МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

БАСИСТЮК ОЛЕГ АНДРІЙОВИЧ

На правах рукопису

УДК 004.652

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ МУЛЬТИМОДАЛЬНИХ ДАНИХ НА ОСНОВІ АНСАМБЛЮ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ

122 – комп’ютерні науки

Дисертація на здобуття наукового ступеня
доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____/О.А.Басистюк/

Науковий керівник

Мельникова Наталія Іванівна

доктор технічних наук, професор

Львів – 2025

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

МС– мультимодальна система.

МД – модальні дані.

ІС – інформаційна система.

ПЗ – програмне забезпечення.

ПП – програмний продукт.

БД – база даних.

GPU – графічний процесор.

CPU – центральний процесор.

RAM – оперативна пам'ять.

RNN – рекурентні нейронні мережі.

АНОТАЦІЯ

Басистюк О.А. Інформаційна технологія аналізу мультимодальних даних на основі ансамблю моделей машинного навчання. – Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 “Комп’ютерні науки”. – Національний університет «Львівська політехніка», Львів, 2025.

Зміст анотації. Дисертація має на меті розробити інформаційну технологію опрацювання мультимодальних даних на основі ансамблю моделей машинного навчання. Дослідження фокусується на розробці методів та засобів штучного інтелекту для попереднього опрацювання, кодування та синхронізації даних різних модальностей (мультимодальних) у єдину структуру даних з подальшим ухваленням остаточного рішення на основі узагальненої репрезентації різно модальних даних. Особливістю цього дослідження є обробка і тренування моделей машинного навчання на українських наборах даних, оскільки ця проблематика є недостатньо дослідженою на сьогодні. Розроблена система, яка стане результатом цієї дисертації, ідеально підходить для завдань, де вимагається висока надійність і ефективність обробки мультимодальних даних, наприклад, у медичній сфері, виробництві, інформаційній безпеці та електронній торгівлі.

У першому розділі «Аналітичний огляд літературних джерел» здійснено аналіз технології, які використовуються для роботи з модальними даними, описані математичні методи, які лежать в основі опрацювання, проаналізовано та описано, що саме собою представляють мультимодальні дані. На підставі проведеного аналізу відомих рішень встановлено, що існуючі технології опрацювання модальних даних, недостатньо пристосовані для опрацювання українських даних, та не можуть забезпечити прийнятного часу їх опрацювання. Результатом даного розділу є постановка задачі, а саме створення інформаційної технології аналізу модальних даних на основі ансамблю моделей машинного навчання.

У другому розділі «Розроблення ансамблю моделей машинного навчання» було здійснено аналіз ансамблів моделей машинного навчання, які придатні й ефективні при опрацюванні даних різних модальностей. Представлено формальний опис структури моделі даних. Описано основні етапи підготовки мультимодальних даних, особливості архітектурного вирішення ансамблю моделей машинного навчання, які використовуватимуться для аналізу модальними даними, а саме текстових, аудіо, відео та мета даних.

У третьому розділі «Розроблення методу аналізу модальних даних» було розроблено концептуальну модель аналізу мультимодальних даних, алгоритм зведення даних різних модальностей до єдиного вигляду, алгоритми обробки модальних даних, а також метод перевірки якості даних. Усі алгоритми було зображено за допомогою блок-схем.

У четвертому розділі «Розроблення архітектури інформаційної технології та апробація результатів» представлено архітектурне вирішення системи, обґрунтовано вибір технологій для реалізації системи. Також визначено функціонал системи та засоби, які забезпечать стабільну роботу системи. Розроблено інтерфейсні рішення. Здійснено аналіз та апробацію отриманих результатів.

Основні наукові результати дисертації опубліковано в 13 працях, зокрема: 3 статті – у виданнях, що індексуються в наукометричних базах даних (дві статі в журналі Q1, одна стаття в журналі Q2), 5 статей – у фахових виданнях України, 5 публікації – у збірниках наукових праць конференцій.

Ключові слова: модальні дані, машинне навчання, глибинне навчання, база даних, цілісність даних.

ABSTRACT

Information technology of multimodal data analysis based on an ensemble of machine learning models. - Qualifying scientific work on the rights of a manuscript. Dissertation for the degree of Doctor of Philosophy in specialty 122 “Computer Science.” - Lviv Polytechnic National University, Lviv, 2025.

Abstract content. The dissertation aims to develop an information technology for processing multimodal data based on an ensemble of machine learning models. The study focuses on the development of methods and tools of artificial intelligence for pre-processing, encoding and synchronizing data of different modalities (multimodal) into a single data structure with further final decision-making based on a generalized representation of multimodal data. The peculiarity of this research is the processing and training of machine learning models on Ukrainian datasets, as this issue is not sufficiently studied today. The developed system, which will be the result of this thesis, is ideally suited for tasks requiring high reliability and efficiency of multimodal data processing, such as in the medical field, manufacturing, information security, and e-commerce.

The first chapter, “Analytical Review of Literature,” analyzes the technologies used to work with modal data, describes the mathematical methods underlying the processing, and analyzes and describes what exactly multimodal data is. Based on the analysis of known solutions, it is established that existing technologies for processing modal data are not sufficiently adapted to Ukrainian data and cannot provide acceptable processing time. The result of this section is the statement of the problem, namely the creation of an information technology for analyzing modal data based on an ensemble of machine learning models.

In the second section, “Development of an ensemble of machine learning models”, we analyze ensembles of machine learning models that are suitable and effective in processing data of different modalities. A formal description of the data model structure is presented. The main stages of multimodal data preparation, features of the architectural solution of the ensemble of machine learning models that

will be used for the analysis of modal data, namely text, audio, video, and meta data, are described.

In the third chapter, “Development of a Method for Analyzing Modal Data,” a conceptual model for analyzing multimodal data, an algorithm for reducing data of different modalities to a single form, algorithms for processing modal data, and a method for checking data quality were developed. All algorithms were depicted using flowcharts.

In the fourth chapter, “Development of the Information Technology Architecture and Testing of the Results,” presents the architectural solution of the system, justifies the choice of technologies for the system implementation. The system functionality and tools that will ensure stable operation of the system are also defined. Interface solutions were developed. The results are analyzed and tested.

The main scientific results of the dissertation are published in 13 papers, in particular: 3 articles - in publications indexed in scientometric databases (two articles in the journal Q1, one article in the journal Q2), 5 articles - in professional publications of Ukraine, 5 publications - in conference proceedings.

Keywords: modal data, machine learning, deep learning, database, data integrity.

