

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

Факультет прикладної математики

Кафедра прикладної математики

«На правах рукопису»

УДК 519.688:004.855.5

«До захисту допущено»

Завідувач кафедри

_____ Олег ЧЕРТОВ

« ____ » _____ 2023 р.

**Магістерська дисертація
на здобуття ступеня магістра
за освітньо-науковою програмою «Наука про дані та математичне
моделювання»
зі спеціальності 113 «Прикладна математика»
на тему: «Математичне та програмне забезпечення системи розпізнавання
емоцій за аудіозаписами»**

Виконала: студент 2 курсу, групи КМ-11мн
Коваленко Олександра Петрівна

Керівник:

Доцент, канд. фіз.-мат. наук, доцент
Третиник Віолета Вікентіївна

Консультант з нормоконтролю:
старший викладач,
Мальчиков Володимир Вікторович

Рецензент:

професор кафедри математичних методів
системного аналізу ІІСА, д-р
техн. наук, доцент
О.Ю.Зайченко

Засвідчую, що в цій магістерській
дисертації немає запозичень із праць інших
авторів без відповідних посилань.
Студент

Київ — 2023

Національний технічний університет України

«Київський політехнічний інститут

імені Ігоря Сікорського»

Факультет прикладної математики

Кафедра прикладної математики

Рівень вищої освіти — другий (магістерський)

Спеціальність – 113 «Прикладна математика»

Освітньо-науковою програма «Наука про дані та математичне моделювання»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Олег ЧЕРТОВ

«__» _____ 2023 р.

ЗАВДАННЯ

на магістерську дисертацію студентці

Коваленко Олександрі Петрівні

1. Тема дисертації: «Математичне та програмне забезпечення системи розпізнавання емоцій за аудіозаписами», науковий керівник дисертації Третиник Віолета Вікентіївна, кандидат фіз.-мат. наук, доцент, затверджені наказом по університету від «30» березня 2023 р. № 1359-С.

2. Термін подання студентом дисертації: «15» травня 2023 р.

3. Об'єктом дослідження є способи організації засобів нейромережевого розпізнавання емоцій за аудіозаписами.

Теоретичні інструменти:

1) методи і алгоритми реалізації архітектур нейронних мереж: нейронні мережі, аналіз та класифікація даних за заданими параметрами; обрання ознак;

2) методи оптимізації гіперпараметрів: гіперпараметри, оптимізація, пошук архітектури нейронної мережі, автоналаштування; обрання моделі.

Існуючі засоби: класи, ознаки, характеристики, область застосування, вартість володіння.

4. Предмет дослідження: математичне та програмне забезпечення системи розпізнавання емоцій за аудіозаписами на основі нейромережевої моделі. Застосування нейронних мереж на етапі попереднього тренування для розпізнавання емоцій за аудіозаписами, порівняльний аналіз методів оптимізації, що використовуються при навчанні нейронних мереж.
5. Перелік завдань, які потрібно розробити: знайти великий та змістовний датасет, проаналізувати на змістовність, виконати аналіз моделей, котрі використовуються для діагностики і вибрати одну. Застосувати алгоритми машинного навчання та нейронні мережі для знаходження закономірності, розробити інтерфейс.
6. Орієнтовний перелік ілюстративного матеріалу: приклади зображень спектрограм, архітектура моделі нейронної мережі, архітектура розробленого програмного забезпечення, зображення з ілюстрацією результатів випробувань.
7. Орієнтовний перелік публікацій: тези «Автоматизація підбору гіперпараметрів lstm для задачі розпізнавання емоцій за аудіозаписами».
8. Дата видачі завдання: «1» лютого 2023 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Грунтовне ознайомлення з предметною областю	15.09.2022	
2	Визначення структури магістерської дисертації; вивчення літератури, пошук додаткової літератури	01.10.2022	
3	Проведення наукового дослідження; робота над статтею за результатами наукового дослідження	20.10.2022	
4	Проведення наукового дослідження; робота над першим розділом магістерської дисертації	15.01.2023	

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
5	Робота над другим розділом магістерської дисертації	15.02.2023	
6	Робота над третім розділом магістерської дисертації; підготовка статті за результатами наукового дослідження; розроблення програмного забезпечення	01.03.2023	
7	Завершення роботи над основною частиною магістерської дисертації	15.04.2023	
8	Оформлення текстової і графічної частин магістерської дисертації	05.05.2023	

Студент

Олександра

КОВАЛЕНКО

Науковий керівник дисертації

Віолета ТРЕТИНИК

РЕФЕРАТ

Дисертацію виконано на 81 аркушах, вона містить 2 додатки та перелік посилань на використані джерела з 22 найменувань. У роботі наведено 40 рисунків та 2 таблиці.

Актуальність теми. Технології сучасності стрімко розвиваються. Сьогодні потужність обчислювальної техніки сильно виросла порівняно з попереднім століттям. Розвиток машинного навчання припадає ще на 60-ті роки минулого століття і швидкість розвитку цієї галузі зростає з кожним роком в геометричній прогресії. Протягом останніх двох десятиліть ця галузь є однією із провідних у комп'ютерних науках. Задачі розпізнавання вирішені використанням нейронних мереж є унікальними. Використання нейронних мереж у багатьох сферах стрімко покращує прогрес. Дана робота, про розпізнавання емоцій людини за аудіозаписами спрямована на вирішення проблем галузях: військовій, розважальній, освітній та медичній.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно з планом науково-дослідних робіт кафедри прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Мета і задачі дослідження. Метою дисертаційної роботи є розробка математичного та програмного забезпечення системи розпізнавання емоцій за аудіозаписами. Для досягнення цієї мети необхідно вирішити наступні задачі:

- дослідити методи навчання нейронних мереж;
- розглянути і проаналізувати архітектури та принципи організації нейронних мереж та методів вирішення задач класифікації;
- способи оптимізації параметрів згорткових мереж;
- запропонувати структуру системи класифікації аудіозаписів;

- розробити модель програмного забезпечення з запропонованою структурою;

- провести експериментальне дослідження характеристик системи.

Об'єктом дослідження є способи організації засобів машинного навчання для розпізнавання емоцій за аудіозаписами.

Теоретичні інструменти:

- методи і алгоритми реалізації архітектур нейронних мереж: нейронні мережі, аналіз та класифікація даних за заданими параметрами; обрання ознак;

- методи оптимізації гіперпараметрів: гіперпараметри, оптимізація, пошук архітектури нейронної мережі, автоналаштування; обрання моделі;

Існуючі засоби: класи, ознаки, характеристики, область застосування, вартість володіння.

Предмет дослідження:

Математичне та програмне забезпечення системи розпізнавання емоцій за аудіозаписами на основі нейромережевої моделі.

Застосування нейронних мереж на етапі попереднього тренування для розпізнавання емоцій за аудіозаписами, порівняльний аналіз методів оптимізації, що використовуються при навчанні нейронних мереж.

Методи дослідження. Для розв'язання поставленої задачі використовувались такі методи: методи машинного навчання (для розробки моделі нейронної мережі); методи оптимізації (для пошуку найкращого налаштування системи); методи обробки даних (для попередньої підготовки вихідних даних); методи теорії алгоритмів та програмування (для програмної реалізації розроблених алгоритмів).

Наукова новизна одержаних результатів складається з таких положень:

- 1) вперше застосовано та поєднано системну інженерію, включено методи та моделі Data Science, які використовуються для реалізації системи розпізнавання емоцій людини за аудіозаписами;

- 2) запропоновано використання інтегрованої моделі класифікації емоцій за аудіозаписами на основі згорткових нейронних мереж з автоматизованим підбором гіперпараметрів.

Практичне значення одержаних результатів. Полягає в тому, що розроблена система розпізнавання емоцій за аудіозаписами на основі згорткової нейронної мережі, яка може використовуватись для аналізу розмови та емоційного стану людини, що в подальшому можна використовувати в рекомендаційних системах, військовій справі і тд.

Апробація результатів дисертації. Основні положення й результати роботи дисертації доповідались та опубліковані у матеріалах XV наукової конференції магістрантів та аспірантів «Прикладна математика та комп'ютинг - ПМК-2022» (Київ 16-17 листопада 2022 року).

Публікації. Результати дисертації викладено в науковій праці: тези «Автоматизація підбору гіперпараметрів lstm для задачі розпізнавання емоцій за аудіозаписами» на XV конференції магістрантів та аспірантів «Прикладна математика та комп'ютинг - ПМК-2022»

Ключові слова: згорткова нейронна мережа, розпізнавання емоцій, класифікація, регуляризація, класифікація, навчання мережі, алгоритми оптимізації, оптимізація гіперпараметрів.

ABSTRACT

The dissertation is made up of 81 sheets, it contains 2 appendices and a list of references to used sources from 22 names. The work contains 40 figures and 2 tables.

Actuality of theme. Modern technologies are developing rapidly. Today, the power of computing has greatly increased compared to the previous century. The development of machine learning dates back to the 60s of the last century, and the speed of development of this field is growing exponentially every year. For the past two decades, this field has been one of the leading in computer science. Recognition tasks solved using neural networks are unique. The use of neural networks in many fields is rapidly improving progress. This work, about recognizing human emotions from audio recordings, is aimed at solving problems in the following fields: military, entertainment, education, and medicine.

Connection of work with scientific programs, plans, topics. The dissertation work was carried out in accordance with the plan of research works of the Department of Applied Mathematics of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".

The purpose and objectives of the research. The aim of the dissertation is to develop a mathematical and software system for recognizing emotions from audio recordings.

To achieve this goal, the following tasks must be solved:

- 1) research methods of learning neural networks;
- 2) consider and analyze the architecture and principles of the organization of neural networks and methods of solving classification problems;
- 3) ways to optimize parameters of convolutional networks;
- 4) propose the structure of the audio recording classification system;
- 5) develop a software model with the proposed structure;
- 6) conduct an experimental study of system characteristics.

The object of research is ways of organizing machine learning tools for recognizing emotions from audio recordings.

Theoretical tools:

- 1) methods and algorithms for implementing neural network architectures: neural networks, data analysis and classification according to specified parameters; selection of signs;
- 2) hyperparameter optimization methods: hyperparameters, optimization, neural network architecture search, auto-tuning; choosing a model.

Existing means: classes, features, characteristics, scope, cost of ownership.

Subject of study:

Mathematical and software development of a system for recognizing emotions from audio recordings based on a neural network model.

Application of neural networks at the pre-training stage for emotion recognition from audio recordings, comparative analysis of optimization methods used in training neural networks.

Research methods. The following methods were used to solve the problem: machine learning methods (for developing a neural network model); optimization methods (to find the best system setting); data processing methods (for preliminary preparation of initial data); algorithm theory and programming methods (for software implementation of developed algorithms).

The scientific novelty of the obtained results consists of the following provisions:

- 1) for the first time, system engineering was applied and combined, Data Science methods and models were included, which are used to implement a system for recognizing human emotions from audio recordings;
- 2) the use of an integrated model of emotion classification based on audio recordings based on convolutional neural networks with automated selection of hyperparameters is proposed.

Practical significance of the obtained results. It consists in the fact that a system for recognizing emotions from audio recordings has been developed based on a convolutional neural network, which can be used to analyze a person's conversation and emotional state, which can later be used in recommender systems, military affairs, etc.

Approbation of the results of the dissertation. The main provisions and results of the dissertation work were reported and published in the materials of the XV scientific conference of master's and postgraduate students "Applied mathematics and computing - PMK-2022" (Kyiv, November 16-17, 2022).

Publications. The results of the dissertation were presented in a scientific work: theses "Automation of the selection of lstm hyperparameters for the problem of emotion recognition from audio recordings" at the XV conference of master's and postgraduate students "Applied mathematics and computing - PMK-2022"

Keywords: convolutional neural network, emotion recognition, classification, regularization, classification, network learning, optimization algorithms, hyperparameter optimization.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	13
Вступ.....	15
1 ОГЛЯД ІСНУЮЧИХ РІШЕНЬ	17
1.1 Spotify	17
1.2 Deemoji	18
1.3 Audeering	20
1.4 Висновок до розділу.....	21
2 ОГЛЯД МАТЕМАТИЧНИХ РІШЕНЬ	22
2.1 Перетворення аудіофайлів.....	23
2.1.1 Метод MFCC.....	24
2.1.1 Метод ІНС	26
2.2 Методи класифікації емоцій.....	27
2.2.1 Decision Tree	28
2.2.2 MLP	31
2.2.3 CNN.....	34
2.3 Огляд методів для оптимізації параметрів	37
2.3.1 Random Search.....	39
2.3.2 BO.....	41
2.3.3 Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization	42
2.4 Висновки до розділу.....	45
3 МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ	47
3.1 Метрики якості	47
3.2 MFCC	52
3.3 CNN.....	53
3.4 Оптимізація	55

3.5 Висновок до розділу.....	57
4 ОПИС СИСТЕМИ	58
4.1 Складові системи.....	58
4.2 Опис функціональних компонентів	60
4.3 Обробка аудіофайлів	60
4.4 Діаграма послідовності.....	64
4.5 Інструкція роботи з системою розпізнавання емоцій.....	66
4.6 Висновки до розділу компонентна модель системи.....	67
5 ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ	68
5.1 Алгоритм роботи програмного забезпечення	69
5.2 Опис даних	70
5.3 Архітектура нейронної мережі	73
5.4 Опис результатів.....	74
5.5 Висновки до розділу програмне забезпечення.....	76
ВИСНОВКИ.....	77
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	79
Додаток А Лістинги програм	82
Додаток Б Ілюстративний матеріал.....	86

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AUDEERING - компанія, що спеціалізується на розробці програмного забезпечення для аналізу голосу та емоційного стану людини.

BO - скорочення від "Bayesian Optimization" - метод оптимізації гіперпараметрів машинного навчання, який базується на байєсівській інференції.

CNN - скорочення від "Convolutional Neural Network" - тип штучної нейронної мережі.

CREMA-D - база даних аудіофайлів, яка використовується в дослідженнях емоцій в голосі людини.

Confusion matrix - матриця, що використовується для оцінки точності класифікатора в машинному навчанні.

Cross-validation - метод оцінки ефективності моделі машинного навчання.

DECISION TREE - алгоритм машинного навчання для прийняття рішень, що базується на дереві рішень.

DeerMojі - модель глибокого навчання, яка використовується для аналізу та генерації емоційних виразів на основі тексту.

F1-score - метрика, яка використовується для оцінки точності моделі машинного навчання, яка об'єднує точність і повноту.

FN - скорочення від "False Negative" - тип помилки в класифікації, коли істинний клас об'єкту є позитивним, але модель визначає його як негативний.

FP - скорочення від "False Positive" - тип помилки в класифікації, коли істинний клас об'єкту є негативним, але модель визначає його як позитивний.

Hyperband - алгоритм оптимізації, який використовується в глибинному навчанні для ефективного гіперпараметричного налаштування моделей.

IHC - імуногістохімічне дослідження, метод візуалізації антигенів або біомолекул в тканинах або клітинах за допомогою специфічних антитіл.

MFCC - мел-частотні кепстральні коефіцієнти.

MLP - багатошаровий перцептрон, нейромережа.

Precision - метрика, що вимірює відношення кількості правильно визначених позитивних класів до загальної кількості класифікованих як позитивні.

RAVDESS - ресурс для аналізу голосу та емоційно-розпізнавальних досліджень.

Random Search - алгоритм оптимізації, що використовується для гіперпараметричного налаштування моделей шляхом випадкового пошуку в просторі гіперпараметрів.

Recall - метрика, що вимірює відношення кількості правильно визначених позитивних класів до загальної кількості позитивних класів у даних.

ReLU - це акронім від "Rectified Linear Unit", який використовується як функція активації в нейронних мережах.

Spotify - це музичний стрімінговий сервіс, який надає доступ до мільйонів пісень, аудіокниг, подкастів та інших змістових матеріалів з усього світу.

TN - це скорочення від "True Negative" і відноситься до значення в матриці плутанини, коли система правильно класифікує відсутність певного об'єкту.

TP - це скорочення від "True Positive" і відноситься до значення в матриці плутанини, коли система правильно класифікує наявність певного об'єкту.

TESS - це скорочення від "Toronto Emotional Speech Set" і представляє собою набір аудіофайлів з різними емоційними викликами, який використовується для досліджень в галузі розпізнавання емоцій в голосі.

Вступ

Існує багато різних підходів до класифікації емоцій. Така задача може здаватись на перший погляд для людини не важкою, оскільки ми використовуємо одночасно всі інструменти для розпізнавання:

- 1) аналізуємо голос (аудіосигнал);
- 2) аналізуємо міміку;
- 3) аналізуємо жести;
- 4) аналізуємо обставини в яких опинились і т.д.

Проте іноді ми опиняємось в ситуації, коли є можливість проаналізувати лише один факт і в цій роботі заплановано розглянути розпізнавання емоцій лише за одним аспектом – аудіозаписом. Система, яка використовує лише одну модальність - голосові записи - стає складнішою і вимагає мікрофон як додатковий засіб.

Обраний підхід є насправді доволі популярним і необхідним. Здатність автоматизованого процесу розуміння настроїв та намірів людей, можуть знайти застосування у безпеці, медицині, екстрених службах, телемаркетингу, соціальних установах колл-центрах та службах підтримки. Це може допомогти людям з обмеженими здібностями розуміти емоції співрозмовника, а власника бізнесів допомагає направити зусилля свої працівників у правильному напрямку.

Надання комп'ютерам здатності розуміти емоційні стани, виражені в людському мовленні, є важливою доповнювальною функцією взаємодії між людиною та машинами. Це може допомогти забезпечити більш повну та точну взаємодію між користувачами та технологіями, а також покращити якість обробки інформації, що вводиться користувачами.

Розпізнавання емоцій у людей є складною задачею, оскільки вона потребує точного визначення ознак, що впливають на цей процес. Визначення цих ознак залежить від використовуваних програм і баз даних, а також від різних властивостей мовлення, таких як зміст, стиль, мовець і стиль. Ці фактори можуть вплинути на

висоту звуку та енергетичні контури, ускладнюючи задачу точного виділення емоційних характеристик.

Існують різні методи виявлення емоцій у мовленні, деякі з яких базуються на акустичних параметрах звуку, таких як його частота, інтенсивність, довжина, амплітуда та спектральний склад. Інші підходи використовують алгоритми машинного навчання для автоматичного розпізнавання емоцій. Однак, вирішення задачі вимагає ретельного аналізування як суб'єктивних, так і об'єктивних факторів, які впливають на розпізнавання емоцій, такі як стать, мовлення та стиль. Найкращі результати можна досягти, використовуючи різноманітні підходи та алгоритми, які можуть адаптуватись до різних контекстів і властивостей мовлення.

Основною метою даного дослідження є вирішення проблем розпізнавання емоцій на основі обраних актуальних та ефективних підходів.

1 ОГЛЯД ІСНУЮЧИХ РІШЕНЬ

В даному розділі подається короткий опис наявних систем для визначення емоційного стану особи, а також аналізуються інструменти та методи, що були використані при розробці таких систем. Також розглядаються перспективні напрямки застосування комплексних систем для визначення осіб та їх емоцій.

Емоції відіграють суттєву роль у міжособистісних відносинах, і багато додатків реалізують можливості інтерпретації емоцій особи на основі його висловленого тексту. Інструменти для розпізнавання емоцій у мовленні допомагають таким додаткам, проте їхнє впровадження стає складним через обмежену наявність збалансованих навчальних даних та неоднозначність щодо необхідних функціональних можливостей для побудови класифікаційної моделі.

1.1 Spotify

Spotify – це додаток, завдяки якому легко знайти потрібну музику чи подкаст на будь-який момент – на телефоні, комп'ютері, планшеті тощо. Spotify містить мільйони треків і подкаст[1].

Така велика компанія, мала наміри, ще у 2018 впровадити до своєї системи розпізнавання емоцій за аудіозаписами, що своєю чергою, стало б складовою системи для ідентифікації смакових ознак користувачів і сприяло б поліпшенню якості складання рекомендацій в системі.

Така розробка мала б працювати не лише з основними охопленими звуками, але і також з шумами, а після опрацювання аудіозаписів ідентифікувати вміст файлу.



Рисунок 1.1 – Додаток Spotify [1]

Ще одна з недосконалостей такого підходу полягає в тому, що він негнучкий, оскільки з часом контент має здатність набридати і в правильний момент дерево не здатне запропонувати альтернативу пісень або жанрів, в той час, як підхід з розпізнаванням емоцій здатне залишатись гнучким, розрізняти емоцію користувача від прослуханого контенту і врешті на основі результату від такого рішення, можна зважитись, чи залишається смак незмінним, або варто користувачу запропонувати нові альтернативи.

Отже, можна прийти до висновку, що система розпізнавання емоцій за аудіозаписами є дійсно корисною та актуальною в рамках системи виявлення вподобань користувачів, проте компанію, Spotify ще не впровадила такий інструмент до свого додатка, тож далі в роботі буде розглянуто інструменти, підходи та засоби, для реалізації такої ідеї.

1.2 Deemoji

Deemoji — це модель, навчена 1,2 мільярда твітів із емодзі, щоб зрозуміти, як мова використовується для вираження емоцій. Завдяки переносному навчанню

модель може отримати найсучаснішу продуктивність у багатьох завданнях текстового моделювання, пов'язаних з емоціями[2].

