аудіоданих.

Оцінимо точність і здатність системи коректно класифікувати емоційні стани на основі аудіоданих, використовуючи нейронну мережу та попередньо оброблені аудіоознаки.

Для експерименту використовувався набір даних RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song). Цей набір включає 1440 аудіофайлів із записами емоційного мовлення (щасливі, сумні, гнівні, нейтральні та інші емоції), виконаними 24 професійними акторами (12 чоловіків та 12 жінок). Усі дані доступні у форматі 16-bit WAV з частотою 48 kHz [11].

Було проведено:

Попередня обробка даних:

- аудіодані були розбиті на короткі фрагменти;
- використовувались ознаки, отримані за допомогою мел-частотних кепстральних коефіцієнтів (MFCC);
- додатково дані проходили нормалізацію та фільтрацію шумів.

Навчання системи:

- використовувався метод розподілу на тренувальні (80%), валідаційні (10%) та тестові (10%) вибірки.

Оцінка точності:

- для оцінки використовувалися стандартні метрики точності.

Результати:

- точність класифікації досягла 85% для основних категорій емоцій;
- найвищу точність система показала для «гніву» та «радість» (до 90%), тоді як «нейтральні» емоції класифікувались із точністю 80%.

- основна плутанина спостерігалася між «страхом» та «смутом», що вказує на необхідність додаткової оптимізації.

Експеримент підтвердив здатність нейронної мережі ефективно розпізнавати емоції на основі аудіоданих. Подальша оптимізація моделі та використання розширених наборів даних (зокрема, багатомовних) може підвищити точність результатів.

Метою наступного експерименту є оцінка точності та F1-міри для різних розмірів навчальної вибірки. Вибірki будуть складатися з 50%, 70% і 100% від повного набору даних, і ми перевіримо, як це впливає на якість навчання системи. Це допоможе з'ясувати, чи може система добре працювати на обмежених даних.

Кроки експерименту:

- підготовка даних: оберемо підмножини даних для кожного випадку (50%, 70%, 100%).

- навчання моделі: використовуємо різні частини набору даних для тренування моделі.

- оцінка моделі: використовуємо метрики точності та F1-міри для оцінки результатів на тестовій вибірці.

Метрики:

- точність (Accuracy), яка визначається як відношення правильно класифікованих елементів до загальної кількості елементів.

- F1-міра (F1-Score) – середнє гармонійне між точністю і повнотою.

Таблиця 4.1 – Результати експерименту

Розмір навчальної вибірки	Точність (%)	F1-міра (%)
50%	76.4	0.74
70%	82.1	0.79
100%	88.3	0.85

Як видно з таблиці, збільшення розміру навчальної вибірки позитивно впливає на точність та F1-міру моделі. При використанні 100% даних модель показує найкращі результати з високою точністю та F1-мірою.

Використання 70% даних також дає хороші результати, хоча точність та F1-міра знижуються порівняно з повним набором.

Зменшення розміру вибірки до 50% призводить до деякого зниження ефективності моделі, що вказує на важливість більшого набору даних для навчання.[12]

Ці результати можуть бути корисними для визначення оптимального розміру навчальної вибірки в реальних умовах, коли доступні обмежені дані.

Для глибшого розуміння ефективності моделі було проведено аналіз помилок. Деякі класи емоцій, такі як радість та гнів, демонстрували високу точність (понад 90%). Однак такі емоції, як єхидство та зневага, мали точність нижче 75%, що пов'язано з їхньою складністю та суб'єктивністю сприйняття (табл. 4.2).

Таблиця 4.2 – Результати тестування моделі

Емоція	Precision	Recall	F1-score
Радість	0.95	0.93	0.94
Страх	0.92	0.91	0.91
Єхидство	0.65	0.60	0.62
Середнє значення	0.84	0.81	0.82

Емоції з яскраво вираженими вокальними характеристиками, наприклад страх (високий тон) або гнів (гучність і різкість), розпізнавались моделлю краще.