СПИСОК ПУБЛІКАЦІЙ

Статті в наукометричних базах даних:

1. Nykoniuk M., Basystiuk O., Shahovska N, Melnykova N. Multimodal Data Fusion for Depression Detection Approach // Computation. – 2024. – Vol. 16, iss. 24. – P. 1–15. (Q2)
2. Zanevych Y., Yovbak V., Basystiuk O., Fedushko S., Shahovska N., Argyroudis S. Evaluation of pothole detection performance using deep learning models under low-light conditions // Sustainability. – 2024. – Vol. 16, iss. 24. – P. 1–15. (Q1)
3. Kohut N., Basystiuk O., Melnykova N., Shahovska N. Detecting Plant Diseases Using Machine Learning Models // Sustainability. – 2024. – Vol. 16, iss. 24. – P. 1–15. (Q1)

Статті у фахових виданнях України

1. Басистюк О. А., Мельникова Н. І. Мультимодальне розпізнавання мовлення на основі звукових і текстових даних // Вісник Хмельницького національного університету. Серія: Технічні науки. – 2022. – № 5 (313). – С. 22–25.
2. Basystiuk O., Melnykova N. Development of the multimodal handling interface based on Google API // Комп'ютерні системи проектування. Теорія і практика. – 2024. – Вип. 6, № 1. – С. 216–223
3. Басистюк, Наталія Мельникова, Ірина Думин. Розроблення мультимодального інтерфейсу на основі Google API // Вісник Хмельницького національного університету. Серія: Технічні науки. – 2024. – № 3 (335). – С. 15–18.
4. Basystiuk O., Rybchak Z. Recommendation systems in e-commerce applications // Вісник Національного університету “Львівська політехніка”. Серія: Інформаційні системи та мережі. – 2024. – Вип. 15. – С. 252–259.

5. Олег Басистюк, Наталія Мельникова, Ірина Думин, Андрій Думин. Ансамблевий підхід у мультимодальній обробці даних на основі Google API // Вісник Хмельницького національного університету. Серія: Технічні науки. – 2024. – № 4 (339). – С. 235–238

Конференції

1. Basystiuk, O., Melnykova, N. (2023, October). Multimodal Learning Analytics: An Overview of the Data Collection Methodology. In 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT) (pp. 1-4). IEEE.
2. Basystiuk O., Melnykova N. Multimodal medical data learning approaches for digital healthcare // CEUR Workshop Proceedings. – 2024. – Vol. 3609: Proceedings of the 6th International conference on informatics & data-driven medicine, Bratislava, Slovakia, November 17-19, 2023. (IDDM 2023) (pp. 332–337).
3. Basystiuk, O., Farmaha I., Rybchak, Z. (2023, February). Exploring Multimodal Data Approach in Natural Language Processing Based on Speech Recognition Algorithms. In 2023, the 17th International Conference on the Experience of Designing and Application of CAD Systems (CADSM) (pp. 1-3). IEEE.
4. Basystiuk, O., Melnykova, N. (2023, December). Multimodal Approaches for Natural Language Processing in Medical Data. In 5th International Conference on Informatics & Data-Driven Medicine (IDDM 2022) (pp. 246-252). IEEE.
5. Kobylyukh, L., Rybchak Z., Basystiuk, O. (2023, April). Analyzing the Accuracy of Speech-to-Text APIs in Transcribing the Ukrainian Language. In The 7th International Conference Computational Linguistics and Intelligent Systems» (CoLInS 2023) (pp. 217-227).

Вступ

Актуальність теми. Стрімкий перехід до епохи великих даних зумовив різке зростання обсягів інформації різних модальностей, що, у свою чергу, породило потребу у пошуку ефективних методів та засобів для їх опрацювання. У цьому контексті технології опрацювання мультимодальних даних стають одними з найактуальніших напрямів на ринку інформаційних технологій. Їхні можливості знаходять широке застосування у прикладних сферах — від систем підтримки користувачів і консультаційних сервісів до спеціалізованих галузей, таких як медицина, торгівля та фінанси. Результати та нові розробки широко обговорюються на конференціях і наукових заходах усіх рівнів.

Під час роботи з мультимодальними даними через походження даних з різних сенсорів, систем та джерел виникає потреба в узгодженні та синхронізації потоків даних, для подальшого опрацювання даних та отримання високоточних результатів. Додатково варто зауважити на доцільності використання ансамблевих підходів для моделей машинного навчання при опрацюванні різнотипових мультимодальних наборів даних. Оскільки, це своєю чергою дозволить створити більш стійкі та комплексні моделі, які зможуть забезпечувати достатній рівень точності при опрацюванні мультимодальних наборів даних.

Насамперед слід наголосити на перевагах, які надають методи опрацювання мультимодальних наборів даних, а саме вони надають можливість аналізувати проблематику більш комплексно та отримувати узагальнені результати з урахуванням особливостей кожної опрацьованої модальності. Первинно, у технологіях опрацювання мультимодальних даних передбачається інтеграція інформації з різних модальностей, очищення від аномалій та надлишковості, а також приведення до єдиного формату та синхронізацію за часовою ознакою. Для опрацювання кожної з модальностей мультимодального набору, доцільно використовувати окрему модель, яка пристосована до особливостей конкретного типу даних, що дозволить досягнути кращих

результатів точності, проте породить проблему агрегації та узагальнення отриманих результатів. Саме в модулі злиття та агрегації отриманих результатів, доцільно застосовувати ансамблеві підходи опрацювання результатів унімодальних моделей машинного навчання, оскільки це дозволить підвищити точність і стійкість отриманих результатів. В основному, це можна досягнути шляхом об'єднання сильних сторін окремих моделей, яке забезпечить усереднення метрик помилок, а також зменшить ризик перенавчання, яке може виникати в окремих унімодальних моделях.

Додатково, варто врахувати, що ансамблевий підхід сприяє кращій адаптації до неоднорідності вхідних даних, що дозволить модулю злиття ефективніше обробляти складні залежності між даними. В результаті, це забезпечує більш якісне узагальнення інформації та прийняття рішень на основі аналізу даних.

Технології, які здатні забезпечити опрацювання мультимодальних даних та ефективно поєднувати результати моделей унімодальної обробки даних, можна використовувати в різних сферах, проте найбільш затребуваною дана технологія буде у сфері комунікацій та взаємодії з користувачами. Система мультимодального опрацювання даних дозволить покращить розуміння контексту повідомлень, дозволить забезпечити більш персоналізовані відповіді, оптимізує та автоматизує процеси управління інформацією, оскільки дозволить об’єднати різні джерела даних для підвищення продуктивності та ефективності комунікації.

За допомогою технології опрацювання мультимодальних даних сфера торгівлі дозволяє створювати інтелектуальні системи рекомендацій, які враховують не лише текстовий опис товарів, але і їхні зображення та відгуки у відеоформаті. Для таких галузей, як медицина, технологія надасть можливість покращити якість передбачення ризиків захворювань, оскільки створить можливість враховувати параметри й особливості мультисенсорних досліджень, та зводити отримані дані до єдиного узагальненого формату.

Отже, задача розроблення інформаційної технології опрацювання мультимодальних даних є актуальною, особливо для підпроблеми україномовних наборів даних.

Зв'язок роботи з науковими програмами, планами і темами. Дисертація була виконана згідно з пріоритетними напрямками наукових досліджень Національного університету "Львівська політехніка", відповідно до координаційних планів Міністерства освіти і науки України. Зокрема, вона вписувалася в рамки наукових досліджень, які здійснювалися відповідно до державного фінансування на кафедрі систем штучного інтелекту «Технології опрацювання мультимодальних україномовних даних для визначення рівня стресу» (№ держ. реєстру **0123U100231**), в міжнародному науковому дослідженні, що включено до звіту за договором «Комплексний догляд для наступного покоління» (**C2020/1-8, iCare4Next**) за договором **№ M/100-2024** від 19.07.2024 р, а також впровадженно у навчальний процес при викладанні дисциплін «Машинне навчання» та «Обробка й аналіз цифрових сигналів» для студентів освітньо-кваліфікаційного рівня «бакалавр», що навчаються за напрямом 122 "Комп’ютерні науки", на кафедрі систем штучного інтелекту.