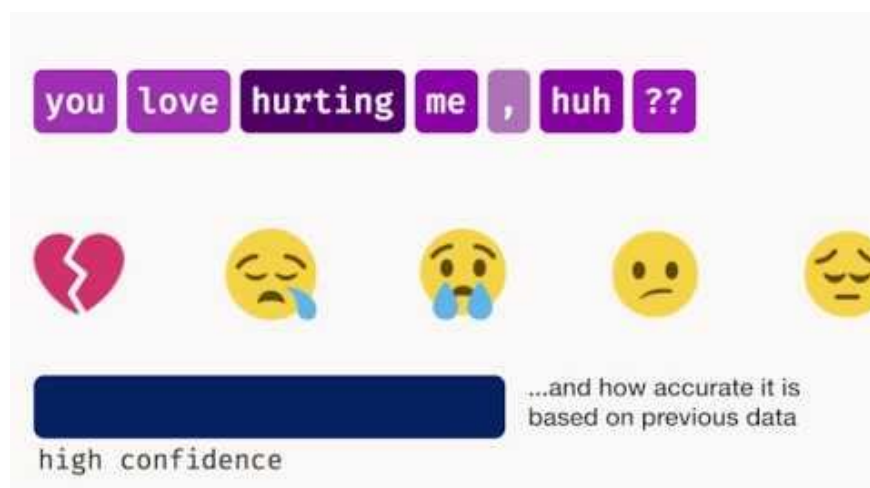


Рисунок 1.2 – Додаток DeerpMoji [2]

Хоча стверджується, що таке програмне забезпечення здатне працювати і з аудіозаписами та розпізнавати емоції за сказаним, проте такий підхід не орієнтується на справжні емоції аудитора. Такий підхід лише визначає характер проговорених слів, оскільки в основі такого підходу лежить обробка тексту. Тобто таке рішення, ніколи не зможе добре працювати з сарказмом та іронією, коли добрі слова проговорені з негативними емоціями та акцентами, чи навпаки доволі негативні використані в якості жарту. Окрім цього алгоритм такого програмного забезпечення для визначення емоцій за аудіозаписом передбачає, що є ще додаткові інструменти, які здатні відтворювати текст, що був прописаний в записі голосу, а після цього лише, опрацьовувати текст і буквально розрізняти емоції за клішованими фразами та характером їх використання.

Тобто така задача є більш затратною і з точки зору ресурсів і з точки зору часу, проте хоча така задача є затратнішою, це не говорить про те, що такий підхід кращий, а навпаки плоскіший, бо не враховує реальний емоційний стан диктора, його особливості, манери та реакції.

1.3 Audeering

Компанія AUDEERING розробляє рішення, що охоплюють широкий спектр напрямків у галузі штучного інтелекту та машинного навчання, в тому числі і досліджень емоцій та їх розпізнавання з аудіо.

AUDEERING — провідний світовий новатор в області штучного інтелекту в роботі з голосом. Їх технології можуть визначити емоції та інформацію навіть про здоров'я за голосом[3].



Рисунок 1.3 – Додаток AUDEERING [3]

Компанія AUDEERING використовує голосовий штучний інтелект для тестування тільки розроблених продуктів або виведення їх якості на новий рівень, а саме аналіз звукових доріжок та виявлення емоцій, того як клієнти сприймають певну продукцію, дозволяє професійно зростати та покращувати розробки.

До того ж розробники стверджують, що з їх системою, програмне забезпечення може працювати в режимі реального часу в пристроях ARM з низьким рівнем споживання ресурсів, оскільки систему оптимізовано.

Однак варто зауважити, що база даних для навчання нейронної мережі у цьому рішенні закрита і таку систему не можна буде оптимізувати та адаптувати для конкретних задач та замовників, до того ж використання такого програмного засобу не є безпечним, оскільки користується записами голосів для класифікації та зберігає їх на серверах іншої країни, що для нашої країни виходить за рамки закону.

1.4 Висновки до розділу

Отже, в цьому розділі було розглянуто три системи розпізнавання емоцій, з різними підходами та для різних цілей. З цього також можна зауважити актуальність та важливість розв'язання такої задачі. Серед головних недоліків можна виділити закритість систем, неможливість протестувати, розібратись з програмної та математичної точки зору з принципами й алгоритмами на який базуються створені програмні забезпечення.

2 ОГЛЯД МАТЕМАТИЧНИХ РІШЕНЬ

В нас час неймовірно актуальним підходом до розв'язання задач стало машинне навчання. З кожним днем зростає, популярність задач розпізнавання, класифікацій та аналізу штучного інтелекту. В цьому розділі буде розглянуто та проаналізовано, які підходи з обраних кращі за інші, а також обґрунтовано вибір допоміжних інструментів, які вплинуть на моделі машинного навчання відповідно до вимог якості. Побудова моделі машинного навчання включає сім етапів:

- 1) збір даних;
- 2) обробка даних;
- 3) виберіть модель;
- 4) навчання модель;
- 5) оцінка точності;
- 6) підбір гіперпараметрів;
- 7) передбачення.

Усі кроки важливі для побудови моделі машинного навчання. Тож в цьому розділі буде розглянуто та охоплено всі етапи на шляху до класифікації емоцій людини за аудіозаписами.

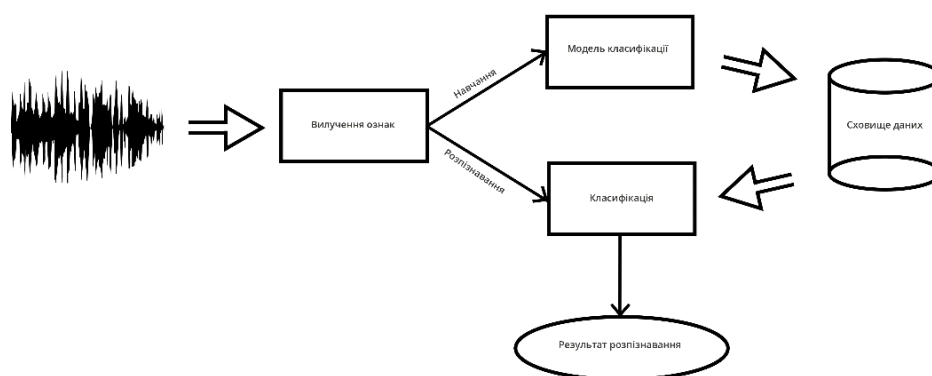


Рисунок 2.1 – Схематичне відображення етапів розпізнавання емоцій

2.1 Перетворення аудіофайлів

Розпізнавання мовлення є завданням, що потребує контрольованого навчання. У даному завданні використовується аудіосигнал як вхідний сигнал, з метою розпізнавання емоції. Однак, в якості інструменту для розпізнавання емоцій будуть використовуватись алгоритми машинного навчання, які потребують даних у вигляді векторного представлення. Тож потрібно розглянути задачу вилучення особливостей та характеристик з аудіо файлів.

Голос є важливою рисою людини в природній взаємодії / спілкуванні між людьми для ідентифікації людини. Отже, голос може розглядатися як біометричний показник для розпізнавання або ідентифікації особи.

Мова людини з точки зору фізики – це передача звукових хвиль з певними частотами. В середньому для чоловіка вона становить 125 Гц, в той час, як для жінок приблизно 210 Гц. Такі хвилі людина створює за допомогою голосових зв'язок. Коли ми видихаємо повітря з легень, воно розкриває голосові зв'язки. Потік повітря прискорюється, а тиск на голосові зв'язки падає. Це знову закриває їх. Як тільки потік повітря знижується, тиск збільшується і знову відкриває їх. Ці відкриті та закриті цикли продовжуються і виробляють ряд частот звукових хвиль [4]. Також варто зазначити, що слова складаються зі складів, тобто з набору звуків.

Окрім того, що звуки можна поділити на голосні та приголосні, в кожного звука є ряд своїх особливостей та характеристик, нижче на рис. 2.2 зображено спектрограму за допомогою, якої можна відобразити звукові хвилі. Такий підхід наочно демонструє суттєву відмінність між двома звуками.

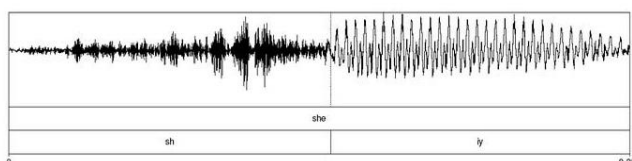


Рисунок 2.2 – Ілюстративне відображення звукового сигналу[5]

Наші артикуляції створюють різноманітні форми голосового тракту, які виробляють різні резонанси. Вони діють як фільтри під час придушення чи посилення частот [4].

Основним завданням будь-якого розпізнавання мовців є виділення корисних функцій і створення значущих шаблонів моделей мовців. Процес виділення ознак включає розрахунок послідовності ознак для кожного короткочасного кадру вхідного сигналу. Цей процес здійснюється з припущенням, що невеликий сегмент мовлення є достатньо стаціонарним, щоб забезпечити значиме моделювання, що дозволяє покращити точність розпізнавання.

Функції обробки мовлення розділяють на:

- 1) короткочасні спектральні;
- 2) джерело голосу;
- 3) спектрально-часові;
- 4) просодичні;
- 5) ознаки високого рівня на основі їх фізичних інтерпретацій[6].

Надалі будуть досліджені деякі короткочасні спектральні характеристики, які використовуються у процесі перетворення даних, отриманих з аудіосигналу :

- 1) MFCC;
- 2) ІНС.

2.1.1 Метод MFCC

Дерево — це техніка виділення звукових характеристик, яка виділяє специфічні параметри мовця з мови [7]. Ця техніка часто використовується для вилучення характеристик із аудіосигналу. Це популярний метод виділення спектральних характеристик мовлення за допомогою фільтрів із перетворенням Фур'є.

Сприйняття звуку машинами відрізняється від сприйняття звуку людьми. У нас є вища роздільна здатність на нижчих частотах, ніж на вищих. Тому ми можемо легко розрізнити звуки на частотах 200 Гц та 300 Гц від звуків на 1500 Гц та 1600 Гц, навіть якщо різниця між ними становить 100 Гц. У машин роздільна здатність однакова на всіх частотах. Як наслідок, застосування моделювання властивостей слуху людини на етапі виділення ознак може покращити продуктивність моделі.

Отже, ми будемо використовувати шкалу мел, щоб зіставити фактичну частоту з частотою, яку сприйматимуть люди. Формула для відображення наведена нижче.

$$Mel(f) = 1127 \ln \left(1 + \frac{f}{700} \right) \quad (1.1)$$

Мел-рекуррентний кепстр (MFCC) показує вхідний звук, як короточасний спектр потужності на основі лінійного косинуса. MFCC використовуються для відновлення звуку з аудіофайлів формату wav. На рис. 2.3 представлено блок-схему алгоритму вилучення ознак MFCC.

Створення звуку людського голосу залежить від форми вокальної доріжки. Форма голосового тракту може впливати на точність розпізнавання голосових команд та мелодій. Використання вбудованих функцій часу, діапазону потужності мовного сигналу та MFCC можуть допомогти виявити голос.



Рисунок 2.3 – Схематичне відображення алгоритму вилучення ознак MFCC

MFCC є популярним інструментом для систем, що працюють з голосом і безпосередньо для систем розпізнавання мовлення.

Обчислення MFCC складається з кількох кроків:

- 1) спочатку проводиться перетворення Фур'є вхідного сигналу;
- 2) використовується функція Ханна для перетворення потужностей спектра на шкалу Mel;
- 3) взяття логарифмічних ступенів Mel на кожній частоті;
- 4) застосування дискретного косинусного перетворення для отримання MFCC, які представляють амплітуди результуючого спектру.

2.1.1 Метод ІНС

ІНС — це техніка виділення звукових характеристик, яка виділяє модель внутрішньої волоскової клітини описує спосіб трансдукції у внутрішній волосковій клітині/слуховому нерві на рівні, досить близькому до фізіології. Нижче на рис.2.4 представлено блок-схему техніки вилучення ознак ІНС:

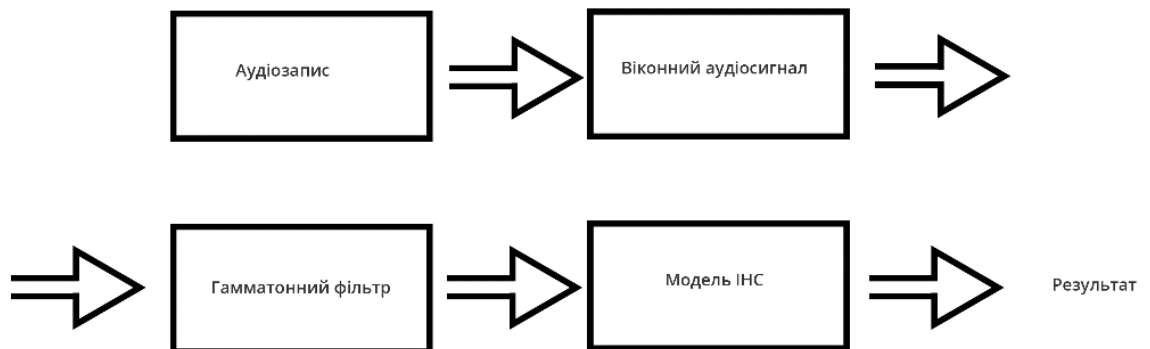


Рисунок 2.4 – Схематичне відображення алгоритму ІНС

Модель Меддіса для внутрішніх волоскових клітин пояснює, як звук перетворюється на нейронні сигнали в слуховому нерві. Коли звукові хвилі потрапляють у вухо, рух внутрішніх волоскових клітин викликає деполяризацію та створює рецепторний потенціал. Це, у свою чергу, призводить до викиду нейромедіаторів у невеликий проміжок між волосковими клітинами та слуховим нервом, відомий як синаптична щілина. Якщо концентрація нейромедіаторів перевищує певний поріг, спрацьовує потенціал дії або сплеск. У процесі моделювання ІНС нейротрансмітери рухаються через проникну мембрану від внутрішньої волоскової клітини до щілини. Ця проникність змінюється залежно від амплітуди стимулу, тобто вона не є постійною.

2.2 Методи класифікації емоцій

Класифікація емоцій є достатньо складним завданням в тому числі і для людини. Не багато хто може з впевненістю ствердити, що вміє точно розпізнавати емоції; однак людям простіше, бо вони мають можливість розпізнавати емоції аналізуючи мову тіла, вираз обличчя і вже потім тембр голосу.

Розпізнавання емоцій за допомогою лише аудіозапису є потенційно значним кроком у майбутнє, оскільки воно представляє величезну різноманітність випадків використання. Такий підхід розглядає розпізнавання емоцій за допомогою лише однієї модальності, голосових записів, які робить його більш складним. Таким чином, він використовує один додатковий носій — мікрофон, який також може захопити деякий шум. Вирішення цієї задачі може сприяти розвитку машин, оскільки емоції - це щось, що найбільше наближує комп'ютер до людського сприйняття. Попередні приклади показують, що такий інструмент може допомогти розуміти емоції не тільки машинами, а й людьми з менш гострою здатністю розрізняти емоції свого співрозмовника або людей з вадами сприйняття.

В цьому розділі мається на меті порівняти алгоритми, які здатні навчатись та розпізнавати, який з них буде давати кращий результат. Оскільки таких досліджень небагато і теоретичне обґрунтування вибору методу може бути не коректним для вузько направленого рішення, тож ефективність методів буде порівняно безпосередньо за результатами класифікації та обраними метриками якості і наочно буде продемонстровано результати, на основі яких і буде визначено, який метод краще обрати для вирішення задачі розпізнавання емоцій людини за аудіозаписами.

Тож у цьому розділі буде проведений якісний та порівняльний аналіз методів машинного навчання та нейронних мереж, щоб визначити в комбінації з яким алгоритмом будуть отримані найкращі результати точності та якості розпізнавання.

2.2.1 Decision Tree

Дерево рішень є одним з найчастіше використовуваних підходів, що застосовують до задач для класифікації та прогнозування (регресії). Навчання таким алгоритмом, як Decision Tree алгоритм, зосереджується на створенні класифікації у формі дерев рішень. За допомогою підходу визначення об'єкта через попереднє значення, такий алгоритм зверху вниз оцінює атрибути у вузлах дерева, оцінює гілки на основі окремих атрибутів вузла та отримує висновки з листових вузлів у дереві рішень. Отже, кожен шлях від кореня до кінцевого вузла - це кон'юнктивний вираз, тоді як усе дерево представляє набір диз'юнктивних виразів.

Вхідним елементом алгоритму побудови дерева рішень є набір класичних позначених прикладів. Результатом структури є бінарне або множинне дерево.

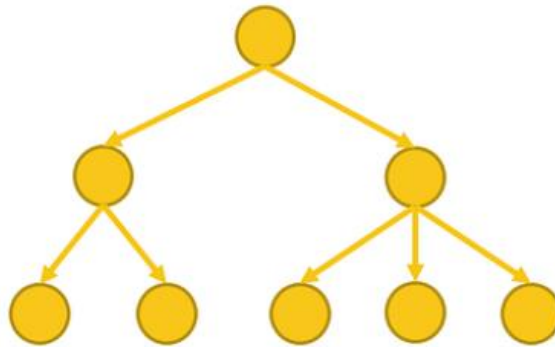


Рисунок 2.5 – Схематичне відображення дерева рішення[8]

У структурі системи рішень, кожен індивідуальний вузол відповідає певній перевірці характеристики, тоді як кожен кінцевий вузол представляє логічне значення. Крім того, кожне розгалуження дерева відображає одне з можливих значень, пов'язаних з властивістю, що перевіряється. Ключовим аспектом побудови системи рішень є вміння обрати найкраще логічне утвердження або характеристику. Навіть для одного і того ж набору прикладів існує багато варіантів. Вивчення показує, що загалом, чим менше дерево, тим краще його передбачувальні здібності. Основним завданням при побудові системи рішень є зменшення її розміру шляхом обрання правильного розгалуження, зумовленого конкретною характеристикою. Вибір характеристик залежить від методу оцінки домішок на різних підмножинах прикладів.

Метод вимірювання включає інформаційне посилення, коефіцієнт посилення, індекс Джині формула (2.1), ентропія формула (2.2) [9], відстань вимірювання, статистика X^2 , вагомість доказів, мінімальна довжина опису тощо. Різні методи вимірювання мають різні наслідки, особливо для атрибутів з багатьма значеннями. Вибір відповідного методу вимірювання суттєво впливає на кінцевий результат.

$$I_G = 1 - \sum_j^c p_j^2 \quad (2.1)$$

$$I_H = -\sum_{j=1}^c p_j \log_2(p_j) \quad (2.2)$$

Об'єктом інтелектуального аналізу даних є реальні дані, які, як правило, не є ідеальними. Вони можуть містити відсутні значення в деяких атрибутах або бути неповними через відсутність важливої інформації. Крім того, дані можуть бути неточними, містити помилки або шум, безпосередньо це стосується і задачі, що вирішується, оскільки аудіозаписи це не найчистіші дані, шуми, вигуки, нетиповий тембр голосу, неякісний пристрій запису звуків – все це має прямий вплив на якість розпізнавання емоції, в той час, як основний алгоритм побудови дерева рішень не приділяє уваги шуму, тому створене дерево повністю враховує навчальні приклади, що може призвести до перенавчання і погіршення прогнозування. Метод обрізки є технікою для управління шумом, одночасно спрощуючи дерево і полегшуючи його інтерпретацію.

Отже, оцінка методу класифікації Decision Tree для розпізнавання емоцій людини представлена нижче:

	precision	recall	f1-score	support
0	0.80	0.79	0.79	2264
1	0.67	0.66	0.67	172
2	0.62	0.59	0.61	1910
3	0.62	0.61	0.61	1878
4	0.62	0.64	0.63	1829
5	0.64	0.65	0.65	1612
6	0.68	0.68	0.68	1820
7	0.65	0.70	0.67	197
micro avg	0.67	0.67	0.67	11682
macro avg	0.66	0.67	0.66	11682
weighted avg	0.67	0.67	0.67	11682
samples avg	0.67	0.67	0.67	11682

Рисунок 2.6 – Оцінка методу класифікації Decision Tree для розпізнавання емоцій людини

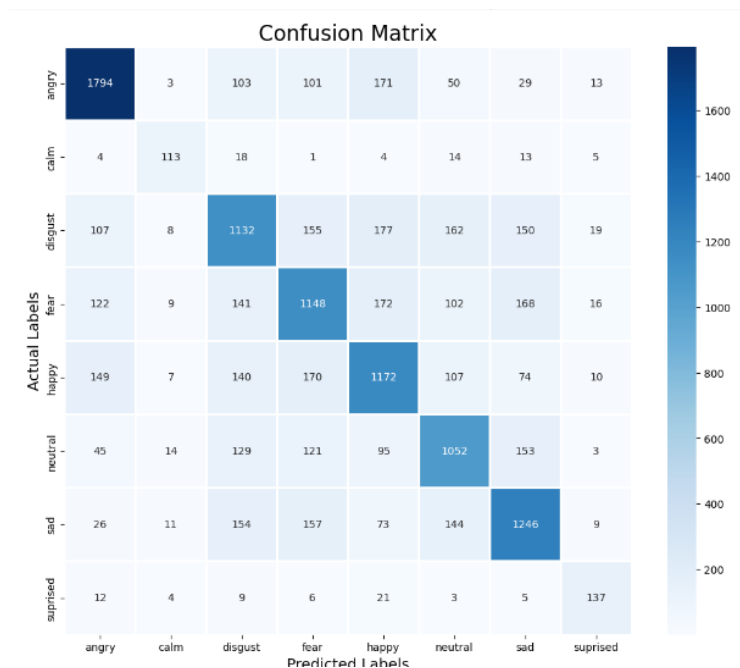


Рисунок 2.7 – Результати розпізнавання з використанням методу Decision Tree

2.2.2 MLP

Багатошаровий перцептрон (MLP) — це контрольований алгоритм навчання, який вивчає функцію навчаючись на наборі даних. [10]

Багатошаровий перцептрон (MLP) належить до типу нейронної мережі, яка складається щонайменше з трьох вузлів. Крім того, кожен вузол у межах MLP, за винятком вхідного вузла, функціонує як нейрон, який використовує нелінійну функцію активації. Вузли багатошарового перцептрона організовані в шари, що підкреслює структурне розташування.

У контексті багатошарового перцептрона (MLP) задіяні різні рівні:

- 1) вхідний рівень: цей рівень отримує початкові вхідні дані;
- 2) рівень виводу: цей рівень виробляє остаточний вихід або прогноз MLP;
- 3) приховані шари: це шари, розташовані між вхідним і вихідним шарами, відповідальні за обробку та перетворення вхідних даних.