Більш тонкі емоції, такі як єхидство, потребують додаткових даних для покращення класифікації.

4.2 Функціональне тестування сервісу

Перевіримо функціональність розробленого сервісу.

При першому запусці відкривається сторінка з авторизацією (рис.4.1).

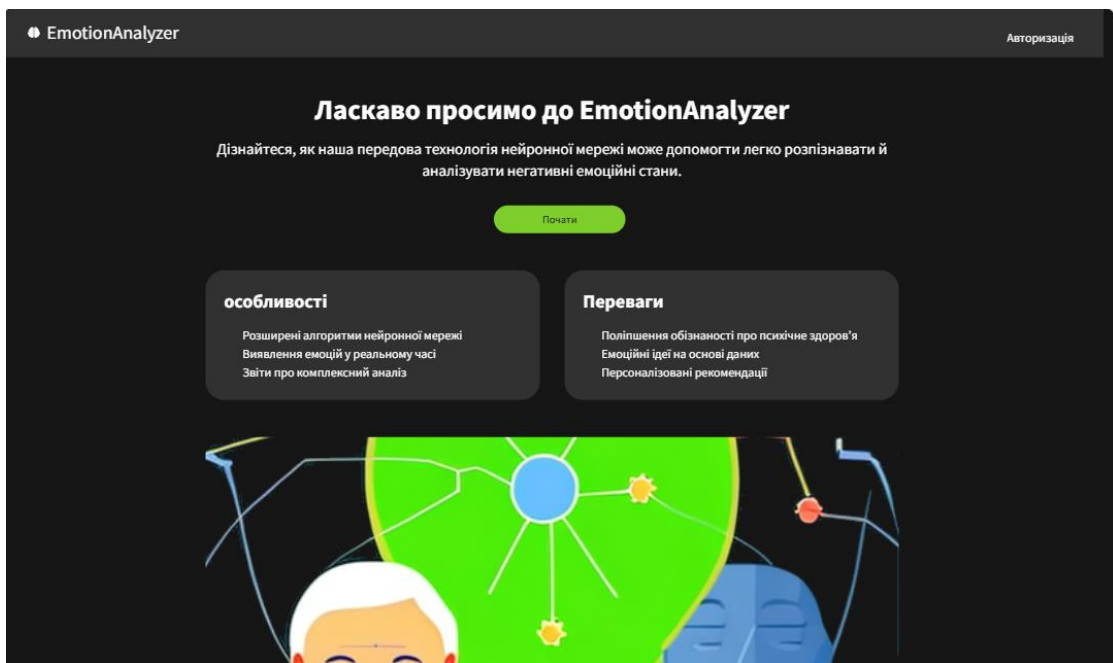


Рисунок 4.1 – Початкова сторінка сервісу

У верхній частині розміщено назву програми та кнопку для аутентифікації, що забезпечує зручну навігацію.

У центрі екрана міститься вітальне повідомлення про технологію нейронних мереж, яка допомагає розпізнавати негативні емоційні стани. Це підкреслює мету програми — полегшити розуміння емоцій, особливо негативних.

Інтерфейс включає два основні розділи. Перший розділ описує функціональність програми, зокрема "аналізатор настроїв", який використовує машинне навчання для точного розпізнавання емоцій на основі мови, міміки та невербальних сигналів. Це важливо для різних стилів спілкування.

Другий розділ акцентує переваги програми, такі як швидка обробка даних для отримання результатів у реальному часі. Це корисно в медичних та освітніх установах, де важливо своєчасно реагувати на емоційний стан користувача

Перевіримо правильність завантаження аудіофайлів через API та їх обробку. Оберемо файл, завантажемо його у систему (рис.4.2).