Метою дослідження є підвищення якості аналізу мультимодальних даних, шляхом створення інформації технології, яка забезпечить цілісність зберігання, швидкий доступ та точне узагальнення інформації для подальшого аналізу даних на основі ансамблю методів машинного навчання.

Для досягнення поставленої мети в роботі потрібно розв’язати такі задачі:

1. Провести аналіз сучасного стану інформаційних технологій аналізу мультимодальних даних;
2. Розробити модель мультимодальних даних;
3. Формалізувати конвеєр опрацювання мультимодальних даних;
4. Дослідити найефективніший спосіб ансамблювання моделей машинного навчання для комплексного аналізу даних різних модальностей;

5. Розробити інформаційну технологію опрацювання мультимодальних даних та апробувати результати дослідження.

Об'єкт дослідження – процеси опрацювання та аналізу мультимодальних даних.

Предмет дослідження – методи машинного навчання для обробки та аналізу мультимодальних даних.

Методи дослідження. У процесі розробки застосовувалися моделі зберігання та передачі мультимодальних даних, методи машинного навчання для опрацювання даних різних модальностей. Методи узгодження і синхронізації даних, використано математичну статистику та теорію множин. Принципи ансамблевих Для розробки алгоритмів запису у базі даних використано теорію інформації та алгоритмів. Для побудови методу перевірки якості введених даних використано теорію кодування.

Наукова новизна одержаних результатів полягає в тому що:

– **вперше** розроблено модель опрацювання унімодальних та мультимодальних даних на основі ансамблевих методів шляхом визначення вагомості кожної з модальностей, що підвищило генералізацію результатів незалежно від кількості опрацьованих модальностей;

– **отримав подальший розвиток** метод синхронізації мультимодальних даних шляхом застосування множини векторизаторів та механізмів уваги, що дало змогу опрацьовувати модальності з різними довжинами та узгоджувати їх між собою

– **вперше** розроблено рекомендаційну систему щодо зберігання модальностей у різних форматах даних, що дало змогу збільшити швидкодію опрацювання таких даних

– **удосконалено** метод попереднього опрацювання україномовних мультимодальних даних шляхом визначення аномальних даних, що дало змогу підвищити точність векторизації даних

Практичне значення одержаних результатів.

1. Розроблені алгоритми оптимізовані для роботи з україномовними мультимодальними даними, що забезпечує вищу точність обробки інформації, усереднений приріст точності класифікаційних задач на україномовних вибірках становить 4%, порівняно з узагальненими міжнародними моделями.
2. Проведено порівняльний аналіз методів кодування та декодування для україномовних наборів мультимодальних даних, які базуються на перетворенні даних із різними структурами в числові масиви унікальних ознак. Попередня обробка кожної модальності, а також застосування методу виявлення аномалій дозволила підвищити достовірність отриманих результатів до понад 0.8.
3. На основі досліджень різнотипових змішувань у методі синхронізації даних, розроблено алгоритм, який передбачає пізнє змішування модальних даних. Таким чином, маючи різні типи мультимодальних даних, у векторній репрезентації, можна їх узгодити за часовим критерієм (синхронізувати), для підвищення комплексності отриманих знань і точності на подальших етапах опрацювання даних. Експериментально визначено, що застосування моделі пізнього змішування даних дозволяє досягнути точності на рівні 0.969, в порівнянні з раннім змішування 0.94 та без застосування змішування 0.912.

Особистий внесок здобувача. Основні положення та результати дисертаційної роботи були отримані автором самостійно. Докторанту належать наступні наукові досягнення: розроблено метод запису мультимодальних даних у базі даних для підготовки подальшого опрацювання [1], розроблено метод перевірки та попередньої обробки даних [2], розроблено модель опрацювання даних ансамблевим методом на основі моделей машинного навчання [3], розроблено метод інтеграції даних [5, 6], розроблено архітектуру даних для пропонованої системи [7, 8].

Апробація результатів дисертації. У результаті дисертації розроблена і запроваджена інформаційну технологію для опрацювання мультимодальних даних. Ця робота являти собою дослідження методів опрацювання і перетворення різнотипових даних у єдиний вектор унікальних ознак, з подальшим опарцюванням їх в ансамблем моделей машинного навчання, для сумаризації і ухвалення рішень на основі фінальної мета моделі. Роботу апробовано на Міжнародній конференції з комп'ютерних наук та інформаційних технологій (CSIT 2024), 5-ій Міжнародній конференції з інформатики та керованої даними медицини (IDDM 2022), 17-ій Міжнародній конференція з проектування та застосування систем автоматизованого проектування (CADSM2024), 7-ій Міжнародній конференції «Обчислювальна лінгвістика та інтелектуальні системи» (CoLInS 2023).

Публікації. За результатами виконаних досліджень опубліковано 13 наукових праць, із них 3 статей – у виданнях, що індексуються в наукометричних базах даних (дві статті в журналах Q1, дві статті в журналах Q2), 5 статті – у фахових виданнях України, 5 публікації – у збірниках наукових праць конференцій.

Структура і обсяг дисертації. Дисертаційна робота викладена на 120 сторінках та складається зі змісту, переліку скорочень, вступу, чотирьох основних розділів, в яких містяться 33 рисунків, списку використаних джерел із 129 найменувань та додатків.

РОЗДІЛ 1. АНАЛІТИЧНИЙ ОГЛЯД ЛІТЕРАТРУРИ

У розділі здійснено аналіз технологій, які використовуються для опрацювання модальних даних, описані математичні методи, які використовуються при перетвореннях, проаналізовано та описано зміст підходів до опрацювання мультимодальних даних.

Матеріали розділу опубліковані у роботах автора [32, 33, 34].

1.1. Аналіз систем опрацювання модальних даних

Мультимодальні дані - це набори інформації, що містять різні типи даних, такі як текст, зображення, аудіо, відео і т. д. У контексті обробки природної мови, мультимодальні дані можуть включати текстові описи, зображення, відео або аудіозаписи, які стосуються одного або кількох об'єктів аналізу [1]. Обробка таких даних є складною через різноманітність форматів та характеристик різних типів даних, що вимагає спеціальних підходів і методів для ефективного аналізу та використання [2].

Штучний інтелект стрімко поширюється та дедалі глибше інтегрується в наше повсякденне життя. Однією з найпопулярніших технологій у цій сфері є розпізнавання мови, яке є складовою концепції мультимодальних даних і охоплює роботу з відео, аудіо та текстовою інформацією.

Сам процес навчання можна уявити як аналогію до реалізації вже відомих моделей, наприклад, створення дерева рішень, але виконується це самостійно. Після завершення навчання якість розпізнавання перевіряється за допомогою нових наборів даних [3]. Штучний інтелект зазвичай використовується для спрощення та автоматизації різних процесів, таких як обробка природної мови, автопілотування автомобілів, розпізнавання обличчя тощо. Штучний інтелект активно застосовується в різних галузях, допомагаючи створювати інновації та оптимізувати процеси, іноді навіть повністю замінюючи людську працю.