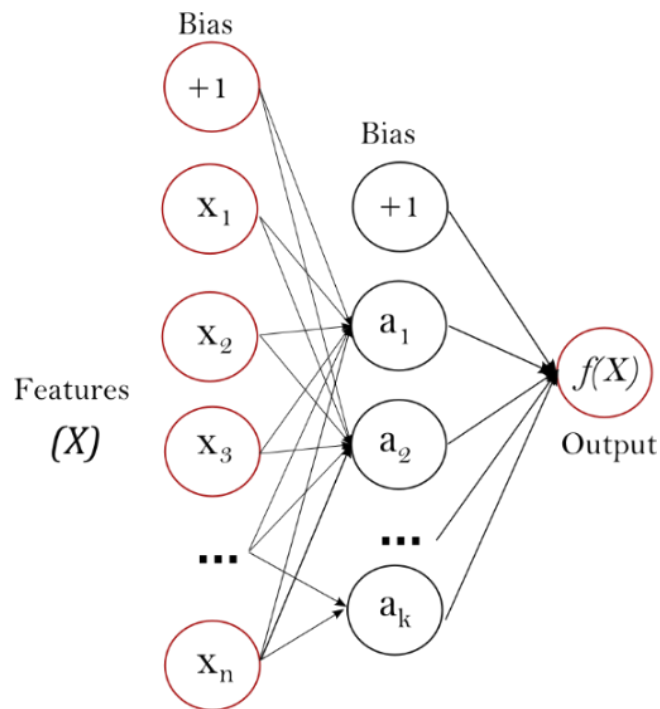


Рисунок 2.8 – Схематичне відображення MLP [10]

Такий підхід відрізняється від логістичної регресії тим, що між вхідним і вихідним шарами може бути один або кілька нелінійних шарів, що називаються прихованими шарами [10]. На рисунку 2.8 показано MLP одного прихованого шару зі скалярним виходом.

У контексті MLP нейрони використовують нелінійні функції активації, призначені для імітації функціональності нейронів людського мозку. І навпаки, у багат шаровому перцептроні кожен нейрон використовує функцію лінійної активації, а процес навчання включає зворотне поширення.

Функція активації відповідає за об'єднання зважених вхідних даних нейрона та включення зміщення для генерації виходу. По суті, він відображає зважені входи на вихід нейрона. Серед різноманітних функцій активації сигмоподібна функція зазвичай використовується для визначення виходу нейрона. Прикладом сигмоїдної функції є логістична функція, формула та графік якої зображено нижче:

$$f(x) = \frac{1}{1 + e^{-\beta x}} \quad (2.3)$$

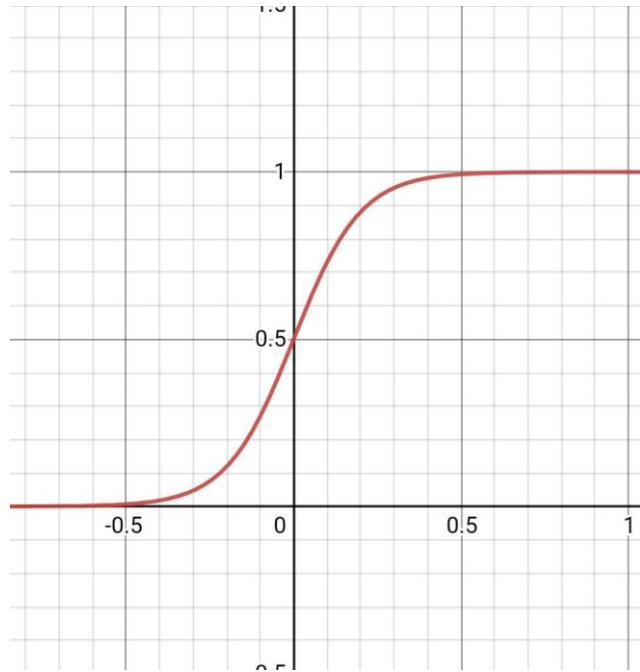


Рисунок 2.9 – Графічне відображення сигмоїдальної функції

Альтернативним прикладом сигмоїдної функції є функція активації гіперболічного тангенса, яка дає результат у діапазоні від -1 до 1.

$$y = th(x) \quad (2.4)$$

Іншим типом функції активації, який можна використовувати, є випрямлена лінійна одиниця (ReLU). ReLU відомий своєю покращеною продуктивністю порівняно з логістичною функцією та функцією гіперболічного тангенса.

Отже, оцінка методу класифікації MPL для розпізнавання емоцій людини представлена нижче.

	precision	recall	f1-score	support
angry	0.81	0.92	0.86	2289
calm	0.83	0.88	0.85	173
disgust	0.66	0.68	0.67	1839
fear	0.83	0.59	0.69	1861
happy	0.76	0.64	0.69	1877
neutral	0.61	0.78	0.68	1570
sad	0.73	0.73	0.73	1889
suprised	0.89	0.90	0.89	184
accuracy			0.74	11682
macro avg	0.76	0.77	0.76	11682
weighted avg	0.74	0.74	0.73	11682

Рисунок 2.10 – оцінка методу класифікації MPL для розпізнавання емоцій людини

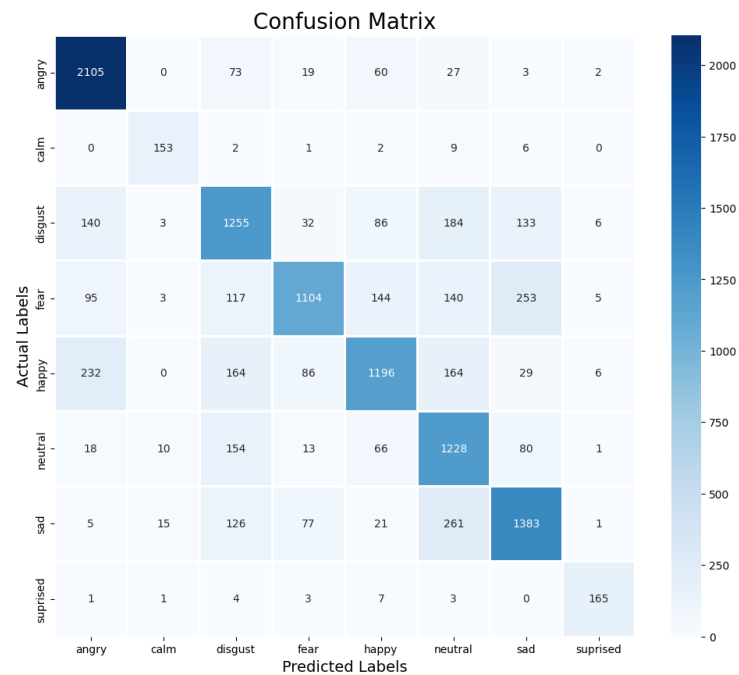


Рисунок 2.11 – Результати розпізнавання з використанням методу MLP

2.2.3 CNN

Мета такої області, як штучний інтелект полягає в тому, щоб надати машинам здатність сприймати світ так, як люди, дозволяючи їм виконувати різні завдання, такі

як розпізнавання зображень і відео, аналіз і класифікація зображень, відтворення медіа, системи рекомендацій і природна мова обробки. З часом було досягнуто значного прогресу в галузі комп'ютерного зору завдяки вдосконаленню спеціального алгоритму, відомого як згорткові нейронні мережі (CNN). CNN складаються з взаємопов'язаних нейронів із регульованими вагами та зміщеннями. В архітектурі CNN нейрони організовані в рівні, включаючи вхідний рівень, кілька прихованих шарів і вихідний рівень. Якщо мережа містить значну кількість прихованих шарів, її зазвичай називають глибокою нейронною мережею. На відміну від повністю зв'язаних мереж, таких як мережі багаторівневого перцептрона (MLP), де нейрони підключаються до всіх входів, нейрони в прихованих шарах CNN підключаються лише до невеликої області вхідного простору, отриманого з попереднього рівня. Цей підхід зменшує кількість ваг (параметрів) з'єднання в CNN порівняно з MLP, що призводить до швидшого часу навчання для мереж порівнянного розміру.

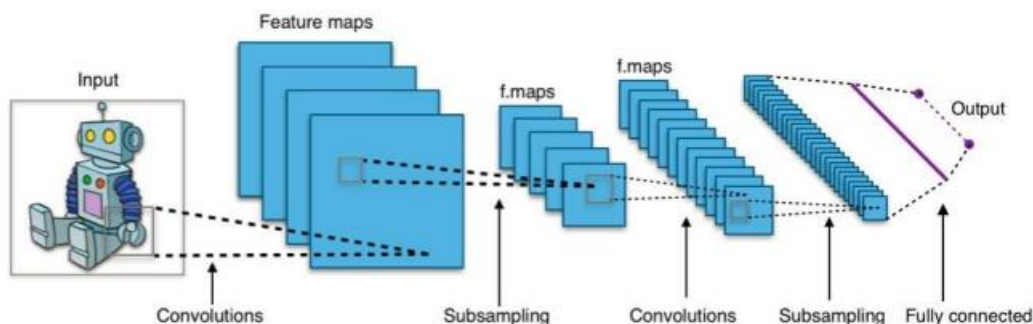


Рисунок 2.12 – Схематичне відображення [15]

Згорткові нейронні мережі (CNN) – це тип штучної мережі прямого зв'язку, натхненна зоровою корою тварин. CNN є формою регуляризованих багатошарових перцептронів, де регуляризовані версії використовуються для пом'якшення надмірного оснащення. На відміну від повністю зв'язаних мереж, де кожен нейрон одного шару з'єднаний з усіма нейронами наступного шару, CNN використовують інший підхід до регуляризації. Замість того, щоб штрафувати параметри чи

зменшувати зв'язок, CNN використовують ієрархічні шаблони, присутні в даних, і створюють складні шаблони, комбінуючи менші та простіші шаблони, вбудовані у їхні фільтри.

Однією з ключових переваг CNN є їхня мінімальна залежність від попередньої обробки порівняно з іншими алгоритмами класифікації зображень. У CNN мережа вчиться оптимізувати фільтри або ядра за допомогою автоматичного навчання, тоді як традиційні алгоритми вимагають ручної розробки цих фільтрів. Ця свобода від попередніх знань і людського втручання у вилучення ознак є значною перевагою.

	precision	recall	f1-score	support
angry	0.89	0.87	0.88	2244
calm	0.69	0.90	0.78	193
disgust	0.67	0.76	0.71	1864
fear	0.81	0.62	0.70	1869
happy	0.73	0.74	0.74	1862
neutral	0.70	0.72	0.71	1598
sad	0.70	0.75	0.73	1864
surprised	0.84	0.84	0.84	188
accuracy			0.75	11682
macro avg	0.75	0.77	0.76	11682
weighted avg	0.76	0.75	0.75	11682

Рисунок 2.13 – Результати розпізнавання з використанням методу CNN

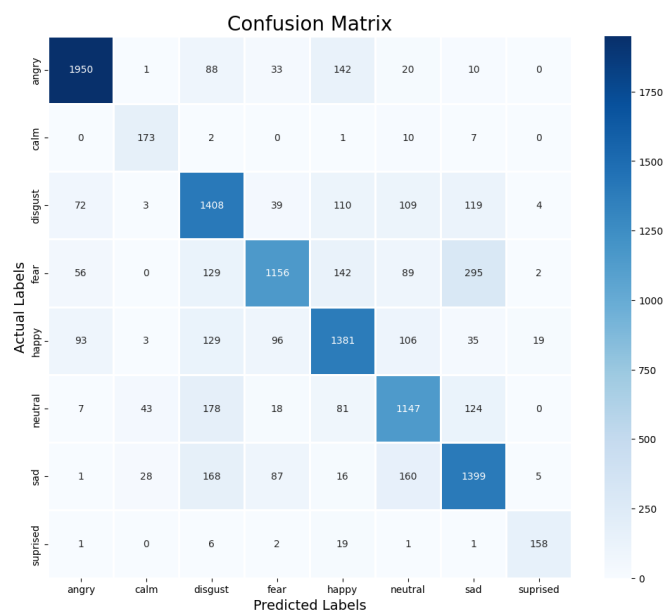


Рисунок 2.14 – Результати розпізнавання з використанням методу CNN

2.3 Огляд методів для оптимізації параметрів

Класифікація гіперпараметри відносяться до значень параметрів, які контролюють процес навчання та значно впливають на продуктивність моделей машинного навчання.

Оптимізація гіперпараметрів включає завдання пошуку оптимальної комбінації значень гіперпараметрів, яка дає найвищу продуктивність заданих даних протягом розумного періоду часу.

Цей процес оптимізації відіграє вирішальну роль у підвищенні точності прогнозування алгоритмів машинного навчання, що робить його складним аспектом розробки моделі.

Хоча багато алгоритмів машинного навчання мають значення гіперпараметрів за замовчуванням, ці значення за замовчуванням не завжди можуть забезпечити оптимальну продуктивність для різних типів проект. Отже, оптимізація стає важливою для визначення найкращої комбінації гіперпараметрів для досягнення чудових результатів.

Алгоритми машинного навчання зазвичай перетворюють задану проблему на задачу оптимізації та використовують різні методи оптимізації для її вирішення. Функція оптимізації містить кілька гіперпараметрів, які попередньо визначені перед процесом навчання та впливають на те, як алгоритм пристосовує модель до даних. На відміну від параметрів внутрішньої моделі (наприклад, вагових коефіцієнтів нейронної мережі), які можна дізнатися під час навчання, гіперпараметри вимагають ретельного вибору перед навчанням. Мета полягає в тому, щоб визначити оптимальний набір значень гіперпараметрів, які забезпечують найкращу продуктивність даних протягом розумного періоду часу. Цей процес, відомий як оптимізація або налаштування гіперпараметрів, значно впливає на точність передбачення алгоритмів машинного навчання.

Останнім часом складність і виразність моделей машинного навчання значно зросла, що супроводжується значними обчислювальними витратами та зростаючою кількістю параметрів налаштування. Ці параметри налаштування, відомі як «гіперпараметри», відіграють вирішальну роль у визначенні того, наскільки добре алгоритм працює з невидимими даними. Вони охоплюють такі фактори, як архітектура моделі, регуляризація та швидкість навчання. Досягнення високоякісної прогнозової моделі значною мірою залежить від вибору відповідних конфігурацій гіперпараметрів.

На жаль, зв'язок між продуктивністю алгоритму та гіперпараметрами часто незрозумілий. На практиці необхідно ітеративно коригувати гіперпараметри, навчати кілька моделей з різними комбінаціями значень і порівнювати їх продуктивність, щоб вибрати найкращу модель. Отже, оптимізація гіперпараметрів стає критичною проблемою в алгоритмах машинного навчання.

Існує в основному два типи методів оптимізації гіперпараметрів, тобто ручний пошук і автоматичний пошук. Ручний пошук перевіряє набори гіперпараметрів вручну. Це залежить від фундаментальної інтуїції та досвіду досвідчених користувачів, які можуть визначити важливі параметри, що мають більший вплив на результати, а потім визначити зв'язок між певними параметрами та кінцевими результатами за допомогою інструментів візуалізації. Ручний пошук вимагає від користувачів більших професійних знань і практичного досвіду. І його важко застосовувати користувачам, які не мають досвіду. Процес налаштування гіперпараметрів нелегко відтворити. Крім того, зі збільшенням кількості гіперпараметрів і діапазону значень стає досить важко керувати, оскільки люди погано вміють працювати з даними великої розмірності та легко неправильно інтерпретують або пропускають тенденції та зв'язки в гіперпараметрах.

Автоматизація оптимізації гіперпараметрів повністю звільняє користувача від необхідності ручного налаштування, що становить значний прогрес. Це свідчить про важливість оптимізації гіперпараметрів у сфері машинного навчання. Існує безліч

різних методів для оптимізації гіперпараметрів моделі. Тому надалі буде проведено порівняльний аналіз різних методів вибору оптимальних значень.

2.3.1 Random Search

Оптимальний підхід для пошуку найкращого набору гіперпараметрів у моделі машинного навчання передбачає використання випадкового пошуку. Метаоцінка рандомізованого пошуку — це алгоритм, який навчає та оцінює кілька моделей шляхом випадкового вибору комбінацій гіперпараметрів із попередньо визначених розподілів. Навчаючи N версій моделі з різними випадковими комбінаціями гіперпараметрів, алгоритм визначає найбільш вдалий варіант моделі, що призводить до майже оптимального набору гіперпараметрів. Важливо зазначити, що випадковий пошук називають метаоцінкою, оскільки це не сама оцінка, а скоріше алгоритм, застосований до існуючої оцінки для налаштування її гіперпараметрів.

Алгоритм випадкового пошуку використовує випадкові комбінації гіперпараметрів для знаходження найкращого рішення для побудованої моделі. У цій техніці модель налаштовується з кожною випадковою комбінацією гіперпараметрів, шляхом вибору випадкових значень для кожної комбінації. Модель, яка демонструє найвищу точність, визначається як найкраща модель, а відповідні гіперпараметри вважаються оптимальними.

Порівняно з пошуком по сітці, випадковий пошук має перевагу, оскільки досліджує розподіли гіперпараметрів, а не попередньо визначені значення кандидатів для кожного гіперпараметра. Ця гнучкість дає змогу виражати переваги щодо очікуваного оптимального значення для гіперпараметра.

Щоб отримати більш надійну оцінку ефективності кожної моделі поза вибіркою, випадковий пошук часто використовується в поєднанні з перехресною перевіркою.

Нижче продемонстровано найкращі результати, яких вдалось досягти використовуючи алгоритм Random Search

	precision	recall	f1-score	support
angry	0.89	0.87	0.88	2195
calm	0.74	0.88	0.80	210
disgust	0.77	0.68	0.72	1908
fear	0.78	0.68	0.73	1851
happy	0.74	0.75	0.75	1869
neutral	0.71	0.76	0.73	1620
sad	0.70	0.82	0.76	1837
suprised	0.79	0.88	0.83	192
accuracy			0.77	11682
macro avg	0.77	0.79	0.77	11682
weighted avg	0.77	0.77	0.77	11682

Рисунок 2.15 – Результати розпізнавання з використанням методу CNN та Random Search

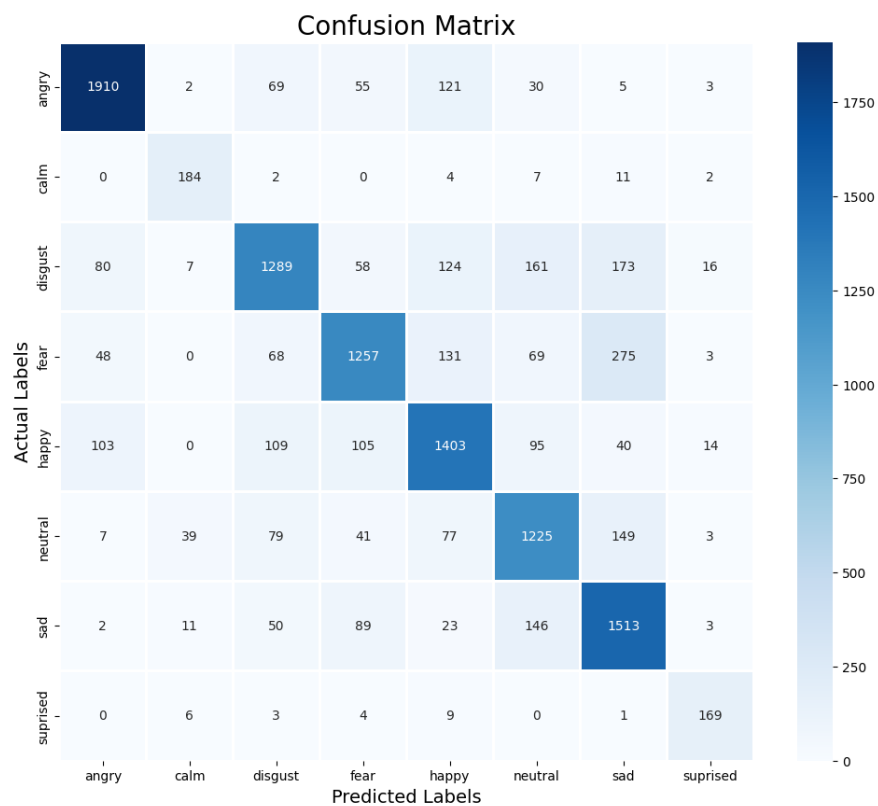


Рисунок 2.15 – Результати розпізнавання з використанням методу CNN та Random Search

2.3.2 BO

Налаштування гіперпараметрів створює серйозну проблему для машинного навчання, але байєсівська оптимізація стала ефективною основою для вирішення цієї проблеми. Такий алгоритм кращий за звичайні методи, такі як пошук у сітці та випадковий пошук, забезпечуючи надійні рішення для оптимізації дорогих функцій чорної скриньки. Байєсова оптимізація використовує непараметричний гаусівський процес формула (2.5) як імовірнісну міру для моделювання невідомої функції, одночасно використовуючи функцію отримання як сурогатну функцію корисності, щоб керувати пошуком наступного спостереження.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.5)$$

Попри ці досягнення, налаштування гіперпараметрів залишається складним, особливо при роботі з великими наборами даних.

Рання спроба покращити ефективність байєсівської оптимізації включала багатозадачну байєсівську оптимізацію, яка оптимізувала гіперпараметри для всього набору даних, розглядаючи допоміжну задачу оптимізації гіперпараметрів для невеликої підмножини даних. Цей підхід оцінював зв'язок між завданнями та намагався знайти найкращі параметри для цих завдань одночасно в одній структурі. Однак цей підхід зіткнувся з двома проблемами. По-перше, це виявилось неефективним, оскільки як недорогі, так і дорогі процеси були виявлені одночасно, що призвело до поганої ранньої оцінки їх взаємозв'язку та затримки конвергенції. По-друге, розробка коваріаційної функції між завданнями для використання зв'язку між гіперпараметрами в різних завданнях була непростою.