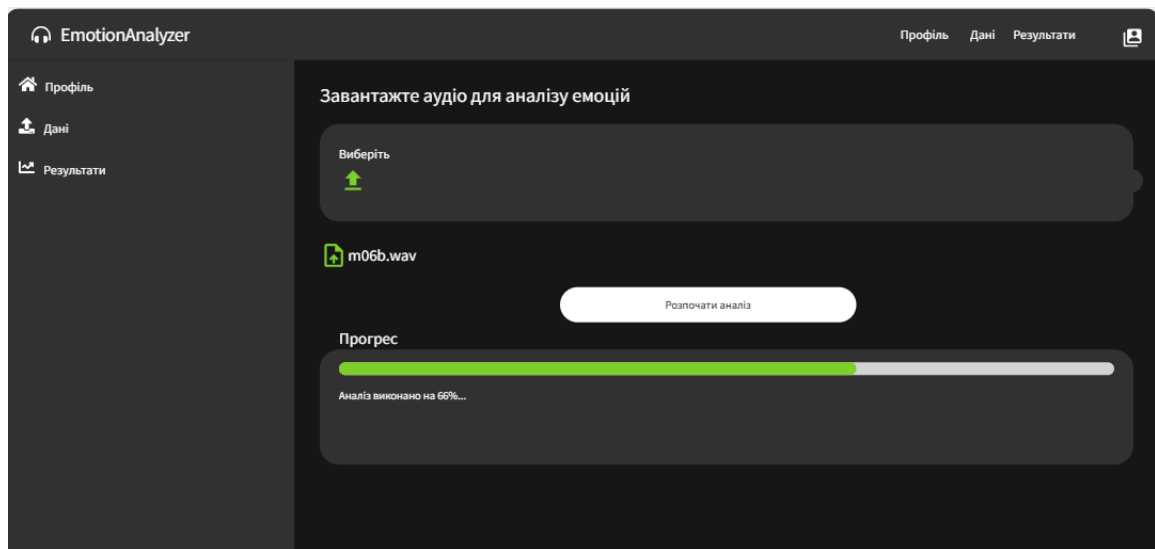


Рисунок 4.2 – Робота з аудіо файлами

Файл успішно завантажувється та зберігається в базі даних.

Рисунок демонструє інтерфейс програми "EmotionAnalyzer", що спеціалізується на аналізі емоційного стану користувача.

У верхній частині знаходиться заголовок "сприйнятий емоційний стан", який привертає увагу до основної функції програми — розпізнавання емоцій. Під заголовком розміщено три смайлика, що символізують страх, смуток і гнів, з відсотковими значеннями, які вказують на силу кожної емоції [14].

Нижче представлений блок "детальний звіт про аналіз", який пропонує користувачеві глибший аналіз емоцій, зосереджуючи увагу на негативних емоціях. Це допомагає користувачам краще усвідомити свої почуття та реакції.

Перевіримо класифікацію емоцій з використанням нейронної мережі та повернення рекомендацій на основі виявленої емоції.