Системи машинного перекладу тексту з однієї мови на іншу імітують роботу перекладача. Однак ефективність таких систем залежить від їх здатності розуміти граматичні правила мови та контекст специфічної галузі [4]. Оскільки кожна галузь має власну термінологію та обмеження, її ефективність залежить від здатності мови розуміти граматику, помічати закономірності та часто співвідносити з цільовою галуззю. Основними одиницями аналізу при перетворенні аудіо в текст є цілі речення або фразеологізми, які пояснюють загальну ідею процесу, а не окреме слово [5]. Їх ефективність залежить від здатності розуміти граматичні правила мови. У перекладі основними одиницями є не окремі слова, а словосполучення або фразеологізми, які виражають різні поняття. Тільки використовуючи їх, можна виразити складніші ідеї за допомогою перекладеного тексту.

Ключовою особливістю машинного перекладу є те, що довжина вхідних і вихідних даних може відрізнятися. Роботу з такими змінними послідовностями забезпечують рекурентні нейронні мережі [6]. RNN є широко використовуваним методом для обробки послідовностей слів як у машинному перекладі, так і в інших завданнях обробки природної мови [7].

З основними компонент мультимодальних даних, джерелом, типами, а також кодування даних можна ознайомитися на таблиці 1.1.

Таблиця 1.1 – Джерело, тип і кодування вхідних даних.

Джерело	Тип	Кодування
Опис події	Текст	UTF-8
Фото з місця	Зображення	JPG, PNG
Коментар очевидців	Звук	WAV
Відео з місця	Відео	AVI

Дане дослідження спрямоване на інтеграцію різних типів модальних даних, таких як текст, зображення, аудіо та відео, для створення комплексних аналітичних рішень [8]. У рамках дослідження буде проаналізовано переваги та

обмеження різних методів інтеграції модальних даних, детальніше ознайомитися зі розподілом компонентів корпусу мультимодальних даних можна на рисунку 1.1.

Рисунок 1.1 – Загальна схема корпусу мультимодальних даних

Додатково, в рамках даної дисертаційної роботи, було проведено порівняльний аналіз наявних підходів з метою вибору найбільш досконалих та надійних методів і технологій для побудови надійних мультимодальних систем перетворення аудіо в текст та перетворення україномовних даних в уніфіковані вектори даних [9]. Це дозволяє краще зрозуміти поточний стан систем обробки модальних даних для української мови й визначити перспективні напрямки для подальших досліджень. Результати дослідження в майбутньому можуть бути використані для створення широкого спектру моделей перетворення мовлення в текст для конкретних галузей, що дозволить зменшити підвищити ефективність використання робочого часу працівниками [10]. З узагальненою структурою системи опрацювання мультимодальних даних можна ознайомитися на рисунку 1.2.

Технології перетворення природних мов в різні типи даних, а саме текст, аудіо, відео, набуває все більшого значення у світі з розвитком цифрової інфраструктури [11]. Проте, у даній сфері точність транскрибування є

критичною для отримання задовільної результативності методів, оскільки саме достатня точність забезпечує адекватність отриманих результатів для подальшого застосування методів в комунікації, освіті, торгівлі та будь-яких інших сферах, де присутня взаємодія та обмін інформацією [12].

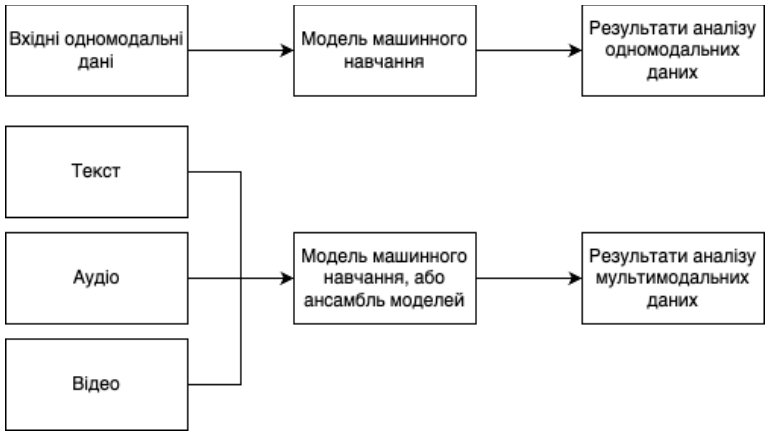


Рисунок 1.2 –Структури системи опрацювання мультимодальних даних

Дане дослідження має на меті запропонувати технологію для опрацювання мультимодальних україномовних даних засобами машинного навчання, а провалідувати отримані результати на підпроблемі перетворення українських аудіоповідомлень в текст [13].

У сфері опрацювання мультимодальних даних, на кожному з етапів використовуються різні метрики для оцінки якості та точності опрацювання. Для прикладу, на етапі векторизації різноманітних даних [14] (перетворення будь-яких даних у вектори), ми порівнюємо різноманітні набори вхідних даних і намагаємося досягнути відповідності у числових вираженнях для кожного конкретного слова і його похідних значень (числових репрезентацій), а також шукати універсальні репрезентації для сталих виразів [15]. Що в майбутньому забезпечить уніфіковане трактування та розуміння цих слів й фраз комп'ютерною системою [16].

Для оцінки якості синхронізації ми порівнюємо отримані синхронізовані вектори з еталонними значеннями, розраховуючи метрики відхилення та втрат [17]. Систем зберігання мультимодальної інформації оцінюються з погляду

ефективності швидкодії, продуктивності та здатності до збереження цілісності й узгодженості отриманих даних.

У статті [18] представлено всебічний огляд останніх розробок у галузі технологій перетворення мови в текст, пов'язаних із транскрипцією української мови, методів, використаних для аналізу, а також результатів та їхніх наслідків. Описуючи на сильні та слабкі сторони систем для перетворення мови в текст, це дослідження має на меті зробити цінний внесок у сферу опрацювання мультимодальних [19] української мови та сприяти розвитку технологій перетворення та інтерпретації мультимодальних україномовних даних.

1.2. Аналіз проблем опрацювання модальних даних

Аналітика даних [20] – це процес аналізу великих і складних джерел даних для виявлення тенденцій, моделей поведінки клієнтів і ринкових переваг, що допомагає приймати більш ефективні бізнес-рішення. Складність аналізу великих даних вимагає нових аналітичних інструментів, таких як аналітика прогнозів, машинне навчання, потокова аналітика, і такі методи, як аналіз в базі даних і в кластері [21].

Додатковою проблемою, окрім величезного обсягу даних, є складність інтеграції та синхронізації зібраних даних, яка створює складнощі управління даними. Компанії, що об'єднують неструктуровані мультимодальні джерела даних [22], мають перевагу в сфері аналізу та узагальнення інформації, оскільки можуть більш косплектно опрацьовувати дані, аналізувати приховані сенси, краще розуміти контекст, на ґрунті отриманих даних ухвалювати більш комплексні та зважені рішення [23-24].

Крім того, специфіка опрацювання великих мультимоадльних даних характеризується підвищеною швидкістю надходження вхідних даних, що поступають з різноманітних джерел, таких як датчики, мобільні пристрої, веб-потоки взаємодій і транзакцій, що призводить до необхідності аналітики в реальному часі. Організації, які можуть отримати вигоду від того, що

відбувається зараз, щоб запобігти відмові обладнання, рекомендувати предмет для покупки, виявити шахрайство з кредитними картками і багато іншого, швидко стають лідерами в своїх галузях.