Нижче продемонстровано найкращі результати, яких вдалось досягти використовуючи алгоритм BO

	precision	recall	f1-score	support
angry	0.91	0.89	0.90	2312
calm	0.82	0.93	0.87	167
disgust	0.79	0.74	0.77	1839
fear	0.85	0.69	0.76	1895
happy	0.82	0.76	0.79	1857
neutral	0.70	0.83	0.76	1627
sad	0.73	0.85	0.79	1801
suprised	0.89	0.86	0.88	184
accuracy			0.80	11682
macro avg	0.81	0.82	0.81	11682
weighted avg	0.81	0.80	0.80	11682

Рисунок 2.15 – Результати розпізнавання з використанням методу CNN та ВО

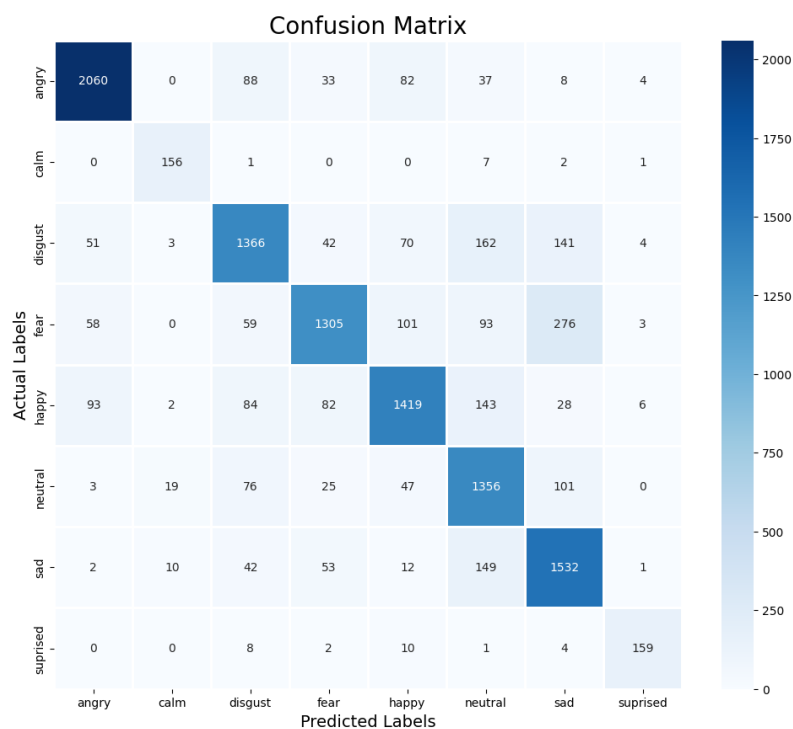


Рисунок 2.16 – Результати розпізнавання з використанням методу CNN та ВО

2.3.3 Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization

Однак взаємодія між цими гіперпараметрами та їхнім впливом на якість моделі залишається недостатньо вивченим. У результаті практики часто вдаються до

підходів грубої сили, таких як випадковий пошук і пошук у сітці, щоб вирішити цю проблему.

Щоб підвищити ефективність процесу пошуку, в останніх досягненнях оптимізації гіперпараметрів домінували байєсівські методи оптимізації. Ці методи спеціально зосереджені на оптимізації вибору конфігурацій гіперпараметрів, спрямовані на покращення загальної продуктивності моделей машинного навчання.

На відміну від цього, пропонуємо інший підхід до оптимізації гіперпараметрів, який наголошує на прискоренні оцінки конфігурації. Ці підходи динамічно виділяють більше ресурсів для перспективних конфігурацій гіперпараметрів. Така адаптивність обчислень дозволяє швидше оцінювати різні конфігурації, що потенційно призводить до значної економії часу.

Основна увага Hyperband зосереджена на тому, як розподілити ресурси між конфігураціями гіперпараметрів із випадковою вибіркою. Застосовуючи принципову стратегію ранньої зупинки, Hyperband може оцінити значно більшу кількість конфігурацій порівняно з процедурами чорної скриньки, такими як байєсівська оптимізація, що відрізняє Hyperband від інших, так це його універсальність, оскільки він робить мінімальні припущення та може застосовуватися до різних сценаріїв оптимізації [11].

Нижче продемонстровано найкращі результати, яких вдалось досягти використовуючи алгоритм байєсівської оптимізація

	precision	recall	f1-score	support
angry	0.89	0.93	0.91	2232
calm	0.88	0.92	0.90	193
disgust	0.81	0.78	0.79	1934
fear	0.88	0.73	0.80	1803
happy	0.84	0.80	0.82	1862
neutral	0.79	0.80	0.79	1570
sad	0.75	0.88	0.81	1905
suprised	0.90	0.90	0.90	183
accuracy			0.83	11682
macro avg	0.84	0.84	0.84	11682
weighted avg	0.83	0.83	0.83	11682

Рисунок 2.16 – Результати розпізнавання з використанням методу CNN та Hyperband

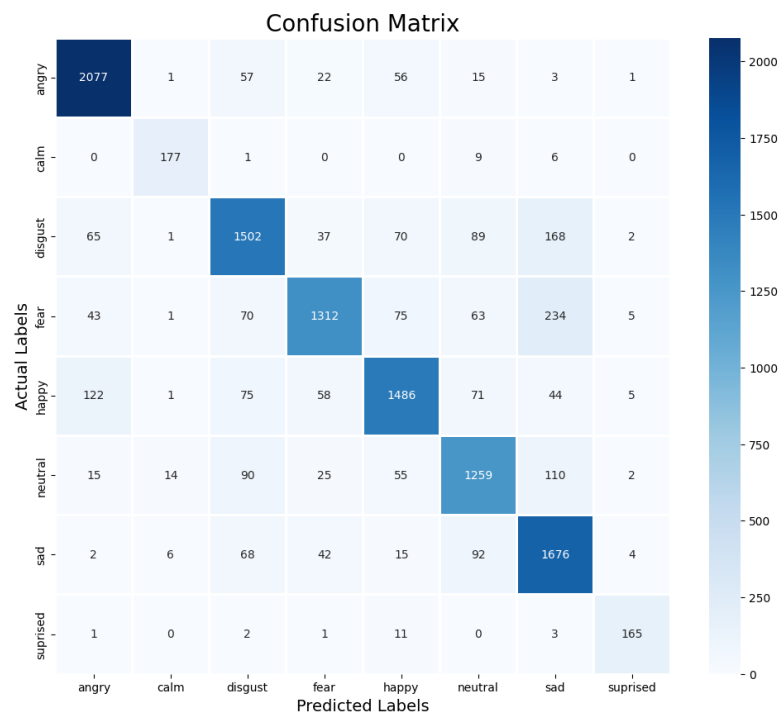


Рисунок 2.16 – Результати розпізнавання з використанням методу CNN та Hyperband

Hyperband використовує стратегію розподілу ресурсів для розподілу ресурсів, таких як час обчислення або кількість епох, для набору конфігурацій з різними гіперпараметрами. Він тренує кожну конфігурацію протягом кількох епох і усуває погані результати за допомогою критеріїв ранньої зупинки. Цей процес повторюється ітеративно зі збільшенням розподілу ресурсів, що дозволяє довше навчатися кращим конфігураціям і таким чином підвищити ефективність оптимізації. Ключовою інновацією Hyperband є його здатність знаходити найефективнішу конфігурацію за певний бюджет часу, використовуючи при цьому значно менше ресурсів порівняно з іншими методами, такими як пошук по сітці або байєсівська оптимізація. Це робить його особливо корисним у ситуаціях, коли обчислювальні ресурси обмежені, наприклад у хмарних середовищах або на периферійних пристроях.

2.4 Висновки до розділу

У відповідному розділі проведено порівняння підходів в розв'язанні задачі класифікації емоцій за аудіозаписами. Аналіз був здійснений в декілька етапів:

- 1) Аналіз методів перетворення аудіозаписів;
- 2) Аналіз методів машинного навчання для класифікації;
- 3) Аналіз методі регуляризації алгоритмів машинного навчання.

З метою досягнення високої точності та продуктивності був обраний метод нейромережі. Проте, цю модель також можна вдосконалити шляхом використання методу оптимізації гіперпараметрів для максимізації класифікаційної точності.

Комбінація такого набору підходів надає найкращі результати якості та точності класифікації разом з продуктивністю.

Таблиця 2.1 – Порівняльний аналіз методів машинного навчання

Метрики	DST	MLP	CNN
Presicion	67%	76%	75%
Recall	67%	77%	77%
F1-score	67%	76%	76%

Таблиця 2.2 – Порівняльний аналіз методів оптимізації гіперпараметрів

Метрики	RS	BO	Hyperband
Presicion	77%	81%	84%
Recall	79%	82%	84%
F1-score	77%	81%	84%

В таблиці 2.1 представлений порівняльний аналіз описаних методів машинного навчання , в таблиці 2.2 представлений порівняльний аналіз описаних методів

регуляризації методів машинного навчання та обґрунтування вибору вищезазначених.

В результат проведеного аналізу, як підсумок можна сказати, що найкращим та актуальними підходами в вирішенні задачі є реалізація системи визначення емоцій за аудіозаписами, що ґрунтується на таких математичних методах:

- 1) MFCC;
- 2) CNN;
- 3) Hyperband.

3 МАТЕМАТИЧНЕ ЗАБЕЗПЕЧЕННЯ

В третьому розділі було проведено аналіз існуючих алгоритмів та підходів в вирішенні задачі класифікації емоцій за аудіозаписами, також були обґрунтовано обрані методи, що будуть реалізовані в системі для вирішення поставленої задачі.

Серед обраних алгоритмів :

- 1) MFCC для перетворення аудіосигналів;
- 2) CNN для здійснення класифікації;
- 3) Hyperband для оптимізації і покращення показників класифікації.

Тож, тепер в даному розділі будуть розглянуті вищезазначені алгоритми більш детально, а саме математичне підґрунтя на якому базуються вони, їх особливості та математичний опис запропонованих підходів.

Також важливим аспектом в задачі класифікації є вірна оцінка якості та точності прогнозування. Тож далі буде продемонстровано та обґрунтовано вибір способів оцінювання результату розпізнавання емоцій, а також наведено математичне представлення та приклади реалізації метрик.

3.1 Метрики якості

Метрики точності – це інструмент та оцінка ефективності для кращого розуміння якості роботи моделі.

Аналізуючи такі показники ефективності, ми можемо отримати уявлення про якість моделі класифікації та прийняти обґрунтовані рішення щодо потенційних покращень або коригувань.

Нижче розглянуті метрики використані у конкретній роботі:

- 1) **Confusion matrix** - це інструмент оцінки ефективності, який зазвичай використовується в задачах машинного навчання та класифікації. Він надає детальну

розбивку прогнозованих і фактичних міток класів у проблемі класифікації. Матриця структурована як таблиця з рядками та стовпцями, що представляють передбачені та фактичні мітки класу відповідно. Основна мета матриці плутанини — показати кількість створених моделлю класифікації показників:

- а) TP - істинно позитивних, тобто модель правильно передбачає позитивні випадки;
- б) TN - істинно негативних, тобто модель правильно передбачає негативні випадки;
- в) FP - хибнопозитивних, тобто модель неправильно передбачає негативні випадки;
- г) FN - хибнонегативних, тобто модель неправильно передбачає негативні випадки.

Матриця плутанини дозволяє зрозуміти продуктивність класифікатора шляхом візуалізації розподілу прогнозованих і фактичних міток. Такий підхід дає уявлення про сильні та слабкі сторони моделі, такі як її здатність розрізняти різні класи та виникнення помилкових класифікацій;

2) Cross-validation - перехресна перевірка є широко використовуваною технікою в машинному навчанні для оцінки ефективності моделі та її здатності до узагальнення. Такий підхід передбачає поділ доступних даних на кілька підмножин або згорток, де кожна згортка використовується як набір для навчання та набір для перевірки.

Основна ідея перехресної перевірки полягає в тому, щоб змодельовати, як модель працюватиме на невидимих даних, навчаючи її на частині доступних даних і оцінюючи її продуктивність на частині, що залишилася. Це допомагає оцінити, наскільки добре модель узагальнюватиме нові, невідомі дані.

Найпоширеніша форма перехресної перевірки називається k-кратною перехресною перевіркою. У k-кратній перехресній перевірці дані поділяються на k підмножин або згорток приблизно однакового розміру. Модель тренується на k-1 згортках і оцінюється на згортках, що залишилися і цей процес повторюється k разів,

причому кожна складка служить набором перевірки рівно один раз [12]. Потім результати продуктивності кожної ітерації усереднюються для отримання більш надійної оцінки продуктивності моделі.

Деякі ключові переваги перехресної перевірки включають:

Надійна оцінка продуктивності. Оцінюючи модель на кількох підмножинах даних, перехресна перевірка забезпечує більш надійну оцінку продуктивності моделі порівняно з одноразовим тестуванням. Це допомагає пом'якшити вплив мінливості даних і забезпечує більш репрезентативну оцінку.

Ефективне використання даних: перехресна перевірка дозволяє нам краще використовувати наявні дані, використовуючи їх як для навчання, так і для перевірки. Це особливо корисно, коли набір даних обмежений, оскільки допомагає отримати максимум інформації з доступних зразків;

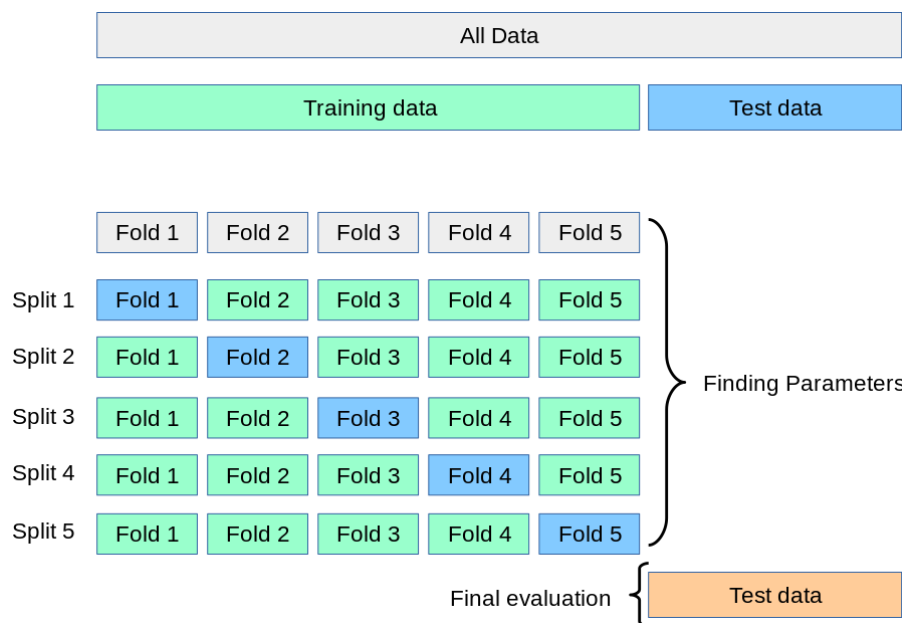


Рисунок 3.1 – Результати розпізнавання з використанням методу CNN та Hyperband [12]

3) Precision

Precision — це показник продуктивності, який використовується в машинному навчанні та пошуку інформації для оцінки точності моделі класифікації, зокрема в

завданнях класифікації. Він вимірює частку правильно передбачених позитивних випадків серед усіх випадків, передбачених моделлю як позитивні.

Точність обчислюється як відношення істинно позитивних (TP) прогнозів до суми істинно позитивних і помилково позитивних (FP):

$$Precision = \frac{TP}{(TP + FP)} \quad (3.1)$$

Простими словами, точність відповідає на запитання: «Скільки з усіх випадків, передбачених моделлю як позитивних, було насправді позитивними?»

Однак важливо розглядати точність у поєднанні з іншими показниками, такими як запам'ятовування, оскільки оптимізація для високої точності може призвести до низького рівня запам'ятовування (висока кількість помилкових від'ємне). Знаходження правильного балансу між Precision та Recall залежить від конкретних вимог проблеми, що розглядається.

4) Recall

Recall, також відоме як чутливість або справжній позитивний показник, є показником продуктивності, який використовується в машинному навчанні та пошуку інформації для оцінки ефективності моделі класифікації, зокрема в завданнях двійкової класифікації. Він вимірює частку правильно передбачених позитивних випадків серед усіх фактичних позитивних випадків у наборі даних.

Recall розраховується як відношення істинно позитивних (TP) прогнозів до суми істинно позитивних і помилково негативних (FN):

$$Recall = \frac{TP}{(TP + FN)} \quad (3.2)$$

Простіше кажучи, Recall відповідає на запитання: «Скільки з усіх фактичних позитивних випадків у наборі даних модель правильно визначила?»

5) F1-score

F1-оцінка — це популярний показник ефективності, який використовується в машинному навчанні та пошуку інформації для оцінки загальної точності моделі класифікації, зокрема в завданнях бінарної класифікації. Він об'єднує показники точності та запам'ятовування в одне значення, забезпечуючи збалансований показник ефективності моделі. Показник F1 розраховується як середнє гармонійної точності та запам'ятовування:

$$F1_{score} = 2 \frac{(Precision * Recall)}{(Precision + Recall)} \quad (3.3)$$

Точність представляє частку правильно передбачених позитивних випадків серед усіх випадків, передбачених як позитивні, тоді як відкликання вимірює частку правильно передбачених позитивних випадків серед усіх фактичних позитивних випадків. Поєднуючи ці два показники, оцінка F1 враховує як здатність моделі мінімізувати хибні спрацьовування (точність), так і її здатність мінімізувати хибні негативи (відкликання).

Показник F1 коливається від 0 до 1, причому значення 1 вказує на ідеальну точність і запам'ятовування, а значення 0 вказує на низьку продуктивність. Вищий показник F1 свідчить про те, що модель має хороший баланс між точністю та запам'ятовуванням, тобто вона може точно класифікувати позитивні випадки, мінімізуючи як помилкові спрацьовування, так і помилкові негативи.

Оцінка F1 особливо корисна, коли набір даних незбалансований, тобто кількість позитивних і негативних випадків значно відрізняється. У таких випадках лише точність може не забезпечити точного представлення продуктивності моделі, оскільки в ній може домінувати більшість класів. Оцінка F1 однаково враховує обидва класи та забезпечує більш надійну оцінку загальної точності моделі.

3.2 MFCC

Техніка, яка найчастіше використовується для виділення характеристик аудіосигналу - MFCC. Виділення ознак є дуже важливою частиною в аналізі та пошуку зв'язків, оскільки моделі нейронних мереж можуть бути навчені лише за даними в зрозумілому для них форматі. Мовні сигнали використовуються, як джерело інформації для виділення ознак і формування набору даних. Попередньо необхідно провести деякі перетворення та аналіз даних перш ніж використовувати головну функцію. Тож, спершу потрібне посилення амплітуди високочастотних сигналів з врахування природного ослаблення високих частот у людському мовленні. Вхідний сигнал фільтрується за допомогою фільтра високих частот першого порядку, що призводить до функції відгуку описаній в формулі[13]:

$$H(z) = 1 - \alpha(z^{-1}) \quad (3.4)$$

Існує припущення, що емоція в аудіозаписах залишається однаковою продовж дуже коротких проміжків часу і рахується в мілісекундах, тож аудіозаписи будуть поділені на проміжки часу 30 мс, тож сигнал розділяється на короткі відрізки.

Далі використовується функція Хеммінга задана формулою (3.5)[13]:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (3.5)$$

де $0 \leq n \leq N-1$, N – довжина вікна.

Далі сигнал $s(n)$ виділяється шляхом множення значення сигналу в момент часу n , з результатом функції Хеммінга в проміжок часу n таке рішення відображено в формулі нижче[13]:

$$y(n) = s(n) * w(n) \quad (3.6)$$

Далі виконується перетворення Фур'є, що застосовується для кожного відрізка спектру.

Наступним кроком є застосування фільтрів формулою (3.7)[13]:

$$Mel(f) = 1127 \ln \left(1 + \frac{f}{700} \right) \quad (3.7)$$

Такий підхід створює імітацію нелінійного сприйняття звуку людським вухом, будучи більш розбірливим на нижчих частотах і меншим на високих частотах. Люди не чують звуку в лінійному масштабі, тому наступним кроком є обчислення логарифма спектру величини. Мовний сигнал не є постійним на кожному відрізку. Враховуються особливості, пов'язані зі зміною часу, різницю можна отримати за формулою (3.8)[13]:

$$\delta(t) = c(t + 1) - c(t) \quad (3.8)$$

Такі отримані характеристики джерела голосу відображають інформацію про мовця. Виділяють висоту тону та форманти для порівняння характеристик джерела голосу з короткими спектральними та часовими характеристиками. Висота визначається основними частотами вібрації голосових складок і форманта визначається резонансними частотами трубки голосового тракту.

3.3 CNN

Моделі глибокого навчання, такі як згорткові нейронні мережі (CNN), показали вражаючі результати в багатьох областях, таких як розпізнавання облич, рукописного

тексту, мови та природної мови. Назва «згортка» пов'язана з математичною операцією, яка є важливою для функціонування цих мереж. CNN складаються з трьох основних блоків: згорткового рівня, рівня зведення та повністю зв'язаного рівня. CNN використовують операції згортки замість простого множення для обробки вхідних даних у згорткових шарах. Цей унікальний підхід дозволяє використовувати локальні рецептивні поля, тобто нейрони в цих шарах реагують на стимули в певних місцях і структурах. Цей дизайн натхненний поведінкою нейронів зорової кори. Реалізуючи згортку, CNN вводять розріджені зв'язки та спільне використання параметрів, що призводить до значного зменшення кількості параметрів, необхідних для глибоких нейронних мереж[15].

Застосування згорткових шарів у CNN полягає у виконанні локальної фільтрації, що дозволяє визначати та виділяти різноманітні характеристики, що є важливими. Цей процес генерує декілька функцій ознак, які потім використовуються на глибшому рівні мережі для створення більш високорівневих представлень вхідних даних. За допомогою витягнутих функцій CNN можуть поступово засвоювати більш складні та абстрактні представлення вхідних даних.

Другим важливим компонентом CNN є пулінг. Основна функція пулінгу полягає в зменшенні чутливості виходів мережі до локальних змін у вхідних даних, при несуттєвій значимості просторових характеристик, пулінг покращує продуктивність CNN, ефективно виділяючи важливі характеристики, крім того, пулінг допомагає зменшити перенавчання шляхом скорочення розмірів та параметрів мережі[14]. У CNN рівні об'єднання використовуються для зменшення розмірності даних, шляхом зведення підвбірок із виходу попереднього шару. Для зведення підвбірок використовується функція агрегації, яка може бути максимумом, середнім значенням, L2-нормою або середньозваженим значенням. Ця функція обчислює суму значень нейронів у кожному з об'єднаних ядер. Розмір об'єднаних ядер, крок зсуву та кількість заповнення є важливими параметрами, які можна налаштувати для досягнення кращої продуктивності CNN. Зазвичай, у повністю підключеному рівні нейронної мережі використовується одиниця softmax як функція активації на виході.