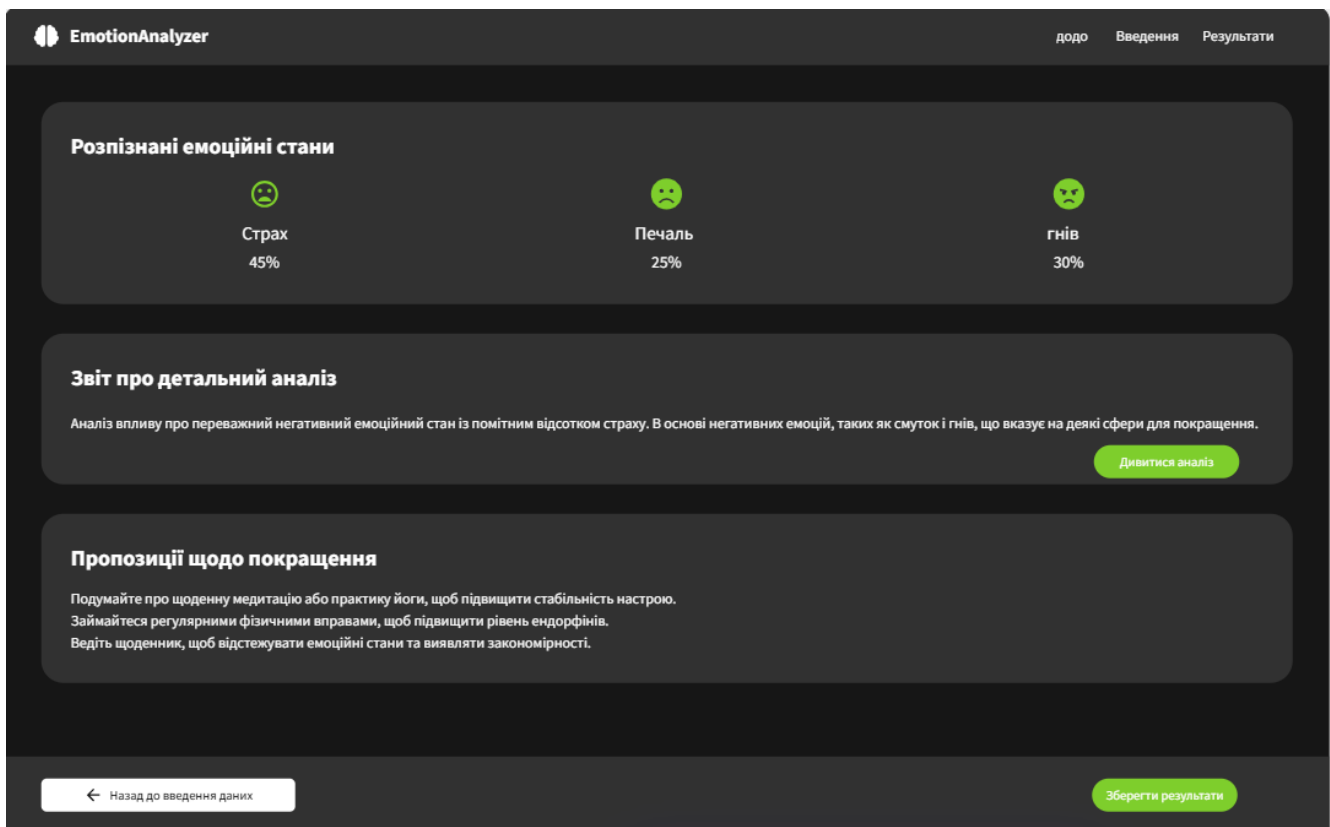


Рисунок 4.3 – Розпізнавання емоцій та надання рекомендацій

У верхній частині екрана розташований заголовок "завантажити аудіо для аналізу настроїв", який вказує на основну функцію програми. Під

заголовком є кнопка "вибрати" для завантаження аудіофайлів, що полегшує процес для користувачів. Поряд із кнопкою відображається ім'я обраного файлу, що підтверджує його успішне завантаження.

Далі на екрані з'являється індикатор виконання, який показує прогрес аналізу в зеленому кольорі, що символізує позитивний прогрес. Текст «почати аналіз» підкреслює, що процес вже розпочато.

Інтерфейс є інтуїтивно зрозумілим, що важливо для користувачів без досвіду роботи з подібними програмами. Програма підтримує різні аудіоформати, такі як WAV і MP3, що робить її універсальною.

ВИСНОВКИ

У ході виконання магістерської роботи було розроблено та протестовано систему для аналізу та розпізнавання негативних емоційних станів на основі аудіо- та відеоданих, що базується на технологіях нейронних мереж. Система продемонструвала високу точність у розпізнаванні емоцій завдяки використанню сучасних методів попередньої обробки даних, таких як спектрограми, мел-спектрограми, та глибоких архітектур, зокрема нейронних мереж.

У рамках кваліфікаційної роботи було розроблено систему, яка дозволяє класифікувати негативні емоції з використанням нейромережевих технологій.