Врешті, мультимодальний підхід до даних, дозволяє забезпечити порівняно високий ступень точності та достовірності даних. Це не означає, що всі дані повинні бути обов’язково попередньо опрацьованими та очищеними, для забезпечення ефективної роботи на подальших етапах витягування унікальних озна. Такий підхід у свою чергу забезпечує додаткову гнучкість в формах і типах даних, які можуть бути застосовані в ході опрацювання інформції, дозволяє отримувати дані з складних для інтерпретації джерела, наприклад з соціальних мереж, неструктурованих сховищ, метаданих тощо.

У основі аналітики лежить принцип перетворення даних в уніфікований цифровий формат, відомий як векторизація, що дозволяє підвищити цінність отриманих послідовнестей для подальшого опрацювання. Проте, зростання структурованих і неструктурованих даних, також відомих як великі дані, радикально змінило функцію аналітики. Реалізація цих можливостей вимагає двох речей: технологічного потенціалу для збору і зберігання мультимодальних даних, а також нових інструментів для перетворення даних і їх інтерпретації.

Аналітика мультимодальних даних може об'єднувати структуровані дані з застосуванням традиційних підходів опрацюваннння, та дані отримані в реальному часі з різних сенсорів для їхнього подальшої синхронізації і узагальнення на основі часової складової [25-27].

Мультимодальні дані вже стали реальністю і активно застосовуються в виробничих процесах для більшості компаній, подальшої ітерацією в розвитку мультимодальних підходів до обробки даних є мультилінгвальні методи. Проте, у випадку опрацювання україномовних даних існує проблема для застосування існуючих методів аналізу до українського контексту даних [28]. Тобто, після досягнення прийнятних результатів у сфері опрацювання мультмодальних україномовних даних, дослідники зможуть провожити удосконалювати рішення

інтегруючи інші існуючі мовні моделі для створення мультилінгвальних та мультимодальних систем опрацювання даних.

Підсумовоючи, побудова ефективних систем опрацювання мультимодальних даних є непростою і актуальною задачею з ряду причин.

1. Такі системи мають складну будову, в яких через варіативніість компонентів не завжди вдається досягнути оптимального результату, та досягнути безвідмовної та стабільної роботи.
2. Відомі системи строюють велику кулькість гетерогенних даних, які не завжди є зв’язані між собою, в рамках даного дослідження, ми приділятимемо увагу і питанням синхронізації мульимоадльних даних та їх узгодження.
3. Через велику варіативність даних не існує уніфікованого методу для побудови системи, саме цим і зумовлений вибір ансемблевого підходу на основі методі машинного навчання, що допоможе
4. Не достатньо досліджена тематика опрацювання саме українських мультимодальних даних, тобто не завжди відомі рішення показують себе ефективно при опрацювання українських наборів, потребують додаткого донавчання та налаштування.

Ця робота присвячена розробці гнучкого методу побудови цифрових двійників, який дозволяє за допомогою наборів мультимодальних спостережень використовувати якомога більше інформації, як відомої, так і невідомої досліднику, при побудові моделі системи [29]. Це забезпечує ефективніше розв’язання поставлених задач.

1.3. Концепція роботи з модальними даними

Розробка інфтерфейсу для опрацювання мультимодальних даних є важливим кроком на шляху до полегшення аналізу та інтерпретації комплексних наборів даних. Він дозволяє дослідникам і практикам об'єднувати дані з різних джерел і модальностей, щоб отримати глибше розуміння складних

явищ [30]. У наступних розділах ми опишемо архітектурне вирішення майбутньої системи, який складається з:

1. Мультимодальний інтерфейс: Модуль інтерфейсу мультимодальних даних - це компонент аналізу мультимодальних даних, який забезпечує зручний інтерфейс для взаємодії та візуалізації мультимодальних даних. Він призначений для того, щоб дозволити користувачам переглядати, маніпулювати та аналізувати дані з різних видів транспорту одночасно.

2. Модуль попередньої обробки: Модуль попередньої обробки мультимодальних даних є важливим компонентом аналізу мультимодальних даних, що забезпечує трансформацію сирих мультимодальних даних у потрібний, який може бути використаний для подальшого аналізу. Цей модуль призначений для одночасної обробки декількох модальностей, таких як текст, мова, зображення, відео та жести.

3. Підключення до бази даних: Підключення до бази даних - це модуль, за допомогою якого ми встановлюємо зв'язок між системою управління базами даних (СУБД) та інтерфейсом обробки мультимодальних даних. Це життєво важливий аспект будь-якої програми, яка взаємодіє з базою даних, дозволяючи програмі отримувати доступ до даних, що зберігаються в базі даних, і маніпулювати ними.

4. Інтерфейс перетворення: Модуль обробки запитів - це мент програмного забезпечення, який приймає запити від клієнтів, обробляє їх і формує відповідь. Зазвичай він є частиною серверної системи, що отримує звернення від різних клієнтських додатків і надсилає відповідні результати у відповідь.

5. Модуль розпізнавання: Модуль мультимодального розпізнавання аудіо та тексту - це компонент мультимодальної системи, який призначений для перетворення аудіосигналів у текст. Цей модуль зазвичай використовується в таких додатках, як розпізнавання мови, транскрипція голосу в текст і аудіокоментарі.

Детальніше з архітектури майбутньої системи, можна ознайомитися на рисунку 1.3.

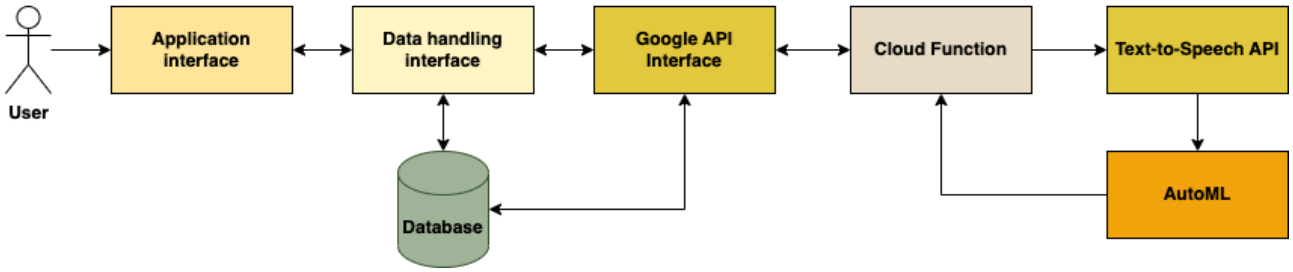


Рисунок 1.3 – Архітектурне вирішення системи [30]

Математичні методи, які лежать в основі опрацювання мультимодальних даних, містити статистичні та машинні алгоритми, які дозволяють інтегрувати, аналізувати та інтерпретувати різні типи даних (текст, зображення, аудіо, відео). До таких методів відносяться кодування, механізм уваги та синхронізації різномодальних даних [31]:

Механізми уваги грунтуються на трьох основних компонентах, а саме: запити Q, ключі K та значення V. Кожен вектор запиту Q, зіставляється з базою даних ключів K для обчислення значення оцінки. Ця операція зіставлення обчислюється як точковий добуток конкретного запиту, що розглядається, з кожним вектором ключів k:

$$E_{(q,\,k)} = Q * K_i \tag{1.1}$$

Отримана оцінки проходять через операцію softmax для отримання ваг:

$$W_{(q,\,k)} = \text{softmax}\,(E_{q,k}) \tag{1.2}$$

Узагальнена увага потім обчислюється, як зважена сума векторів значень, де кожен вектор значень пов'язаний з відповідним ключем:

$$\text{Attention}_{(q,\,K,\,V)} = \sum W_{(q,\,k)} * V \tag{1.3}$$

Опісля отримання узагальнених значень уваги, ми можемо виконувати операції синхронізації співставляючи вектори унікальних значень для даних різних модальностей за допомогою співставлення екстремуми векторів і пошуку найбільш відповідних сегментів даних і узгоджуючи їх між собою.