Ця функція використовується для визначення ймовірностного розподілу відносно к класів для вхідного екземпляру, і забезпечує нормалізацію цих значень, тобто сума всіх значень ймовірностей дорівнює 1. Функція softmax є розширенням сигмоїдної функції на випадок багатокласової класифікації, обидві функції використовуються для прогнозування ймовірності належності вхідного зразка до певного класу [15].

$$\text{softmax}(\bar{z})_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (3.9)$$

де $\bar{z}$ — вектор вихідного сигналу.

В штучних нейронних мережах (ШНМ) функції активації грають важливу роль, додавши нелінійність, що дозволяє ШНМ вивчати складні моделі. У цьому контексті функції ReLU стали популярним вибором, оскільки вони забезпечують баланс між нелінійністю і лінійністю, яка полегшує навчання на основі градієнта. У згорткових шарах CNN функції ReLU використовуються перед підвибіркою в шарах об'єднання. Однак, недоліком функцій ReLU є їх неефективність під час навчання, коли нейрони видають нуль для від'ємних значень.

3.4 Оптимізація

Гіперпараметрами у машинному навчанні називають параметри алгоритмів, значення яких встановлюються перед запуском процесу навчання. Гіперпараметри використовуються для управління процесом навчання. Вони надають можливість використовувати експериментальний підхід, щоб знайти найкращу конфігурацію для «структури» моделі [16].

Нуретband — це модифікована форма випадкового пошуку, яка включає теорію дослідження й використання для визначення оптимального розподілу часу для кожної

конфігурації. Основна ідея полягає в тому, що алгоритм виконує багато коротких експериментів, щоб швидко знайти проміжні результати, і використовує ці результати, щоб краще сконцентруватись на областях простору гіперпараметрів, які показують найкращі результати. Алгоритм передбачає випадковий вибір набору гіперпараметрів на першому етапі та оцінку їх втрати підтвердження після певної кількості ітерацій, комбінації відкидаються, залишаючи лише половину початкового набору і тоді конфігурації, що залишилися, запускаються на додаткових ітераціях, і процес повторюється, половина з них відкидається і процес повторюється, поки не залишиться лише одна модель.

Для ефективного використання цього алгоритму необхідно визначити гіперпараметр, однак, є можливість скористатися компромісом " n проти V/n ", який використовує послідовний підхід зменшення вдвічі. У цьому контексті, значення V відповідає загальній кількості можливих ітерацій для пошуку оптимальних параметрів, а значення n - кількості комбінацій гіперпараметрів, що будуть досліджені. Відношення V до n є середнім значенням.

Оптимальний вибір залежить від проблеми. Якщо конфігурації гіперпараметрів легко розрізнити, наприклад, якщо набір даних швидко сходиться або погані конфігурації легко ідентифікувати, доцільно вибрати велике значення для n , якщо ж вони повільно диференціюються, або якщо простір пошуку невеликий, але найкраща конфігурація шукається з високою впевненістю, краще вибрати велике значення для V/n [17]. Таким чином, компроміс " n проти V/n " може допомогти знайти найкращу комбінацію параметрів відповідно до конкретної задачі.

Однак обидві ці стратегії мають свої недоліки. Якщо n велике, хороші конфігурації, для зближення яких може знадобитися більше часу, можуть бути відкинуті занадто рано. З іншого боку, якщо V/n великий, то будуть затрачені ресурси на розрахунки комбінацій, які можна було б відкинути раніше.

Nhyperband вирішує цю проблему, розглядаючи кілька можливих значень n для V , по суті, виконуючи пошук у сітці за можливими значеннями n . У довгостроковій перспективі це дозволяє Nhyperband знайти оптимальну стратегію розподілу.

З практичної точки зору, використання Hyperband дозволяє контролювати свою модель під час навчання, щоб її можна було зупинити та відновити за потреби.

3.5 Висновки до розділу

В розділі було проведено аналіз математичних інструментів для вирішення поставленої задачі, описано методологію застосування запропонованих математичних алгоритмів та наведено метрики для оцінки ефективності розробленої моделі класифікації. Було досліджено особливості реалізації згорткової нейронної мережі та вплив гіперпараметрів на її продуктивність. Дослідження показало, що правильний вибір гіперпараметрів є ключовим для досягнення оптимальних результатів. Однак, через велику кількість параметрів, ручне налаштування гіперпараметрів є затратною та складною задачею. Щоб уникнути цих проблем, запропоновано автоматизувати процес з використанням методу Hyperband для оптимізації гіперпараметрів.

4 ОПИС СИСТЕМИ

4.1 Складові системи

Розроблювальна система окрім головної задачі, такої як розпізнавання емоцій за аудіозаписами повинна бути доступна для користувачів без спеціальних знань, а саме мати зручний та зрозумілий інтерфейс системи, доступну навігацію.

Архітектурний дизайн системи повинен відповідати призначенню та необхідним характеристикам системи, забезпечуючи ефективне розпізнавання емоцій за аудіозаписами. Модель автоматизованої системи розпізнавання емоцій за аудіозаписами надає можливість формалізувати процеси аналізу великого обсягу даних, встановлювати взаємодію між компонентами, визначати правила реалізації компонентів та моделювати як статичні, так і динамічні аспекти функціонування системи. Також в процесі реалізації системи розпізнавання емоцій за аудіозаписами потрібно організувати можливість завантаження файлів з записом голосу, а також реалізувати можливість залишати записи та результати аналізу вхідних аудіофайлів в базі даних системи, тож архітектура системи повинна бути реалізована таким чином, щоб була можливість запису даних.

Розроблювальна система представлена у вигляді веб-додатку. У багатьох аспектах успіх веб-рішення залежить від якісно проробленої архітектури, яка є фундаментом проєкту. Компонування різних елементів відповідно до цілей вебрішення гарантує його правильну функціональність і привабливість для цільової аудиторії.

Під час розробки даного проєкту було використано декілька Python фреймворків. Наприклад, для реалізації моделі розпізнавання ми скористалися бібліотекою, яка містить багато алгоритмів машинного навчання. Також для зберігання даних ми використали PostgreSQL, що забезпечує зручне локальне зберігання інформації. Крім цього, ми використали різноманітні допоміжні

інструменти для підтримки та розширення функціональності проєкту. Нижче представлено схематичне відображення архітектури розробленої системи:

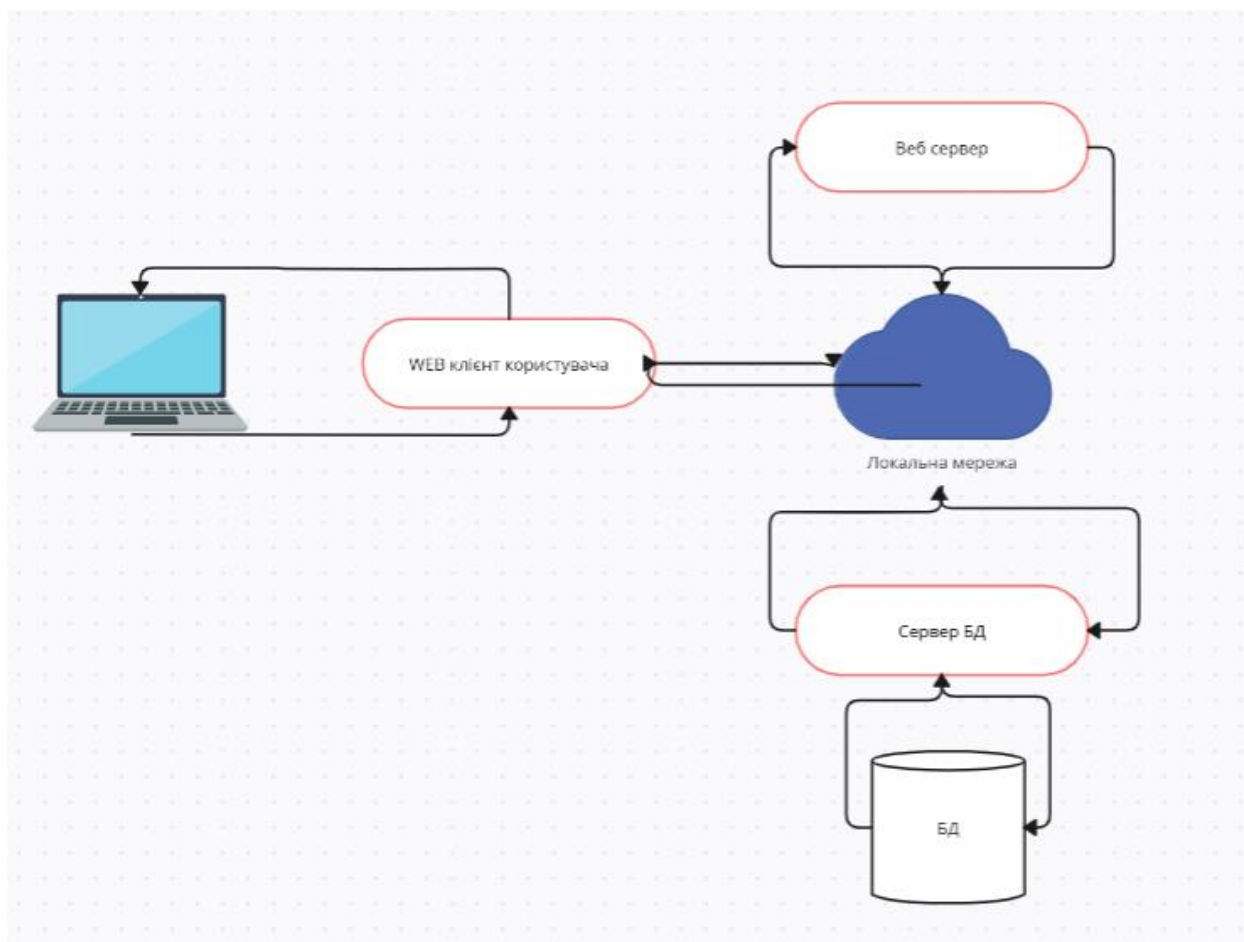


Рисунок 4.1 – Схематичне відображення архітектури системи

Задачею програмного забезпечення є класифікація емоцій за аудіозаписами.

Роботу такої системи можна описати розбиттям на такі етапи:

- 1) збір даних;
- 2) база даних;
- 3) попередня обробка аудіофайлів, виділення ознак;
- 4) розпізнавання емоцій.

4.2 Опис функціональних компонентів

Архітектура системи має відповідати призначенню та необхідним властивостям системи, забезпечувати класифікацію емоцій людини за аудіозаписом, що був наданий системі. Такий підхід дозволить проводити якісний аналіз в багатьох сферах, де відбувається безпосереднє спілкування з людьми.

Архітектура компонентної моделі:

- 1) збір даних : завантаження аудіофайлів до системи;
- 2) зберігання: зберігання аудіофайлів в системі;
- 3) перетворення: форматування отриманих файлів до відповідного формату, що здатні обробляти нейронні мережі;
- 4) класифікація: розпізнавання емоцій з використанням алгоритму машинного навчання.

Автоматизована система, яка використовується для класифікації емоцій людини, дозволяє структурувати процеси аналізу значної кількості даних, створювати інтерфейси взаємодії між компонентами, встановлювати правила для реалізації компонентів.

4.3 Обробка аудіофайлів

Розроблювана система розпізнавання емоцій за аудіозаписами працює лише з файлами в форматі (.wav):

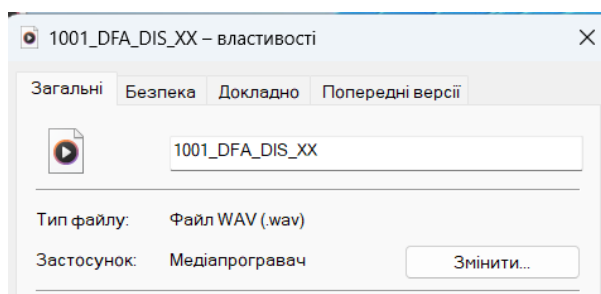


Рисунок 4.2 – Приклад формату аудіофайлу

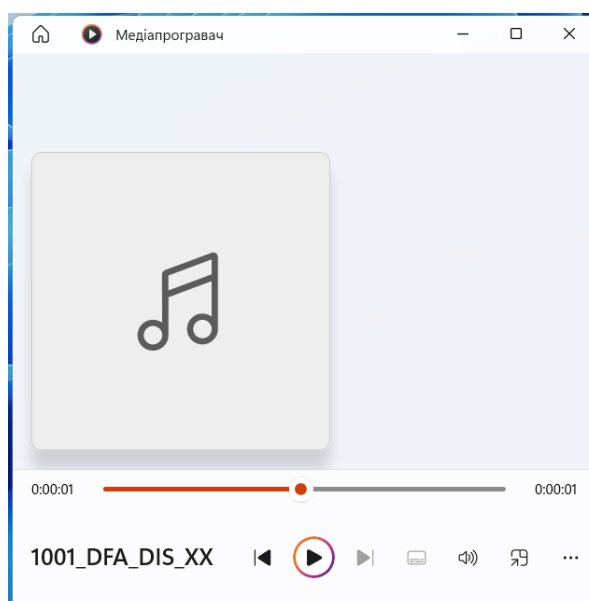


Рисунок 4.3 – Приклад програвання аудіофайлу

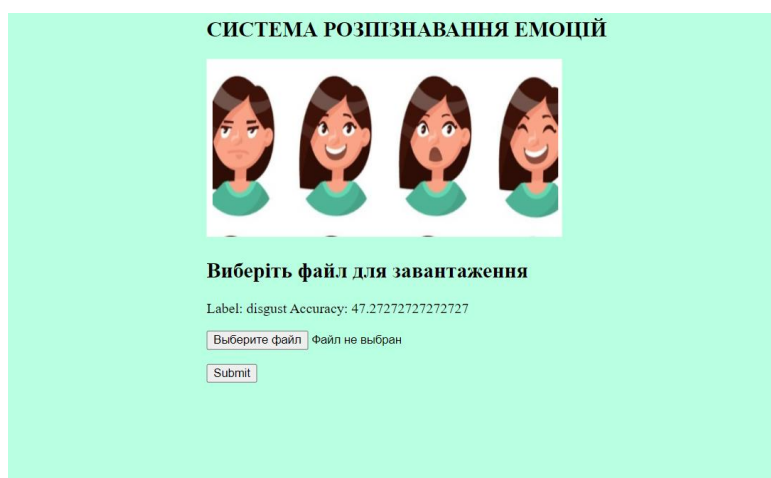


Рисунок 4.4 – Приклад зчитування системою аудіосигналу

Для здійснення класифікації емоції на аудіофайлі, попередньо відбувається форматування даних до масиву числових характеристик з яким можна працювати з нейронними мережами. Дані такого формату були завантажені та передані безпосередньо до Python, де можна прослухати, ознайомитись та проаналізувати, модифікувати (розтягнути, звужити, знешумити чи навпаки і тд.) , а також можна було побачити спектрограму таких аудіозаписів для якісного аналізу даних та навчання в моделі нейронної мережі. Нижче на рис. 4.5 та рис. 4.6 приклад спектрограми одного з аудіозаписів:

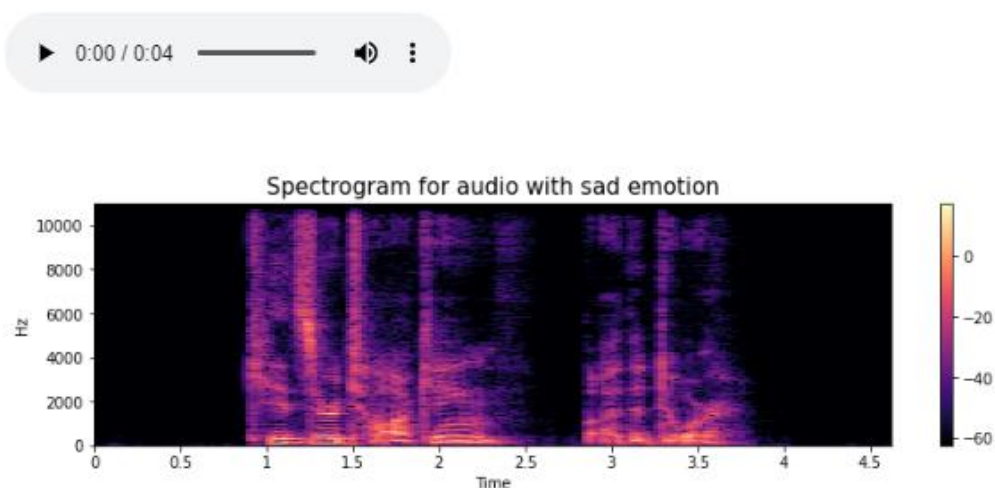


Рисунок 4.5 – Приклад відображення спектрограми

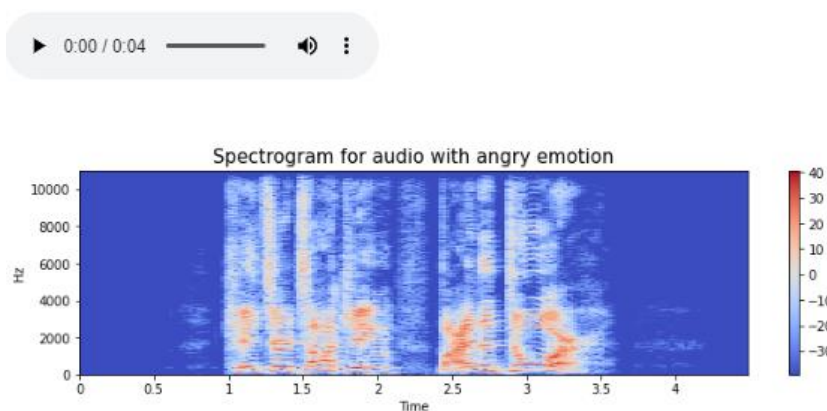


Рисунок 4.6 – Приклад відображення спектрограми

Після проведення аналізу, відбулось перетворення аудіозаписів до векторного вигляду зібраних у масиві даних. На рис. 4.7 представлення записів у такому форматі.

Для такого форматування був використаний фреймворк *librosa* це популярний інструмент для роботи зі звуком і після цього набір отриманих даних проходить процес стандартизації, оскільки стандартизація набору даних є загальною вимогою для багатьох оцінювачів машинного навчання: вони можуть працювати погано, якщо окремі функції більш-менш не схожі на стандартні дані з нормальним розподілом[18].

	0	1	2	3	4	5	6	7	8	9	...	30	31	32	33	34	35	36	37	38	39
0	-290.411469	90.227158	-14.306788	40.434174	-17.645830	3.924506	-5.248758	-4.460320	-13.240655	1.658785	...	3.634449	-3.958090	2.929164	2.366761	4.685206	3.099382	2.718958	0.987018	-2.105847	3.496448
1	-95.386702	13.887843	-5.133540	-0.475272	-7.500907	-8.790511	-4.070734	-7.247482	-7.079582	-2.067318	...	0.294736	-1.790897	-0.326462	1.881231	2.029864	2.009378	2.425562	0.388704	0.179827	1.593879
2	-341.083923	72.970833	-12.015630	33.847786	-20.121098	2.188679	-5.693862	-5.878659	-11.560396	5.038205	...	-1.404880	-0.651900	6.483722	3.780522	4.481521	1.239374	1.832815	-1.895768	2.031978	-1.688317
3	-311.221191	85.546539	-18.043245	39.109692	-19.696962	2.480769	-5.808243	-5.094702	-13.628708	1.218207	...	3.581122	-4.390938	2.922006	3.406705	5.895154	3.283037	2.773895	1.753276	-2.546918	2.855142
4	-311.221191	85.546539	-18.043245	39.109692	-19.696962	2.480769	-5.808243	-5.094702	-13.628708	1.218207	...	3.581122	-4.390938	2.922006	3.406705	5.895154	3.283037	2.773895	1.753276	-2.546918	2.855142
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
365	-290.411469	90.227158	-14.306788	40.434174	-17.645830	3.924506	-5.248758	-4.460320	-13.240655	1.658785	...	3.634449	-3.958090	2.929164	2.366761	4.685206	3.099382	2.718958	0.987018	-2.105847	3.496448
366	-264.831182	72.115293	-9.994488	31.859417	-12.232098	0.790542	-2.453887	-5.795645	-11.946329	0.581234	...	2.936360	-3.231917	2.329897	3.272208	4.109806	3.442018	2.656977	0.637581	-1.670078	3.087957
367	-341.083923	72.970833	-12.015630	33.847786	-20.121098	2.188679	-5.693862	-5.878659	-11.560396	5.038205	...	-1.404880	-0.651900	6.483722	3.780522	4.481521	1.239374	1.832815	-1.895768	2.031978	-1.688317
368	-311.221191	85.546539	-18.043245	39.109692	-19.696962	2.480769	-5.808243	-5.094702	-13.628708	1.218207	...	3.581122	-4.390938	2.922006	3.406705	5.895154	3.283037	2.773895	1.753276	-2.546918	2.855142
369	-311.221191	85.546539	-18.043245	39.109692	-19.696962	2.480769	-5.808243	-5.094702	-13.628708	1.218207	...	3.581122	-4.390938	2.922006	3.406705	5.895154	3.283037	2.773895	1.753276	-2.546918	2.855142

370 rows x 40 columns

Рисунок 4.7 – Приклад відформатованих даних

Додатковою особливістю такого рішення було те, що один файл був представлений в п'яти інтерпретаціях:

- 1) звичайні дані;
- 2) дані з шумом;
- 3) дані з розтягуванням і кроком;
- 4) дані з високою швидкістю;
- 5) дані з низькою швидкістю;

і на тестування серед набору розпізнаних емоцій обирається найчастіше розпізнана.