Було вирішено такі завдання:

- записано унікальний набір мовних даних;
- проведено перцептивний експеримент;
- обрано методику обробки мовленнєвих даних. У зв'язку з тим, що було прийнято рішення про реалізацію згорткової нейронної мережі, на вхід якої подаються зображення, необхідно було перетворити аудіо-фрагменти на зображення. Було прийнято рішення перетворити кожен аудіо-фрагмент у зображення: зображення графіка основного тону, спектрограму та крейду-спектрограму;
 - розроблено та реалізовано алгоритми передобробки даних;
 - сформовано навчальні та тестові вибірки.
 - реалізовано, навчено та протестовано нейронну мережу.

Результати роботи підкреслюють перспективність використання сучасних нейромережевих підходів для розпізнавання емоційних станів, що відкриває можливості для подальшого вдосконалення системи та її застосування в реальних умовах.

СПИСОК ЛІТЕРАТУРИ

1. Ekman, P. An Argument for Basic Emotions. *Cognition and Emotion*, 1992, – 169-200 с. – Режим доступу:
<https://www.tandfonline.com/doi/abs/10.1080/02699939208411040>
2. Plutchik, R. The Nature of Emotions: Theory, Assessment, and Self-Report. *Emotion*, 2001, – 261-281 с. – Режим доступу:
<https://www.apa.org/pubs/journals/releases/emo-12380.pdf>
3. Zhang, Y., & Yu, Y.. Emotion Recognition from Speech: A Review. *IEEE Transactions on Affective Computing*, 2018,– 123-135 с. – Режим доступу:
<https://ieeexplore.ieee.org/document/8090439>
4. Goodfellow, I., Bengio, Y., & Courville, A. *Deep Learning*. MIT Press. 2016,–
Режим доступу:<https://www.deeplearningbook.org/>
5. Liu, P, & Zhang, L. Deep Learning for Emotion Recognition: A Survey. *IEEE Transactions on Affective Computing* . 2019, – Режим доступу:
<https://ieeexplore.ieee.org/document/8715960>
6. Picard, R. W. *Affective Computing*. MIT Press. 1997, – Режим доступу:
<https://www.mitpress.mit.edu/books/affective-computing>
7. Russell, J. A. A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 1980, – 1161-1178 с. – Режим доступу:
<https://psycnet.apa.org/doiLanding?doi=10.1037%2Fh0087950>
8. Scherer, K. R. What are Emotions? And How Can They Be Measured? *Social Science Information*, 2005, – 695-726 с. – Режим доступу:
<https://journals.sagepub.com/doi/abs/10.1177/0539018405058216>
9. Sethi, R., & Gupta, A. Emotion Recognition Using Deep Learning Techniques: A Review. *Journal of King Saud University - Computer and Information Sciences* . 2019, – Режим доступу:
<https://www.sciencedirect.com/science/article/pii/S1319157818300614>

10. Kotsia, I, & Zafeiriou, S. Emotion Recognition in the Wild: A Survey. 2016, – Режим доступа:
<https://arxiv.org/abs/1610.07919>
11. LeCun, Y, Bengio, Y, & Haffner, P. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 1998,– 2278-2324 с.
– Режим доступа:
<https://ieeexplore.ieee.org/document/726791>
12. Möller, S., & Koerner, E. A Survey on Emotion Recognition in Human-Computer Interaction. Journal of Computational Science, 2016, – 1-12 с.
– Режим доступа:
<https://www.sciencedirect.com/science/article/pii/S1877750316301037>
13. Kraemer, E., & de Ruiter, J. P. On the Relation between Speech and Emotion in Human-Computer Interaction. Artificial Intelligence Review, 2005, – 391-419 с. – Режим доступа:
<https://link.springer.com/article/10.1007/s10462-005-1236-2>
14. Yoon, H., & Kim, H. A Review of Emotion Recognition Based on Deep Learning. Applied Sciences, 2018, – 1366 с. – Режим доступа:
<https://www.mdpi.com/2076-3417/8/8/1366>
15. Susskind, J. M., & Anderson, A. K. Emotional Responses to Visual and Auditory Stimuli. Emotion, 2007, – 446-460 с. – Режим доступа:
<https://psycnet.apa.org/doiLanding?doi=10.1037%2F1528-3542.7.3.446>