Конвеєр опрацювання мультимодальних даних складається з декількох основних етапів, які забезпечують уніфіковану обробку кожного типу даних, комплексування цих даних та класифікацію на основі інтегрованих моделей.

1. Уніфікована обробка кожного типу даних: Кожен тип даних (текст, зображення, аудіо, відео) обробляється за допомогою спеціалізованих алгоритмів. Наприклад:
   - Текст: Токенізація, виділення характеристик (TF-IDF, word embeddings).
   - Зображення: Попередня обробка (нормалізація, аугментація), виділення характеристик (feature extraction).
   - Аудіо: Попередня обробка (фільтрація шуму), виділення характеристик (MFCC, спектрограми).
   - Відео: Розбиття на кадри, виділення характеристик для кожного кадру (CNN+RNN).
2. Модель комплексування: Після попередньої обробки дані об'єднуються в єдину репрезентацію за допомогою моделей, які інтегрують різні модальності. Для цього можуть використовуватись гібридні моделі, які поєднують в собі різні підходи, такі як нейронні мережі.
3. Класифікаційна мережа: Інтегровані дані подаються на вхід класифікаційної мережі, яка використовує складні алгоритми, такі як глибокі нейронні мережі, для класифікації або регресії. Ця мережа навчається на основі великої кількості прикладів, щоб забезпечити високу точність та надійність результатів.
4. Результат: Після обробки та класифікації даних система генерує результати. Зазвичай результатом є дані, які перетворені в інший формат, наприклад аудіо в текст, або навпаки. Результати також можуть бути представлені у вигляді звітів, візуалізацій або інтерактивних панелей для зручності користувачів.

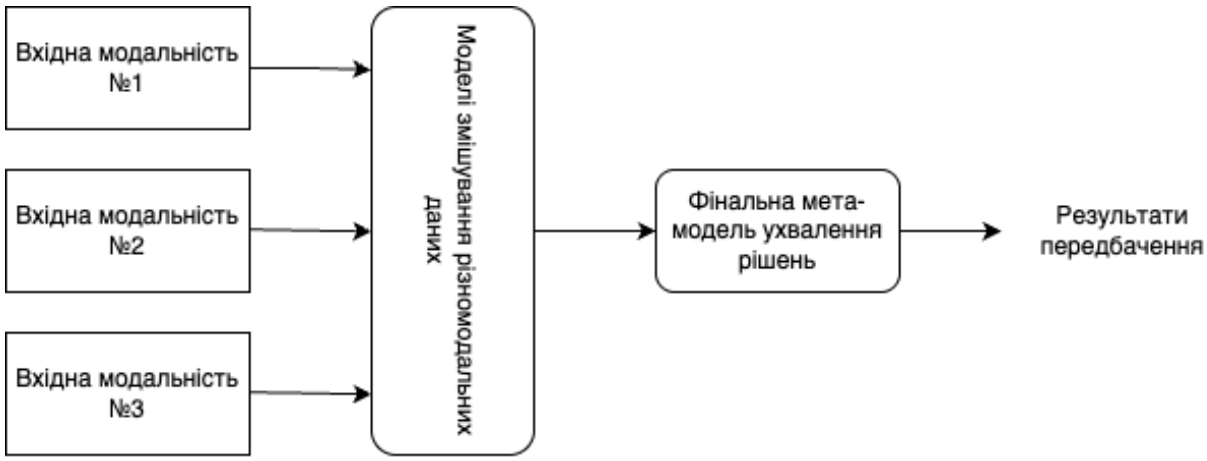


Рисунок 1.4 – Робота конвеєру опрацювання мультимодальних даних

1.4. Постановка задачі дослідження

Безліч чинників і складність опрацювання гетерогенних даних, що зокрема і може складати з себе мультимодальні набори даних, роблять область їх обробки перспективною для проведення досліджень з метою знаходження оптимальних рішень. Під час обробки даних виникає необхідність розв'язання проблеми синхронізації та узгодженості масивів даних, що особливо ускладнюється відсутністю стандартизації. Зумовлена складність варіативністю форматів, розмірів та кодувань у різних наборах даних, які використовуються у сферах та проєктах з опрацювання мультимодальних даних.

Суть дисертаційного дослідження полягає у розробленні інформаційної технології для опрацювання природних мов, з анцентом на забезпечення ефективності опрацювання саме української мови, ґрунтуючись на використанні мультимодальних підходів опрацювання даних, для підвищення точності й узагальнення досліджуваних даних і знань.

На підставі проведеного аналізу відомих рішень встановлено, що існуючі технології опрацювання мультимодальних даних, не можуть бути використані для опрацювання україномовних наборів даних через неможливість забезпечення прийнятну точність і часову ефективність опрацювання відповідних даних (Таблиця 1.2).

Таблиця 1.2. Порівняльний аналіз сучасного стану сфер використаннях мультимодальних даних

Назва моделі	**Модальності**	**Ключові особливості**	**Додатки**
CLIP	Текст та Зображення	Попередньо навчені енкодери зображень і тексту, мультимодальні нейрони	Підпис до зображення, генерація тексту до зображення
DALL-E	Текст та Зображення	Варіант GPT-3 з параметром 13B генерує зображення з тексту	Генерація тексту до зображення, маніпулювання зображеннями
GLIDE	Текст та Зображення	Використовує модель дифузії для генерації зображень, ранжує зображення	Генерація тексту до зображення, маніпулювання зображеннями
VLMo	Текст та Зображення	Модульна трансформаторна мережа зі спільною самоуважністю	Візуальні відповіді на запитання
ALIGN	Текст та Зображення	Використовує зашумлені дані альтернативного тексту для навчання окремих енкодерів	Мультимодальний пошук
ClipBERT	Відео і Текст	Поєднання CNN і трансформаторної моделі, наскрізне	Відео-мовні завдання: відповіді на запитання, пошук, підпис до зображення
VIOLET	Відео і Текст	Масковане візуально-символьне моделювання, розріджена увага	Відповіді на запитання, пошук відео

SwinBERT	Відео і Текст	Використовує архітектуру Swin Transformer для відео завдань	Відповіді на запитання, пошук відео, підпис до відео.

На сьогодні практично не існує аналогів систем, які б надали користувачів можливість перетворити мультимодальну україномовну інформацію в узагальнену структуру знань, для подальшого опрацювання. У зв’язку з цим створення технології та архітектури платформи для подальшого опрацювання мультимодальних україномовних даних дозволить покращити точність та швидкість опрацювання саме україномовних наборів даних для різних корпусів даних.

В майбутньому, пропонована технологія може знайти широке застосування в системах підтримки прийняття рішень для комунікації, журналістиці, освіті та інших. Створює передумови для подальших розробок та пошуку покращень та удосконалення наявного розв'язання науково-дослідної проблеми.