Отже, розподіл даних в початковому наборі даних продемонстровано на рисунку 4.8, а перетвореного та того за допомогою, якого буде навчатись нейронна мережа відображено на рисунку 4.8

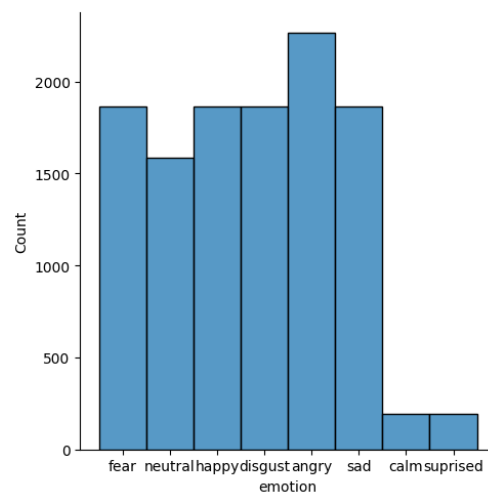


Рисунок 4.8 – Розподіл даних для початкових даних

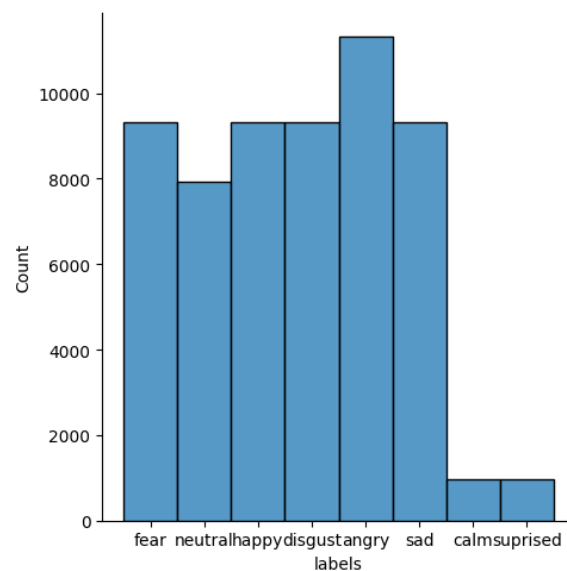


Рисунок 4.9 – Розподіл даних для відформатованих даних

4.4 Діаграма послідовності

Діаграма послідовності є графічною ілюстрацією, яка демонструє взаємодію компонентів системи, переходи від одного стану до іншого та послідовність виконання операцій. Вона чітко відображає алгоритм роботи системи. На діаграмі

зображено процес завантаження даних на сервер, розпізнавання емоцій за аудіозаписами та збереження результату в сховище даних. Це відображає послідовні дії, де дані спочатку передаються на сервер, потім відбувається аналіз та розпізнавання емоцій, а результат зберігається для подальшого використання.



Рисунок 4.10 – Діаграма послідовності

- 1) користувач надсилає файл з аудіосигналом на сервер;
- 2) перетворення аудіосигналу до векторної форми;
- 3) класифікація емоції за отриманим файлом запису голосу;
- 4) результат надсилається на сервер та зберігається в сховищі даних.

Так само на діаграмі показано процес завантаження даних, перетворення даних та розпізнавання емоції на конкретному аудіофайлі. Врешті, результати зберігаються у сховищі даних. Тут наочно продемонстровано послідовність операцій і їх взаємодію в системі, що сприяє розумінню та аналізу їх функціонування.

4.5 Інструкція роботи з системою розпізнавання емоцій

Для роботи з системою необхідно мати доступ до її вебверсії. Користувачу буде запропоновано лише два кроки на екрані:

- 1) «Обрати»: обрати та завантажити файл формату .wav, який містить запис голосу, емоцію носія, якого хотілося б розпізнати;
- 2) «Надіслати»: надіслати файл на сервер для подальшої обробки сигналу та класифікації емоції;

Після виконання заданих кроків, користувач має змогу отримати результат, а саме відповідь на питання, які емоції характерні диктору з завантаженого аудіофайлу. На рис. 4.11 та рис. 4.12 проілюстровано такі дії.

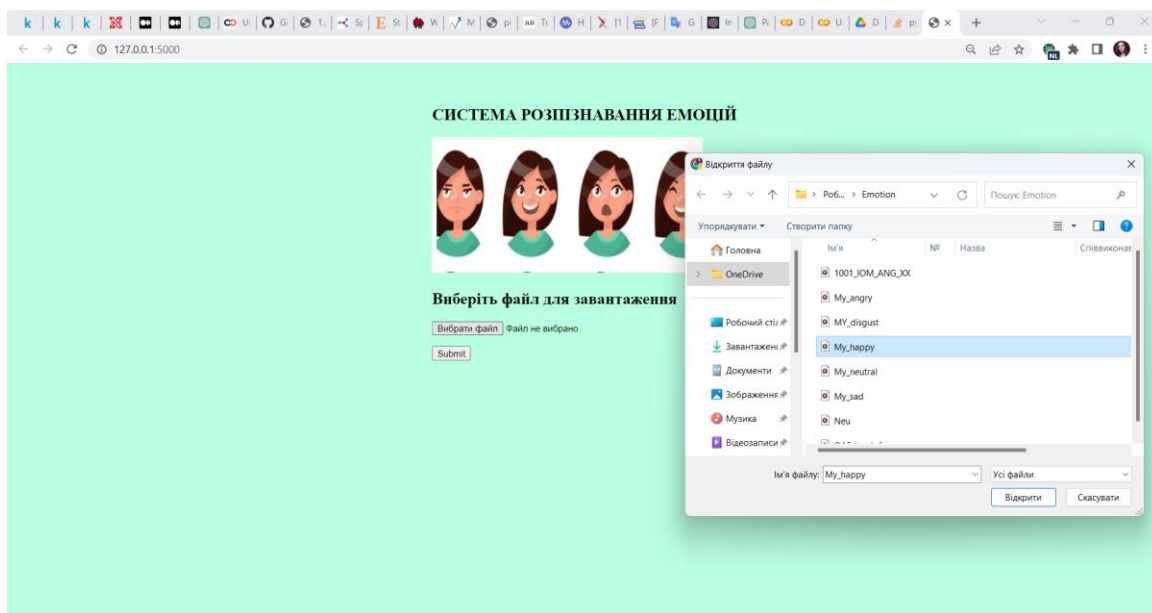


Рисунок 4.11 – Демонстрація роботи системи

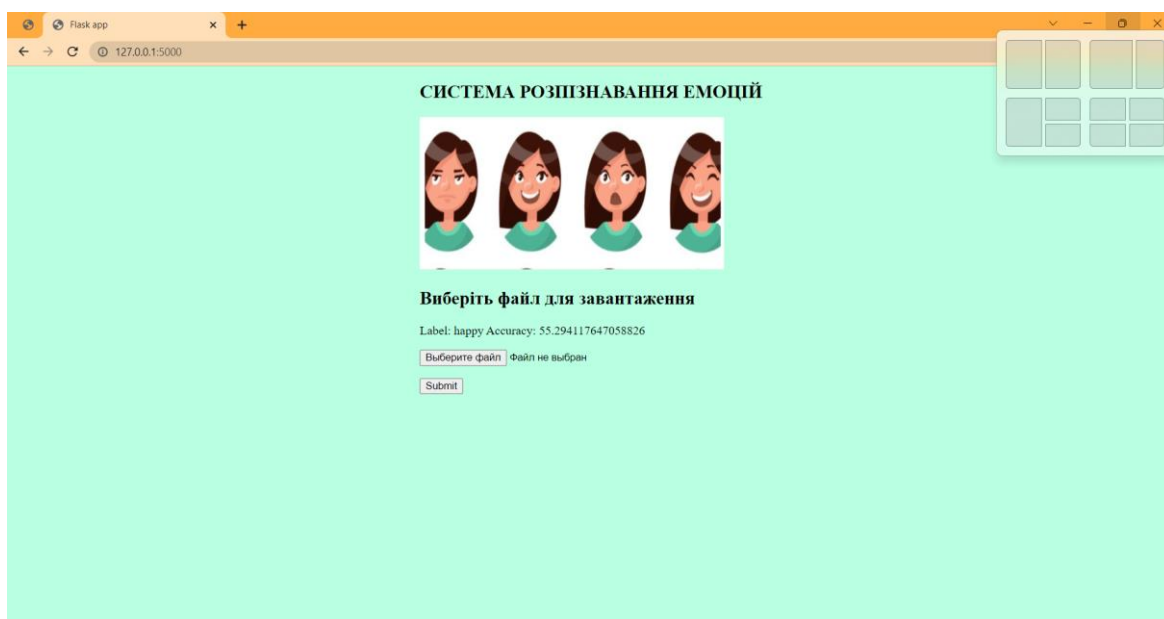


Рисунок 4.11 – Демонстрація роботи системи

4.6 Висновки до розділу компонентна модель системи

В цьому розділі було надано опис компонентів, що входять до системи розпізнавання емоцій за аудіозаписами, як вони взаємодіють між собою та цілому в системі. Також було детально розказано, з яким файлами та типом даних взаємодіє система, а також складову системи, як перетворення аудіосигналів, оскільки – це один з найважливіших етапів в розпізнаванні емоцій, бо нейронна мережа здатна навчатись та тренуватись лише на даних відповідного векторного представлення. Окрім того, надано детальну інформацію щодо утворення збільшення набору даних для подальшого навчання класифікаторів емоцій за аудіосигналами. Було розказано про створення системи з використанням фреймворку для роботи з API, а також надано зразки її використання. І врешті продемонстровано інструкцію користування системою. Ці приклади показують, ступінь ефективності системи.

5 ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

Основною метою цієї роботи є створення елементарної, ефективної та практичної моделі, яка використовує техніку машинного навчання як його основну функцію, для створення яких ми можемо довіряти системі точні та безпомилкові результати. За допомогою мови програмування Python та фреймворку для обробки аудіофайлів Librosa було реалізовано програмне забезпечення розпізнавання емоцій за аудіозаписами.

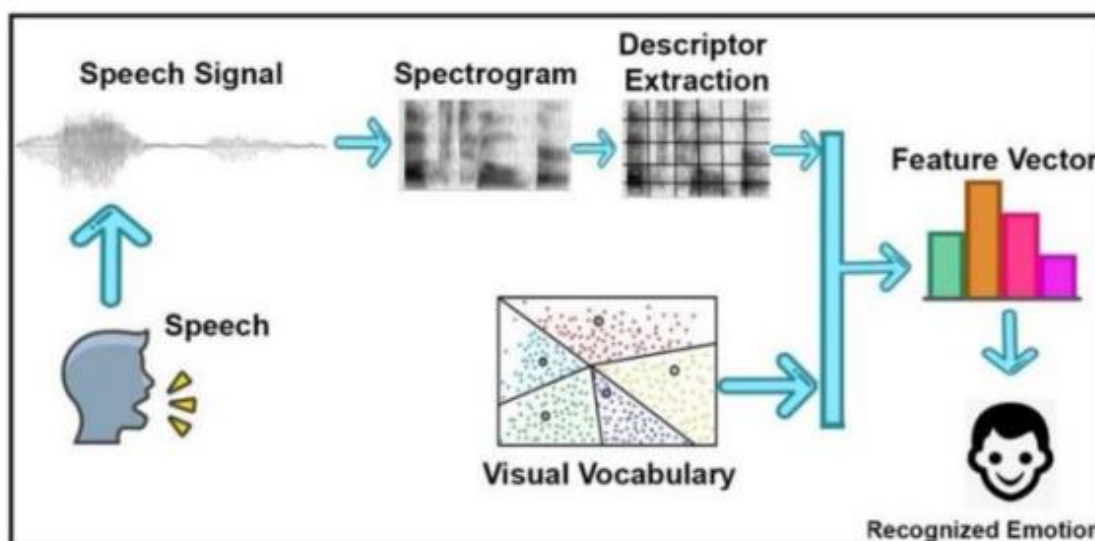


Рисунок 5.1 – Схематичне відображення етапів системи розпізнавання емоцій[19]

Було створено систему, щоб автоматизовано розпізнати емоцію диктора, додатково зібрати більшу кількість зразків записів для подальшого навчання та покращення системи для аналізу, і в результаті створено класифікацію емоцій за короткими звуковими доріжками на основі шаблону. Така система є автоматизованою, тож не потребує налаштувань, а від користувача необхідне лише завантажити файл з аудіосигналом. Обрана стратегія є кращою для розпізнавання аудіосигналу. Вхідні змінні для цього класифікатора жодним чином не обмежені, тож

можна надавати обрані аудіосигнали, оскільки вони проходять попереднє форматування.

5.1 Алгоритм роботи програмного забезпечення

Програмне забезпечення було реалізовано за допомогою мови програмування Python та з використанням бібліотек opencv, torch, tensorflow та keras.

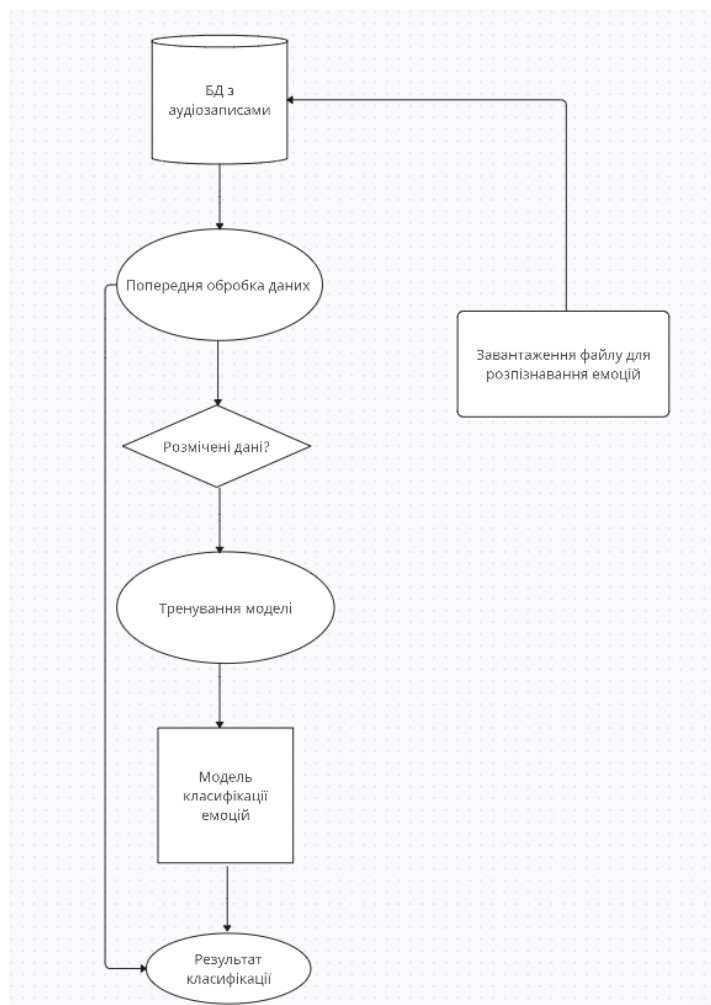


Рисунок 5.2 – Схематичне відображення роботи програмного забезпечення

Алгоритм роботи програмного забезпечення можна описати:

- 1) для навчання дані зчитуються з бази даних системи, яка містить аудіозаписи;
- 2) вибірки для навчання та тренування моделі були отримані внаслідок розділення набору даних, що буде описаний в наступному розділі. Для цього була використана готова функція з фреймворку sklearn, а саме `train_test_split`, тренувальна вибірка була використана для навчання моделі класифікації, а призначення тестової вибірки перевірити точність та продуктивність навченої моделі. Розбиття набору даних на тренувальну та тестову вибірки було зроблено в відношенні 80 до 20 відповідно;
- 3) для класифікації емоцій було використано згорткові нейронні мережі;
- 4) файли, що були завантажені до системи для розпізнавання зберігають у її БД.

5.2 Опис даних

Для реалізації такої задачі та навчання моделі розпізнавання емоцій за аудіозаписами було використано комбінацію наборів даних аудіозаписів з відкритих та доступних джерел, утворений набір містить файли з такими типами емоцій:

- 1) спокою;
- 2) радості;
- 3) суму;
- 4) сердитості;
- 5) страху;
- 6) здивування;
- 7) огиди.

Першим набором даних з комбінації був RAVDESS — це база даних емоційної мови. База даних збалансована та складається з 24 професійних акторів, які озвучують лексично відповідні твердження з нейтральним північноамериканським акцентом,

кожен вираз створюється на двох рівнях емоційної інтенсивності з додатковим нейтральним виразом[20]. На рисунку 5.3, графічне відображення розподілу емоцій у даній вибірці.

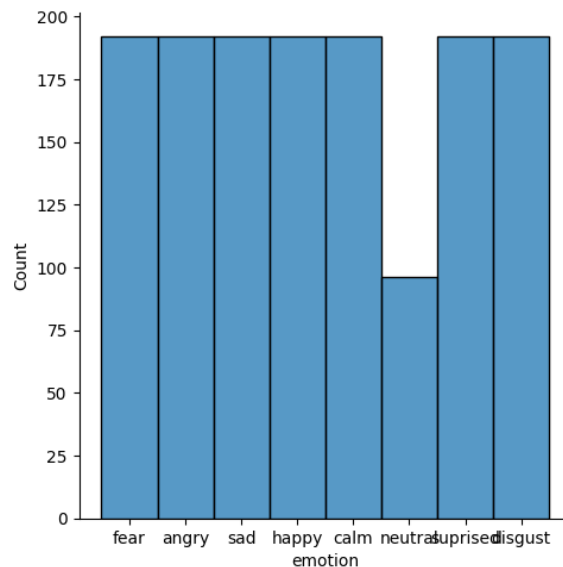


Рисунок 5.3 – Розподіл даних в наборі RAVDESS

Наступним набором даних з комбінації був CREMA-D це база даних, де люди передають свій емоційний стан голосом. Набір даних, який унікально підходить для вивчення мультимодального вираження емоцій і сприйняття. 7442 запис 91 актора різного етнічного походження були оцінені кількома експертами. Категоричні позначки емоцій і реальні значення інтенсивності емоцій, що сприймаються, були зібрані за допомогою краудсорсингу від 2443 оцінювачів[21]. На рисунку 5.4, графічне відображення розподілу емоцій у даній вибірці.

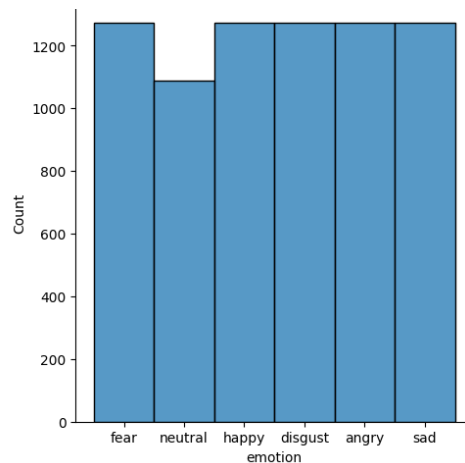


Рисунок 5.4 – Розподіл даних в наборі CREMA-D

Останнім набором даних з комбінації був Toronto emotional speech set (TESS) – набір аудіозаписів, що були змодельовані на основі слухового тесту № 6 Північно-західного університету (NU-6; Tillman & Carhart, 1966). Дві актриси (віком 26 і 64 роки) вимовили набір із 200 цільових слів у фразі-носії «Скажи слово _____», і було зроблено записи, що відображають кожен із семи емоцій [22]. На рисунку 5.5, графічне відображення розподілу емоцій у даній вибірці.

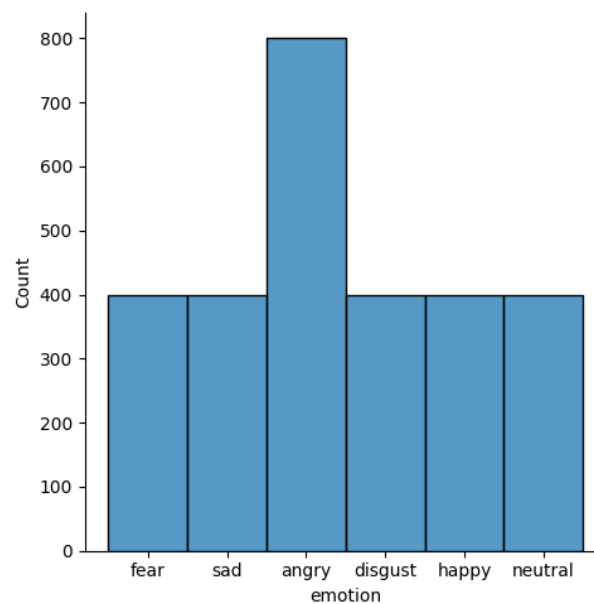


Рисунок 5.5 – Розподіл даних в наборі TESS

5.3 Архітектура нейронної мережі

В якості моделі машинного навчання на основі експериментальних випробувань було обрано CNN з автоматизованим підбором гіперпараметрів. Нижче на рис.5.6 буде продемонстровано архітектуру моделі розпізнавання емоцій.

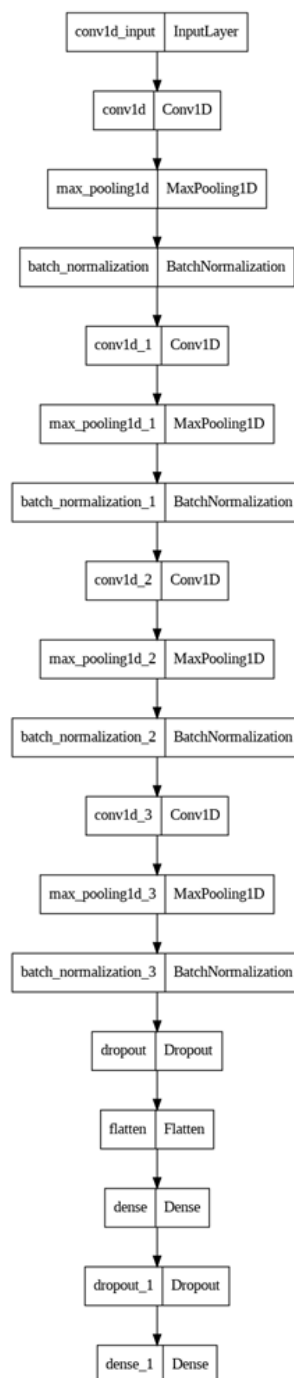


Рисунок 5.6 – Архітектура нейронної мережі

5.4 Опис результатів

В першу чергу варто зазначити, що набір даних був розбитий на навчальну та тестові вибірки в пропорції 80 до 20 відповідно. Також для тестування та в майбутньому для демонстрації були записані власні короткі аудіофайли.

Для перевірки результатів було використано декілька підходів перевірки ефективності моделі розпізнавання - в першу чергу – це матриця помилок. Такий підхід наглядно демонструє яка кількість аудіофайлів з тестової вибірки були розпізнані вірно й хибно та за якою саме характеристикою. Нижче на рис. 5.7 можна ознайомитись з матрицею помилок для навченої моделі.