Тому для досягнення мети задачами, що поставлені у роботі, є наступні:

1. Провести аналіз сучасного стану інформаційних технологій аналізу мультимодальних даних;
2. Розробити модель мультимодальних даних;
3. Формалізувати конвеєр опрацювання мультимодальних даних;
4. Дослідити найефективніший спосіб ансамблювання моделей машинного навчання для комплексного аналізу даних різних модальностей;
5. Розробити інформаційну технологію опрацювання мультимодальних даних та апробувати результати дослідження.

Виконашви вищеперелічені завдання, результатому буде технологія модель опрацювання мультимодальних даних, що забезпечить цілісність інформації, ефективність комплексування різномодальних даних та забезпечить покращення точності отриманих результатів.

У дисертаційній роботі також потрібно розробити систему, яка функціонуватиме як універсальна платформа. Даний підхід, дозволить використання запропонованої системи в різних галузях для обробки мультимодальної інформації та зберігання повноти і цілісності зібраних даних. Мультимодальні дані, використані у дисертаційній роботі для ілюстрації роботи системи, отримані з галузі охорони здоров'я (медичні дані, перетворення мультимодальних даних в уніфіковану структуру для визначення рівня стресу пацієнта) та сфери комунікацій (взаємодія користувачів на основі бімодальних даних, для перетворення аудіо в текст і навпаки). Така апробація дозволяє системі демонструвати її можливості й потенційне застосування у майбутньому.

1.5. Висновки

У цьому розділі надано визначення мультимодальним даним, проведено аналіз моделей цих даних, описано методи, що лежать в основі технологій їх обробки, окреслено задачі, які необхідно і заплановано вирішити за допомогою даного дослідження, та визначено перспективи подальшого використання і впровадження пропонованої системи.

Також виявлені недоліки наявних методів обробки даних у наявних системах, додатково розглянута проблематика обробки українських мультимодальних даних за допомогою ансамблів моделей машинного навчання. Проведений аналіз математичного підґрунтя використання ансамблевих методів для покращення існуючих моделей в межах опрацьовуваної сфери.

На основі цього аналізу встановлено, що теперішній технології обробки мультимодальних даних, які використовують методи машинного навчання для опрацювання даних ізних модальностей, не можуть бути ефективно використані та адекватно опрацьовані через проблеми з доступом до даних та їх

обробкою у прийнятний час, особливо у випадках, коли передбачено використання українського корпусу мультимодальних даних.

Результати розділу опубліковано у таких наукових роботах [32, 33, 34].

РОЗДІЛ 2. РОЗРОБЛЕННЯ АНСАМБЛЮ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ

У розділі розглянуто підходи до опрацювання унімодальних наборів даних з подальшим злиттям і опрацювання вже в форматі мультимодальних наборів. Здійснено аналіз методів комплексування мультимодальних даних. Розроблено метод оцінки якості перетворених наборів даних. Оцінено результативність різних видів злиття векторів гетерогенних даних. Розроблено архітектуру зберігання мультимодальних даних та методи додавання даних у відповідні структури даних.

Матеріали розділу опубліковані у роботах автора [84, 85, 86, 87].

2.1. Вибір типів моделей даних для представлення модальних даних

Основою будь-якої інформаційної системи є модель даних, яка містить у собі структури для зберігання, обмеження цілісності й дозволяє виконувати операції маніпуляції над даними в системи. При роботі з мультимодальними даними, доводиться враховувати особливості для зберігання різнотипової (унімодальної) інформації, яка забезпечує повноцінну роботу мультимодальної системи опрацювання даних [35].

Мультимодальна та унімодальна моделі представляють два різних підходи до розробки систем штучного інтелекту. У той час як унімодальна модель фокусується на навчанні систем для виконання одного завдання з використанням одного джерела даних, мультимодальна модель прагне інтегрувати кілька джерел даних для всебічного аналізу заданої проблеми.

Нижче наведено детальне порівняння цих двох підходів [36]:

1. Одномодальні структура даних призначені для обробки одного типу даних, таких як зображення, текст або аудіо.
2. Мультимодальні структури даних призначені для інтеграції декількох джерел даних, включаючи зображення, текст, аудіо та

відео у єдину узагальнену структуру з можливістю часової та змістової синхронізації наборів.

Вибір моделі даних є ключовим етапом у проєктуванні ефективних систем штучного інтелекту, які працюють з модальними даними. У цьому контексті розрізняють два основних підходи: унімодальні та мультимодальні моделі.

Унімодальні моделі зосереджені на одному типі даних — тексті, зображеннях, аудіо тощо. Вони менш складні в реалізації, але часто обмежені у своїх аналітичних можливостях через відсутність контексту та взаємопідсилюючої інформації з інших джерел [37]. Приклад унімодальних типів даних — текст та зображення. Ілюстрація демонструє окреме представлення двох типів даних в контексті унімодального підходу, що може бути доцільним у випадках однорідних даних або вузькоспеціалізованих завдань. Представлення виконане у форматі, сумісному з мультимодальними репрезентаціями, наприклад, через уніфіковані векторні просторові уявлення на Рисунку 2.1.

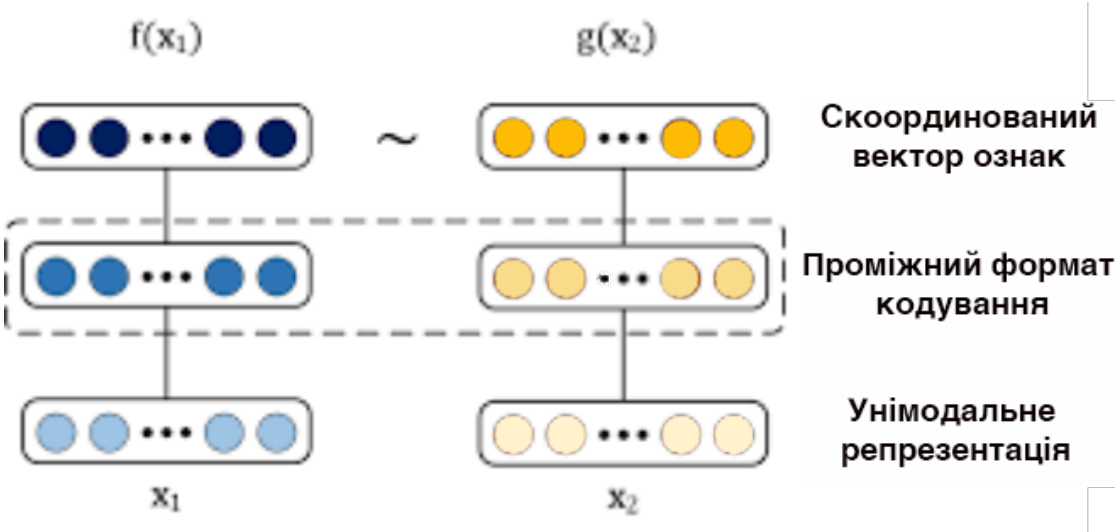


Рисунок 2.1 – Приклад екстракції векторів унікальних ознак

Натомість мультимодальні моделі дозволяють інтегрувати декілька типів даних (текст, зображення, відео, аудіо), забезпечуючи глибший і ширший аналіз завдяки синхронізації в часовому та змістовому контексті. Наприклад, системи

розпізнавання емоцій або автоматичного опису зображень отримують перевагу саме через багатство інтегрованих джерел інформації [38].