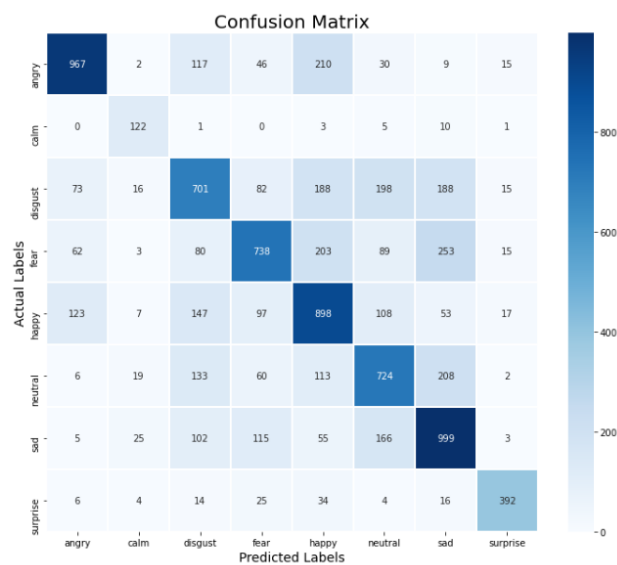


Рисунок 5.7 – Матриця помилок отриманої моделі

Також в якості метрик якості були використані Precision, Recall та F1-score - важливі показниками для оцінки продуктивності моделі класифікації, і вони особливо корисні в класифікації.

	precision	recall	f1-score	support
angry	0.78	0.69	0.73	1396
calm	0.62	0.86	0.72	142
disgust	0.54	0.48	0.51	1461
fear	0.63	0.51	0.57	1443
happy	0.53	0.62	0.57	1450
neutral	0.55	0.57	0.56	1265
sad	0.58	0.68	0.62	1470
surprise	0.85	0.79	0.82	495
accuracy			0.61	9122
macro avg	0.63	0.65	0.64	9122
weighted avg	0.61	0.61	0.61	9122

Рисунок 5.8 – Точність класифікації отриманої моделі

Також був використаний підхід перехресної валідації для оцінки якості розпізнавання. Така метрика працює, як вибір середнього значення серед отриманих десяти результатів за десятьма розбиттями, тож середнє значення між заданими розбиттями становить 84%, що є доволі хорошим результатом та задовольняє вимогам до даної роботи.

Також нижче представлено приклади результатів розпізнавання емоцій системою.

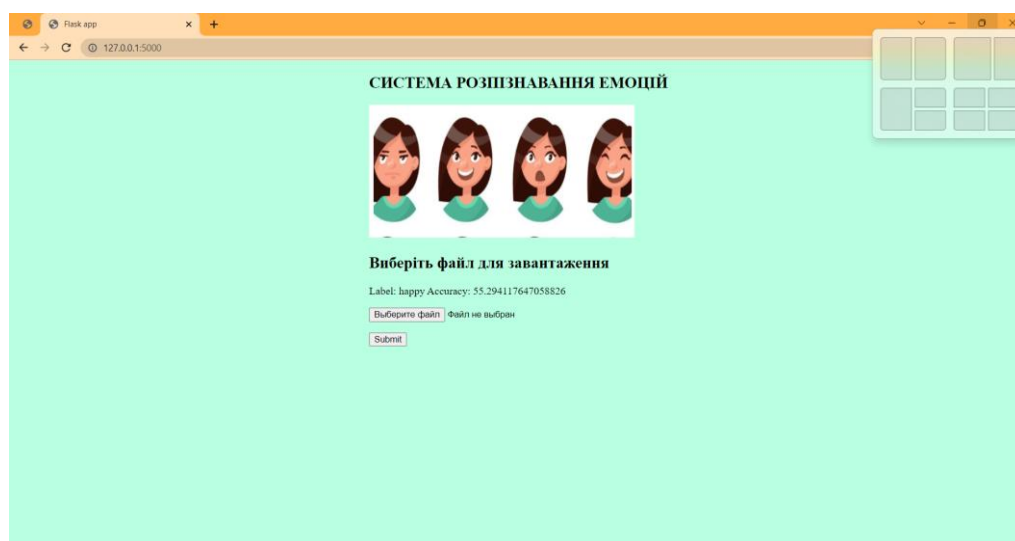


Рисунок 5.9 – Приклад розпізнавання емоції за аудіозаписом.

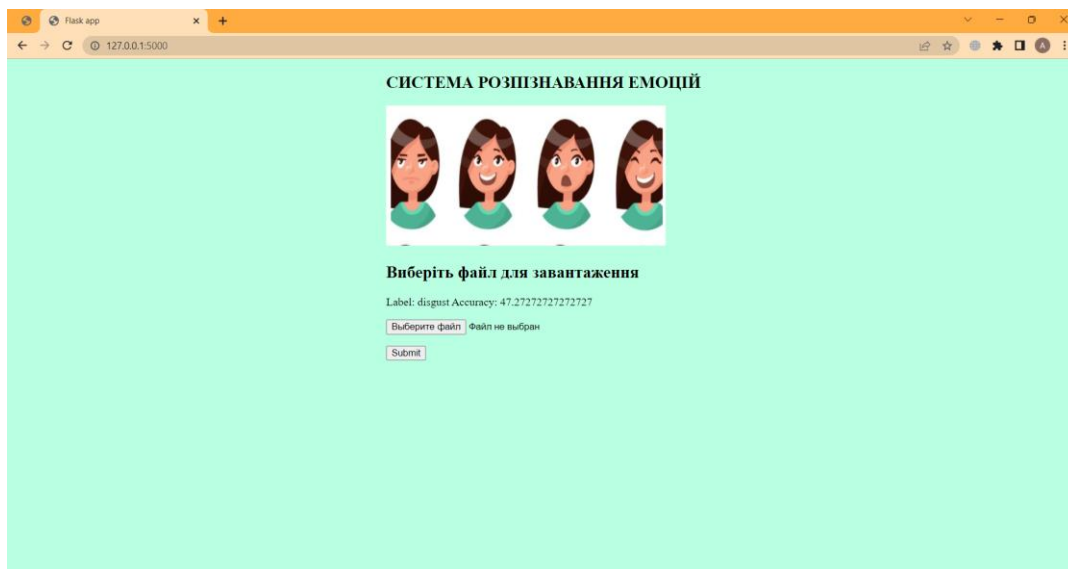


Рисунок 5.10 – Приклад розпізнавання емоції за аудіозаписом.

5.5 Висновки до розділу програмне забезпечення

У розділі 5:

- 1) представлено програмне забезпечення для системи розпізнавання емоцій за аудіозаписами;
- 2) окремо описано етап отримання та опрацювання даних для подальшого вирішення задачі;
- 3) розказано про обґрунтований вибір нейронної мережі, підбір гіперпараметрів та проілюстровано архітектуру отриманої моделі нейронної мережі;
- 4) на етапі оцінювання було обрано декілька стратегій для якісної перевірки ефективності роботи розробки та проілюстровано отримані результати.

ВИСНОВКИ

У дисертаційній роботі одержано такі нові теоретичні та практичні результати:

1. Досліджено існуючі системи розпізнавання емоцій за аудіозаписами, методи класифікації та алгоритми оптимізації для підбору найкращої комбінації параметрів цих методів, а також існуючі датасети, що стосуються цієї тематики і прийнято рішення використання їх комбінації задля кращих результатів ефективності моделі. Встановлено, що використання такої системи допомагає краще розуміти емоції людини, а також зберігає ресурси, через автоматизацію цього процесу, а для здійснення задачі класифікації використовується: методи обробки аудіофайлу до векторної форми, методи машинного навчання для розпізнавання емоцій та метод оптимізації гіперпараметрів для пошуку найкращої комбінації налаштувань, а також встановлено що існуючі системи є мають більш складні підходи, хоча не є ефективними, також недоліком є те, що вони закриті й не доступні кожному. На основі порівняння характеристик цих методів було встановлено, що найкращим вибором для побудови системи метод перетворення аудіосигналу до векторної форми MFCC, згорткові нейронні мережі для розпізнавання емоцій та метод оптимізації гіперпараметрів Hyperband для пошуку найкращої комбінації налаштувань

2. Розроблено архітектуру, систему, модель і програмне забезпечення для вирішення задачі розпізнавання емоцій за аудіозаписами, а також математичне забезпечення системи. Воно представлено, як опис окремих компонентів моделі: методу MFCC, згорткових нейронних мереж, а також методу оптимізації Hyperband. Описано також формули показників ефективності за якими, були оцінені та порівняні моделі.

3. Програмне забезпечення було реалізовано мовою програмування python з допоміжними інструментами: pandas, matplotlib, seaborn, skikit-learn, librosa з використанням яких було реалізовано систему розпізнавання емоцій людини за аудіофайлам , у якій користувачу надається можливість завантажити аудіофайл в

певному форматі для визначення емоції на основі згорткових нейронних мереж. Було розроблено методичні матеріали, що надають користувачам можливість налаштовувати програмне забезпечення з метою подальшої його ефективної використання. Ці засоби дозволяють користувачам налагоджувати належне зчитування завантажених файлів для розпізнавання емоцій та забезпечувати правильну роботу всього функціоналу, включаючи переформатування аудіосигналу, розпізнавання та взагалі нормальне функціонування веб-застосунку.

4. Результати аналізу розробленого програмного забезпечення були отримані шляхом проведення тестування на спеціально створених аудіозаписах та тестовій вибірці. При цьому використовувалися описані у дослідженні метрики якості. Отримані показники ефективності класифікації були представлені у вигляді графіків та таблиць порівнянь. Для оцінки результатів були розраховані значення Precision, Recall та F1-показників, а також побудовано матрицю помилок. Це дозволяє оцінити ефективність розпізнавання аудіосигналу та визначити, наскільки можна довіряти отриманим результатам. Такий підхід надає можливість представникам підприємств та установ вплинути на ефективність та якість роботи своєї організації та покращити їх показники, якщо знадобиться.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Stitchlabs.com [електронний ресурс]: інтернет сервіс для вибору музики. Режим доступу: <https://open.spotify.com/>
2. Armand S. Rotaru, Gabriella Vigliocco, Constructing Semantic Models From Words, Images, and Emojis // Cognitive Science. – 2020. – Т. 44. – С. e12830
3. audeering.com [електронний ресурс]: інтернет сервіс розпізнавання емоцій за аудіозаписами. Режим доступу: <https://www.audeering.com/technology/>
4. Luxton D.D., Bickmore T. Intelligent Mobile, Wearable, and Ambient Technologies for Behavioral Health Care // Artificial Intelligence in Behavioral and Mental Health Care. – 2016. – P. 121–140.
5. Ван І. Consonant-vowel interaction and phonological structure processing in cross-language perceptual assimilation: Australian English CV syllables perceived by L1 Mandarin listeners / І. Ван // Master of Arts in Applied Linguistics. – University of Melbourne, 2018.
6. Sithara A., Thomas A., Mathew D. Study of MFCC and IHC Feature Extraction Methods With Probabilistic Acoustic Models for Speaker Biometric Applications / 8th International Conference on Advances in Computing and Communication (ICACC-2018). – 2018. – P. 267–276.
7. Молау С., Піц М., Шлютер Р., Ней Г. Computing Mel-frequency cepstral coefficients on the power spectrum // 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. – 2001. – С. 73-76.
8. Mahesh T. R., Kumar D., Kumar V. V., Asghar J., Bazezew B. M., Natarajan R., Vivek V. Blended Ensemble Learning Prediction Model for Strengthening Diagnosis and Treatment of Chronic Diabetes Disease // Computational Intelligence and Neuroscience. 2022. Vol. 2022. P. 4451792. DOI: 10.1155/2022/4451792.

9. Hossain S. K. M., Ema S. A., Sohn H. Rule-Based Classification Based on Ant Colony Optimization: A Comprehensive Review // *Applied Computational Intelligence and Soft Computing*. 2022. Vol. 2022. P. 2232000. DOI: 10.1155/2022/2232000.

10. scikit-learn.org [електронний ресурс]: документація до методу класифікації MLPClassifier. Режим доступу: https://scikit-learn.org/stable/modules/neural_networks_supervised.html#

11. Джемисон К., Талвалкар А. Non-stochastic Best Arm Identification and Hyperparameter Optimization // *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. – 2016. – Т. 51. – С. 240-248.

12. scikit-learn.org [електронний ресурс]: документація до методу перехресної перевірки. Режим доступу: https://scikit-learn.org/stable/modules/cross_validation.html

13. Sithara A., Thomas A., Mathew D. Study of MFCC and IHC Feature Extraction Methods With Probabilistic Acoustic Models for Speaker Biometric Applications / A. Sithara, A. Thomas, D. Mathew // *Procedia Computer Science*. – 2018. – Vol. 143. – P. 267-276.

14. Indolia S., Goswami A.K., Mishra S.P., Asopa P. Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach / S. Indolia, A.K. Goswami, S.P. Mishra, P. Asopa // *Procedia Computer Science*. – 2018. – Vol. 132. – P. 679-688.

15. Wang, Jiayun, et al. "Orthogonal convolutional neural networks." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

16. Третиник В.В., Коваленко О.П. Автоматизація підбору гіперпараметрів lstm для задачі розпізнавання емоцій за аудіозаписами // *Прикладна математика та комп'ютинг. ПМК, 2022 : п'ятнадцята наук. конф. магістрантів та аспірантів, Київ, 16—18 лист. 2022 р. : зб. тез доп. / [редкол.: Дичка І. А. та ін.]. — К. : Просвіта, 2022. — с.174-179* ISBN 978-617-7010-23-3

17. Лі Л., Джемисон К., Де Сальво Дж., Ростамізадех А., Талвалкар А. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization // *Journal of Machine Learning Research*. – 2018. – Т. 18. – С. 1-52.

18. scikit-learn.org [електронний ресурс]: документація до методу стандартизації даних. Режим доступу: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
19. Babu P., Nagaraju V. S., Vallabhuni R. R. Speech Emotion Recognition System With Librosa // 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT). 2021. P. 18-19 June 2021. DOI: 10.1109/CSNT51715.2021.9509714.
20. Livingstone SR, Russo FA (2018) The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. PLoS ONE 13(5): e0196391
21. Houwei Cao, David G. Cooper, Michael K. Keutmann, Ruben C. Gur, Ani Nenkova, and Ragini Verma, CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset, IEEE Trans Affect Comput. Author manuscript; available in PMC 2015 Feb 2. IEEE Trans Affect Comput. 2014 Oct-Dec; 5(4): 377–390.
22. Дюпюї К., Пікора-Фуллер М. К. Toronto emotional speech set (TESS) // TSpace Repository. – 2010.

Додаток А

Лістинги програм

Лістинг файлу `diplom.py`

```
from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense

from tensorflow.keras import utils

import numpy as np

#Importing The Dependencies

import pandas as pd

import numpy as np

#import matplotlib.pyplot as plt

import librosa

import IPython.display as ipd

#from librosa.display import waveshow

#import seaborn as sns

from sklearn.preprocessing import OneHotEncoder,StandardScaler

from sklearn.neural_network import MLPClassifier

from sklearn.model_selection import train_test_split

import tensorflow

import tensorflow as tf

#import tensorflow_hub as hub

from tensorflow.keras import Sequential

from tensorflow.keras.layers import Conv1D,Dense,AveragePooling1D,Dropout,add,BatchNormalization,LSTM,Flatten,MaxPooling1D,GRU

from tensorflow.keras.optimizers import Adam

from tensorflow.keras.activations import relu, softmax

from keras.activations import relu, softmax

from sklearn.metrics import confusion_matrix,classification_report

import pickle
```

```

my_model = Sequential()

my_model.add(Conv1D(256, kernel_size=7, strides=1, padding='same', activation='relu', input_shape=(40, 1)))

my_model.add(MaxPooling1D(pool_size=2, strides=2, padding='same'))

my_model.add(BatchNormalization())

my_model.add(Conv1D(128, kernel_size=7, strides=1, padding='same', activation='relu'))

my_model.add(MaxPooling1D(pool_size=2, strides=2, padding='same'))

my_model.add(BatchNormalization())

my_model.add(Conv1D(128, kernel_size=7, strides=1, padding='same', activation='relu'))

my_model.add(MaxPooling1D(pool_size=2, strides=2, padding='same'))

my_model.add(BatchNormalization())

my_model.add(Conv1D(64, kernel_size=7, strides=1, padding='same', activation='relu'))

my_model.add(MaxPooling1D(pool_size=2, strides=2, padding='same'))

my_model.add(BatchNormalization())

my_model.add(Dropout(0.5))

my_model.add(Flatten())

my_model.add(Dense(units=32, activation='relu'))

my_model.add(Dropout(0.5))

my_model.add(Dense(units=8, activation='softmax'))

my_model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

my_model.load_weights('SOUND21_RELU')

def getPrediction(filename):

    data, sample_rate = librosa.load('uploads/'+filename)

    def feature_extraction(file):

        mfcc_features = librosa.feature.mfcc(y=file, sr=sample_rate, n_mfcc=40)

        mfcc_scaled_feature = np.mean(mfcc_features.T, axis=0)

        return mfcc_scaled_feature

    # NOISE

    def noise(data):

        noise_amp = 0.035 * np.random.uniform() * np.amax(data)

        data = data + noise_amp * np.random.normal(size=data.shape[0])

        return data

    # STRETCH

```

```

def stretch(data, rate=0.8):

    return librosa.effects.time_stretch(data, rate)

# SHIFT

def shift(data):

    shift_range = int(np.random.uniform(low=-5, high=5) * 1000)

    return np.roll(data, shift_range)

# PITCH

def pitch(data, sampling_rate, pitch_factor=0.7):

    return librosa.effects.pitch_shift(data, sampling_rate, pitch_factor)

# higher speed

def higher_speed(data, speed_factor=1.25):

    return librosa.effects.time_stretch(data, speed_factor)

# lower speed

def lower_speed(data, speed_factor=0.75):

    return librosa.effects.time_stretch(data, speed_factor)

def get_feat(path):

    data, sample_rate = librosa.load(path, duration=2.5, offset=0.6)

    # normal data

    res1 = feature_extraction(data)

    result = np.array(res1)

    # data with noise

    noise_data = noise(data)

    res2 = feature_extraction(noise_data)

    result = np.vstack((result, res2))

    # data with stretch and pitch

    new_data = stretch(data)

    data_stretch_pitch = pitch(new_data, sample_rate)

    res3 = feature_extraction(data_stretch_pitch)

    result = np.vstack((result, res3))

    # data with high speed

    high_speed = higher_speed(data)

    res4 = feature_extraction(high_speed)

```

```

result = np.vstack((result, res4))

# data with low speed

low_speed = lower_speed(data)

res5 = feature_extraction(high_speed)

result = np.vstack((result, res5))

return result

x = []

for path in (zip('uploads/'+filename)):

    feature = get_feat('uploads/'+filename)

    for ele in feature:

        x.append(ele)

x_test = pd.DataFrame(x)

scaler = StandardScaler()

x_test = scaler.fit_transform(x_test)

x_testcnn = np.expand_dims(x_test, axis=2)

with open('data.pickle', 'rb') as f:

    Y = pickle.load(f)

encoder = OneHotEncoder()

Y = encoder.fit_transform(np.array(Y).reshape(-1, 1)).toarray()

pred_test = my_model.predict(x_testcnn)

y_pred = encoder.inverse_transform(pred_test)

vv = pd.DataFrame(y_pred).value_counts()

result = ((vv.loc[vv == vv[0]].index)[0])[0]

print((pd.DataFrame(y_pred).value_counts()))

print(x_testcnn.shape)

print(feature.shape)

return result, 100 * vv[0]/vv.sum()

```

Додаток Б
Ілюстративний матеріал

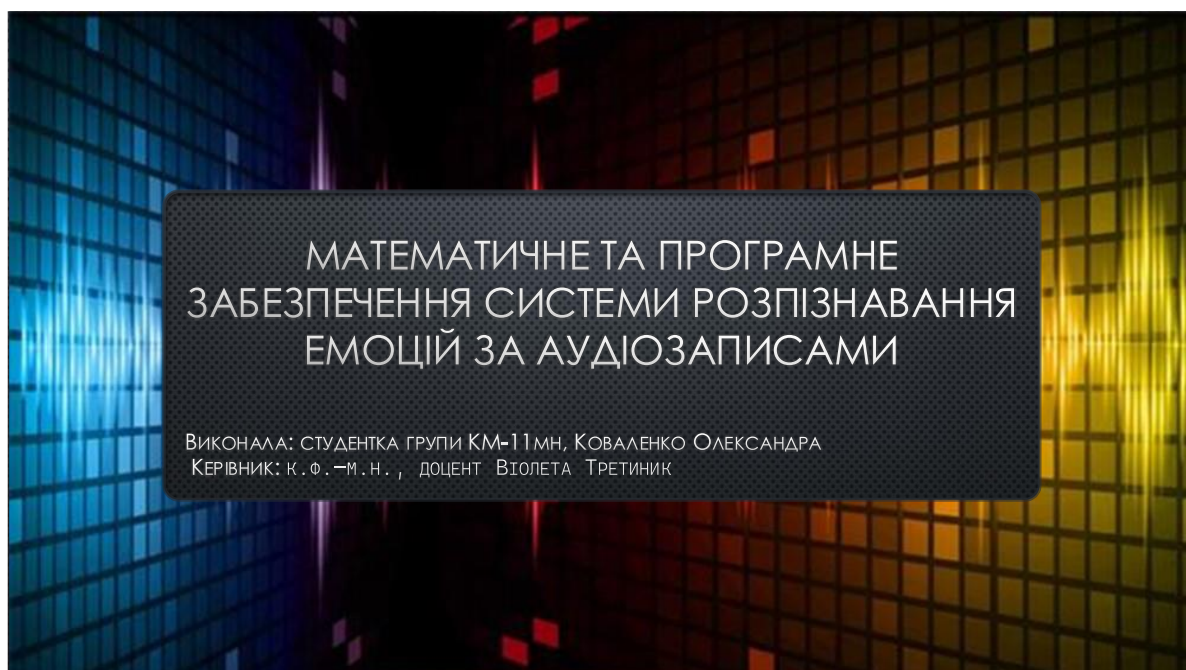


Рисунок Б.1 – Слайд 1

ВСТУП

- РОЗПІЗНАВАННЯ ЕМОЦІЙ Є ВИКЛИКОМ НАВІТЬ ДЛЯ ЛЮДЕЙ І ВИМАГАЄ УВАГИ ДО КІЛЬКОХ АСПЕКТІВ, ТАКИХ ЯК МОВА ТІЛА, ВИРАЗ ОБЛИЧЧЯ ТА ГОЛОСОВИЙ ТОН. РОЗПІЗНАВАННЯ МОВНОЇ ЕМОЦІЇ МОЖЕ БУТИ ЗНАЧНИМ КРОКОМ УПЕРЕД, ОСКІЛЬКИ ВОНО ВКЛЮЧАЄ ШИРОКИЙ СПЕКТР СИТУАЦІЙ.
- СИСТЕМА, ЯКА ВИКОРИСТОВУЄ ЛИШЕ ОДНУ МОДАЛЬНІСТЬ - ГОЛОСОВІ ЗАПИСИ - СТАЄ СКЛАДНІШОЮ І ВИМАГАЄ МІКРОФОН ЯК ДОДАТКОВИЙ ЗАСІБ. ПРАВИЛЬНІ РЕЗУЛЬТАТИ В ЦІЙ ОБЛАСТІ МОЖУТЬ СПРИЯТИ БІЛЬШ ГУМАНІЗОВАНОМУ РОЗВИТКУ МАШИН, ОСКІЛЬКИ ЕМОЦІЇ Є НЕВІД'ЄМНОЮ ЧАСТИНОЮ ЛЮДСЬКОСТІ. МАШИНИ, ЗДАТНІ РОЗУМІТИ НАСТРОЇ ТА НАМІРИ ЛЮДЕЙ, МОЖУТЬ ЗНАЙТИ ЗАСТОСУВАННЯ У БЕЗПЕЦІ, МЕДИЦИНІ, ЕКСТРЕНИХ СЛУЖБАХ, ТЕЛЕМАРКЕТИНГУ ТА СОЦІАЛЬНИХ УСТАНОВАХ. ЦЕ МОЖЕ ДОПОМОГТИ ЛЮДЯМ З ОБМЕЖЕНИМИ ЗДІБНОСТЯМИ РОЗУМІТИ ЕМОЦІЇ СПІВРОЗМОВНИКА.
- ДАНА РОБОТА, ПРО РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ ЗА АУДІОЗАПИСАМИ СПРЯМОВАНА НА ВИРІШЕННЯ ПРОБЛЕМ ГАЛУЗЯХ: ВІЙСЬКОВІЙ, РОЗВАЖАЛЬНІЙ, ОСВІТНІЙ ТА МЕДИЧНІЙ.

2

Рисунок Б.2 – Слайд 2

ПОСТАНОВКА ЗАДАЧІ

- МЕТА ДОСЛІДЖЕННЯ – АНАЛІЗ, ОБРОБКА АУДІОЗАПИСІВ І ЇХ КЛАСИФІКАЦІЯ ЗА ЕМОЦІЙНИМ СТАНОМ ДИКТОРА.
- ПРЕДМЕТ ДОСЛІДЖЕННЯ Є МАТЕМАТИЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ СИСТЕМИ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЗА АУДІОЗАПИСАМИ І ОБРОБКА АУДІОСИГНАЛІВ, МЕТОДИ КЛАСИФІКАЦІЇ, ВЕКТОРИЗАЦІЇ ТА ОПТИМІЗАЦІЇ ГІПЕРПАРАМЕТРІВ
- ОБ'ЄКТОМ ДОСЛІДЖЕННЯ Є СПОСОБИ ОРГАНІЗАЦІЇ ЗАСОБІВ МАШИННОГО НАВЧАННЯ ДЛЯ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЗА АУДІОЗАПИСАМИ.
- ЗАВДАННЯМ Є РОЗРОБКА СИСТЕМИ ДЛЯ ВИРІШЕННЯ ЗАДАЧІ КЛАСИФІКАЦІЇ, ІЗ ВИКОРИСТАННЯМ МЕТОДІВ ОБРОБКИ АУДІОСИГНАЛІВ ДЛЯ ПРЕДСТАВЛЕННЯ ОБРОБЛЕНИХ ФАЙЛІВ У ВИГЛЯДІ ЧИСЛОВИХ ВЕКТОРІВ ТА МЕТОДІВ РЕГУЛЯРИЗАЦІЇ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПІДБОРУ ПРОДУКТИВНІШОЇ КОМБІНАЦІЇ ЗНАЧЕНЬ ГІПЕРПАРАМЕТРІВ, З ПОДАЛЬШОЮ КЛАСИФІКАЦІЄЮ ВСІХ ОТРИМАНИХ СТРУКТУРОВАНИХ ДАНИХ.
- КІНЦЕВИМ РЕЗУЛЬТАТОМ Є МОДЕЛЬ КЛАСИФІКАЦІЇ ЕМОЦІЙ ЗА АУДІОЗАПИСОМ.

3

Рисунок Б.3 – Слайд 3

БІЗНЕС ПРОБЛЕМА



4

Рисунок Б.4 – Слайд 4

ВИМОГИ ДО СИСТЕМ

- ЗДАТНІСТЬ СИСТЕМИ ОБРОБЛЯТИ СИРІ ДАНІ ТА ТРАНСФОРМУВАТИ ДАНІ У ФОРМАТ ЗРУЧНИЙ ДЛЯ РОЗПІЗНАВАННЯ
- НАЯВНІСТЬ АЛГОРИТМУ МАШИННОГО НАВЧАННЯ, ЩО ВИРІШУЄ ЗАДАЧУ КЛАСИФІКАЦІЇ АУДІОЗАПИСІВ ЗА ЕМОЦІЯМИ;
- ЗДАТНІСТЬ СИСТЕМИ КЛАСИФІКУВАТИ ЗВУКОВІ ЗАПИСИ ЗА НАСТУПНИМИ КЛАСАМИ: СПОКІЙ, РАДІСТЬ, СУМ, ЗЛІСТЬ, СТРАХ, ЗДИВУВАННЯ ТА ОГИДА.
- РОЗРОБЛЕНИЙ ПРОГРАМНИЙ ПРОДУКТ В КІНЦЕВОМУ ВИГЛЯДІ ПОВИНЕН БУТИ ПРЕДСТАВЛЕНИЙ У ВИДІ ВІКНА З МОЖЛИВІСТЮ ЗАВАНТАЖЕННЯ АУДІОЗАПИСІВ;
- ЗДАТНІСТЬ СИСТЕМИ ПРОВОДИТИ ПРОЦЕДУРУ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ ДОКУМЕНТІВ З ТОЧНІСТЮ НЕ НИЖЧЕ 80%;
- ДОСТУП ДО СИСТЕМИ НАДАВАТИ ЧЕРЕЗ WEB-ІНТЕРФЕЙС.

5

Рисунок Б.5 – Слайд 5

ВИБІР МЕТОДУ КЛАСИФІКАЦІЙ

Decision Tree

Decision Tree алгоритм, який зосереджується на створенні класифікації у формі дерев рішень, отриманих із набору неупорядкованих і нерегулярних екземплярів. За допомогою підходу визначення об'єкту через попереднє значення, такий алгоритм зверху вниз оцінює атрибути у вузлах дерева, гілки на основі окремих атрибутів вузла та отримує висновки з листових вузлів.

MLP

У контексті MLP нейрони використовують нелінійні функції активації, призначені для імітації функціональності нейронів людського мозку. І навпаки, у багатошаровому перцептроні кожен нейрон використовує функцію лінійної активації, а процес навчання включає зворотне поширення.

CNN

Згорткові нейронні мережі (CNN) – це тип штучної мережі прямого зв'язку, яка черпає натхнення із зорової кори тварин. CNN є формою регуляризованих багатошарових перцептронів, де регуляризовані версії використовуються для пом'якшення надмірного оснащення.

8

Рисунок Б.8 – Слайд 8

ЕКСПЕРЕМЕНТАЛЬНІ РЕЗУЛЬТАТИ(DECISION TREE)

	precision	recall	f1-score	support
0	0.80	0.79	0.79	2264
1	0.67	0.66	0.67	172
2	0.62	0.59	0.61	1910
3	0.62	0.61	0.61	1878
4	0.62	0.64	0.63	1829
5	0.64	0.65	0.65	1612
6	0.68	0.68	0.68	1820
7	0.65	0.70	0.67	197
micro avg	0.67	0.67	0.67	11682
macro avg	0.66	0.67	0.66	11682
weighted avg	0.67	0.67	0.67	11682
samples avg	0.67	0.67	0.67	11682

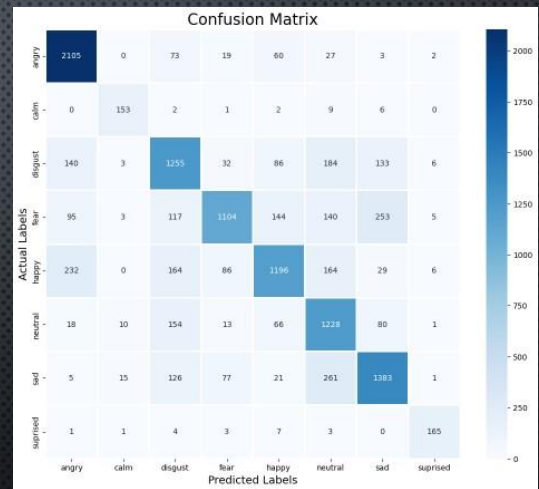
Confusion Matrix

	0	1	2	3	4	5	6	7
0	1794	3	169	103	171	50	26	32
1	8	183	28	1	4	14	10	5
2	187	8	1192	155	177	182	150	16
3	132	9	143	1146	172	132	186	16
4	143	7	148	170	1175	137	74	36
5	45	14	116	123	85	1053	133	1
6	26	11	114	111	11	189	1196	6
7	12	5	9	6	21	1	5	117

Рисунок Б.9 – Слайд 9

ЕКСПЕРЕМЕНТАЛЬНІ РЕЗУЛЬТАТИ (MLP)

	precision	recall	f1-score	support
angry	0.81	0.92	0.86	2289
calm	0.83	0.88	0.85	173
disgust	0.66	0.68	0.67	1839
fear	0.83	0.59	0.69	1861
happy	0.76	0.64	0.69	1877
neutral	0.61	0.78	0.68	1570
sad	0.73	0.73	0.73	1889
surprised	0.89	0.90	0.89	184
accuracy			0.74	11682
macro avg	0.76	0.77	0.76	11682
weighted avg	0.74	0.74	0.73	11682

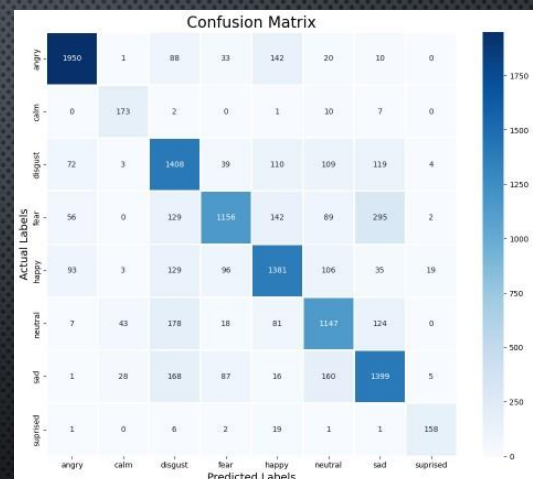


10

Рисунок Б.10 – Слайд 10

ЕКСПЕРЕМЕНТАЛЬНІ РЕЗУЛЬТАТИ (CNN)

	precision	recall	f1-score	support
angry	0.89	0.87	0.88	2244
calm	0.69	0.90	0.78	193
disgust	0.67	0.76	0.71	1864
fear	0.81	0.62	0.70	1869
happy	0.73	0.74	0.74	1862
neutral	0.70	0.72	0.71	1598
sad	0.70	0.75	0.73	1864
surprised	0.84	0.84	0.84	188
accuracy			0.75	11682
macro avg	0.75	0.77	0.76	11682
weighted avg	0.76	0.75	0.75	11682

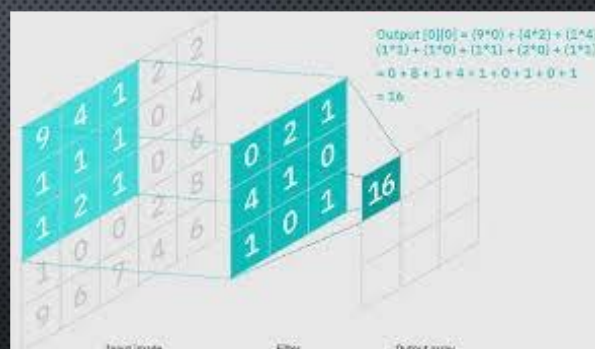


11

Рисунок Б.11 – Слайд 11

МЕТОД

- CNN стали високоефективними моделями глибокого навчання, досягаючи вражаючих результатів у різних областях досліджень, таких як розпізнавання об'єктів, розпізнавання облич, розпізнавання рукописного тексту, розпізнавання мови та обробка природної мови.
- CNN використовують операції згортки замість простого множення для обробки вхідних даних у згорткових шарах.
- Цей дизайн натхненний поведінкою нейронів зорової кори.



12

Рисунок Б.12 – Слайд 12

ГІПЕРПАРАМЕТРИ

- Налаштування гіперпараметрів полягає в пошуку набору оптимальних значень гіперпараметрів для алгоритму навчання під час застосування цього оптимізованого алгоритму до будь-якого набору даних. Ця комбінація гіперпараметрів максимізує продуктивність моделі, мінімізуючи попередньо визначену функцію втрат, щоб отримати кращі результати з меншою кількістю помилок. Алгоритм навчання оптимізує втрати на основі вхідних даних і намагається знайти оптимальне рішення в межах заданого налаштування.

■ Кількість прихованих шарів

■ Функція активації

■ Метод оптимізації

13

Рисунок Б.13 – Слайд 13

ВИБІР МЕТОДУ ПІДБОРУ ГІПЕРПАРАМЕТРІВ

Випадковий пошук

Випадковий пошук намагається використовувати випадкову комбінацію гіперпараметрів у кожній ітерації та записує продуктивність моделі. Після кількох ітерацій він повертає набір, яка дала найкращий результат.

Баєсівська оптимізація

Алгоритм спостерігає за результатами, які генеруються. На основі цих спостережень оптимізатор використовує середнє цього Гаусового процесу в якості наближення функції.

Hyperband

Гени формують індивіда в популяції. Створюється пошуковий простір, в якому накопичуються всі особи. Усі особи закодовані в межах кінцевої довжини в просторі пошуку. Кожній особині в просторі пошуку (популяції) присвоюється оцінка придатності, яка говорить про її здатність конкурувати з іншими особинами.

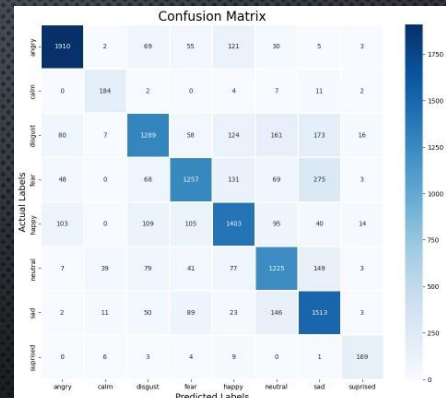
14

Рисунок Б.14 – Слайд 14

ЕКСПЕРЕМЕНТАЛЬНІ РЕЗУЛЬТАТИ(ВИПАДКОВИЙ ПОШУК)

- Вхідний шар: 256
- Приховані шари: 128, 128, 128
- Функція активації: sigmoid
- Оптимізатор: adam

	precision	recall	f1-score	support
angry	0.89	0.87	0.88	2195
calm	0.74	0.88	0.80	210
disgust	0.77	0.68	0.72	1908
fear	0.78	0.68	0.73	1851
happy	0.74	0.75	0.75	1869
neutral	0.71	0.76	0.73	1620
sad	0.70	0.82	0.76	1837
surprised	0.79	0.88	0.83	192
accuracy			0.77	11682
macro avg	0.77	0.79	0.77	11682
weighted avg	0.77	0.77	0.77	11682



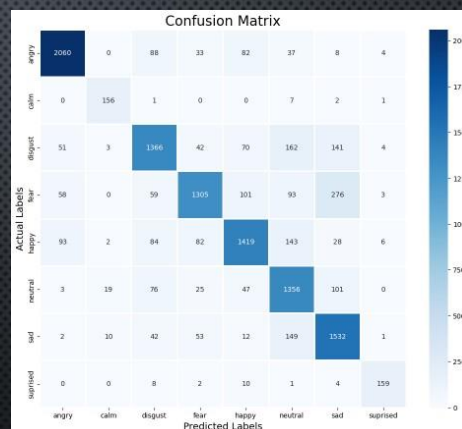
15

Рисунок Б.15 – Слайд 15

ЕКСПЕРЕМЕНТАЛЬНІ РЕЗУЛЬТАТИ (БАЄСІВСЬКА ОПТИМІЗАЦІЯ)

- Вхідний шар: 256
- Приховані шари: 128, 128, 64
- Функція активації: relu
- Оптимізатор: adam

	precision	recall	f1-score	support
angry	0.91	0.89	0.90	2312
calm	0.82	0.93	0.87	167
disgust	0.79	0.74	0.77	1839
fear	0.85	0.69	0.76	1895
happy	0.82	0.76	0.79	1857
neutral	0.70	0.83	0.76	1627
sad	0.73	0.85	0.79	1801
surprised	0.89	0.86	0.88	184
accuracy			0.80	11682
macro avg	0.81	0.82	0.81	11682
weighted avg	0.81	0.80	0.80	11682



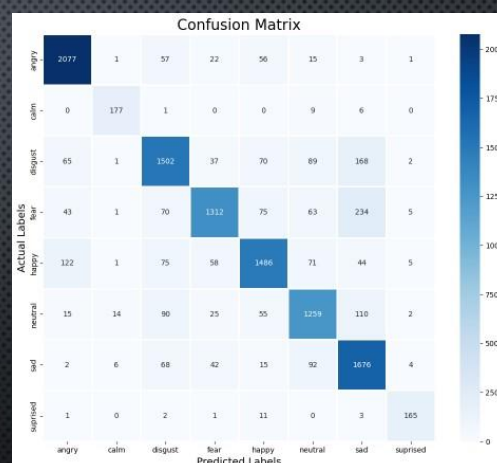
16

Рисунок Б.16 – Слайд 16

ЕКСПЕРЕМЕНТАЛЬНІ РЕЗУЛЬТАТИ (HYPERBAND)

- Вхідний шар: 256
- Приховані шари: 128, 256, 64
- Функція активації: relu
- Оптимізатор: adam

	precision	recall	f1-score	support
angry	0.89	0.93	0.91	2232
calm	0.88	0.92	0.90	193
disgust	0.81	0.78	0.79	1934
fear	0.88	0.73	0.80	1803
happy	0.84	0.80	0.82	1862
neutral	0.79	0.80	0.79	1570
sad	0.75	0.88	0.81	1905
surprised	0.90	0.90	0.90	183
accuracy			0.83	11682
macro avg	0.84	0.84	0.84	11682
weighted avg	0.83	0.83	0.83	11682



17

Рисунок Б.17 – Слайд 17

НАЙТОЧНІШІ ПОКАЗНИКИ БУЛИ ОТРИМАНІ ПРИ ВИКОРИСТАННІ НЕЙРОМЕРЕЖЕВОГО ПІДХОДУ
З ВИКОРИСТАННЯМ CONVOLUTIONAL NEURAL NETWORK
ТА МЕТОДОМ ПІДБОРІ ГІПЕРПАРАМЕТРІВ HYPERBAND

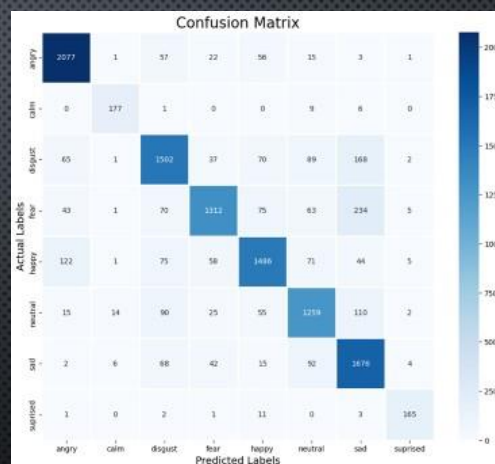
- Вхідний шар: 256
- Приховані шари: 128, 256, 64
- Функція активації: relu
- Оптимізатор: adam
- Precision: 84%
- Recall: 84%
- F1 - score: 84%

18

Рисунок Б.18 – Слайд 18

РЕЗУЛЬТАТ РОЗПІЗНАВАННЯ НА ТЕСТОВІЙ ВИБІРЦІ

	Predicted Labels	Actual Labels
0	disgust	disgust
1	happy	happy
2	disgust	disgust
3	disgust	disgust
4	angry	angry
...	...	...
11773	happy	disgust
11774	angry	angry
11775	disgust	fear
11776	angry	angry
11777	angry	angry



19

Рисунок Б.19 – Слайд 19

РЕЗУЛЬТАТ РОБОТИ ДЕМОНСТАЦІЙНОЇ ВЕРСІЙ СИСТЕМИ



20

Рисунок Б.20 – Слайд 20

РЕЗУЛЬТАТ РОБОТИ ДЕМОНСТАЦІЙНОЇ ВЕРСІЙ СИСТЕМИ



21

Рисунок Б.21 – Слайд 21

ВИСНОВКИ

- **ДОСЛІДЖЕНО**, наявні застосунки та теоретичні методи розпізнавання емоцій за аудіозаписами і розроблено застосунок, який виконує такий функціонал.
- **ВСТАНОВЛЕНО**, що використання такої системи допомагає краще розуміти емоції людини, а також зберігає ресурси, через автоматизацію цього процесу.
- **ПРОВЕДЕНО**, порівняльний аналіз підходів класифікації та методів оптимізації.
- Модель була навчена на комбінації декількох наборів даних та здатна розрізняти класи емоцій: спокій, радість, сум, злість, страх, здивування та огиду.
- Подальші перспективи в цьому напрямку полягають в випробуванні нових більш сучасних алгоритмів та застосування їх для підвищення точності розпізнавання емоцій.

22

Рисунок Б.22 – Слайд 22

ДЯКУЮ ЗА УВАГУ!

23

Рисунок Б.23 – Слайд 23