До переваги та обмеження мультимодальних моделей відносять такі:

1. Складність реалізації - Унімодальні системи штучного інтелекту зазвичай мають простішу архітектуру, оскільки працюють лише з одним типом даних — наприклад, лише з текстом або лише з зображенням. Це спрощує розробку, навчання та обчислення. Натомість мультимодальні системи потребують складнішої інфраструктури для синхронного аналізу декількох модальностей, що підвищує вимоги до ресурсів і архітектурних рішень.
2. Контекстуальність - унімодальні моделі обмежені рамками однієї модальності, через що їм часто бракує повного контексту, необхідного для точного аналізу. Мультимодальні системи інтегрують інформацію з різних джерел (наприклад, поєднують текст із зображенням чи аудіо), завдяки чому можуть формувати ширше уявлення про ситуацію та покращувати якість прогнозів або класифікацій.
3. Продуктивність у різних умовах - у задачах, де є лише один домінантний тип даних (наприклад, розпізнавання мови або переклад тексту), унімодальні моделі можуть показувати високу ефективність. Проте в більш комплексних завданнях — таких як опис зображення, відеоаналіз або аналіз емоцій — унімодальні підходи можуть втрачати релевантність через обмежену доступність контексту. Мультимодальні системи в таких випадках демонструють значно кращу адаптивність і точність.

Репрезентація модальностей має вирішальне значення для ефективності моделі. Для кожної модальності застосовуються відповідні техніки:

- Текст: Word2Vec [39], BERT [40], GPT [41] — для отримання контекстуальних семантичних векторів.

- Зображення: CNN [42] — для виявлення просторових ознак.
- Аудіо: спектрограми, MFCC [43] — для захоплення ритму, тону та вмісту.

Відео: 3D-CNN [44], RNN [45] — для врахування просторово-часових патернів.

Кожна модальність має власний спеціалізований кодер: для зображень використовується CNN, для текстових даних — трансформери GPT або BERT. На етапі обробки застосовуються механізми уваги (attention) для виокремлення найбільш інформативних фрагментів даних. Отримані векторні представлення зводяться до уніфікованого формату, після чого об'єднаний вектор ознак передається на фінальну мета-модель для ухвалення кінцевого рішення, на рисунку 2.2. детальніше представлено узагальнену схему конвеєра обробки бімодальних (текст та зображення) даних, з подальшим зведенням їх до одного формату і передавання об'єднаного вектору унікальних ознак на фінальну мета модель для ухвалення фінального рішення.

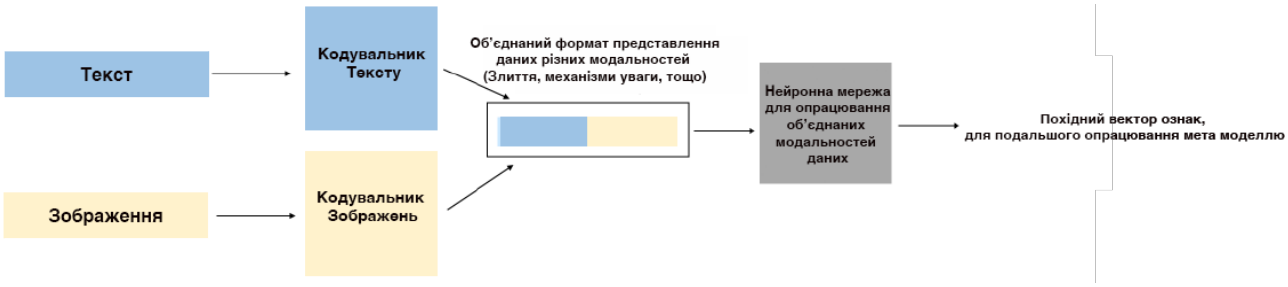


Рисунок 2.2 – Візуалізація узагальненого конвеєра опрацювання мультимодальних даних

Інтеграція даних різних типів дозволяє моделі компенсувати недоліки кожної окремої модальності, посилюючи загальну точність та надійність. Наприклад, якщо текст є неоднозначним, контекст із зображення або відео допоможе розкрити значення. Крім того, це критично для вирішення завдань, що потребують контекстуального розуміння, як-от розпізнавання намірів користувача, сарказму або емоцій [46].

Приклади мультимодальних моделей

- Data2Vec [47] — універсальна модель від MetaAI, яка застосовується до тексту, аудіо та зображень у єдиній архітектурі, заснованій на трансформерах.
- VilBERT [48] — обробляє перехресні завдання між мовою та візуальною інформацією.
- Flamingo [49[] — модель з можливостями few-shot навчання, яка поєднує LLM з мультимодальним вхідним потоком.

Злиття даних у рамках мультимодальних підходів має за мету максимізувати інформативність наявних даних шляхом поєднання сильних сторін різних джерел. Такий підхід дозволяє сформувати більш цілісне, узгоджене та точне уявлення про об'єкт чи явище, що вивчається. Завдяки синергетичному ефекту, досягнутому шляхом поєднання гетерогенних даних, система набуває здатності виявляти закономірності та залежності, які були б недоступні в умовах аналізу лише однієї модальності [50].

Мультимодальні дані, такі як аудіо, текст та фізіологічні сигнали, забезпечують різні погляди на досліджуваний об'єкт або поведінковий патерн. Кожна модальність несе унікальну інформацію, а їх поєднання дозволяє моделі навчатися на взаємодоповнюючих характеристиках, знижуючи ймовірність помилок, викликаних шумами або пропущеними даними в окремих джерелах. Це забезпечує більш глибоке розуміння контексту, в якому відбувається взаємодія, та підвищує ситуаційну обізнаність у прикладних системах. (див. рисунок 2.3)

Крім того, інтеграція різних модальностей дає змогу компенсувати слабкі сторони окремих джерел даних. Наприклад, у випадку недостатньої чіткості тексту, аудіоінформація може містити додаткові паралінгвістичні ознаки, такі як інтонація чи тембр голосу, що дозволяє покращити точність висновків. У свою чергу, аналіз тексту надає важливу семантичну складову, яка доповнює емоційне забарвлення аудіоданих. Така взаємодія дає змогу забезпечити глибший аналіз психоемоційного стану, що є особливо важливим у завданнях

скринінгу психічного здоров’я [51].

У процесі злиття даних використовуються різні техніки, такі як раннє (на рівні ознак), середнє (на рівні латентних представлень) та пізнє (на рівні прийняття рішень) злиття. Кожен підхід має свої переваги залежно від специфіки задачі та типу модальностей..

Загалом, мультимодальна інтеграція формує більш надійну аналітичну основу для прийняття рішень у складних середовищах. Вона підвищує адаптивність моделей, дозволяє краще справлятися з неоднозначністю, а також забезпечує багатовимірну перспективу аналізу. Внаслідок цього, прийняті рішення є більш поінформованими, точними та релевантними поставленим цілям [52].

Таким чином, злиття даних у мультимодальних системах відіграє критично важливу роль у формуванні стійких і продуктивних моделей машинного навчання, що є особливо актуальним у сферах медицини, психодіагностики, безпеки та користувацької аналітики, де точність і надійність даних мають вирішальне значення.

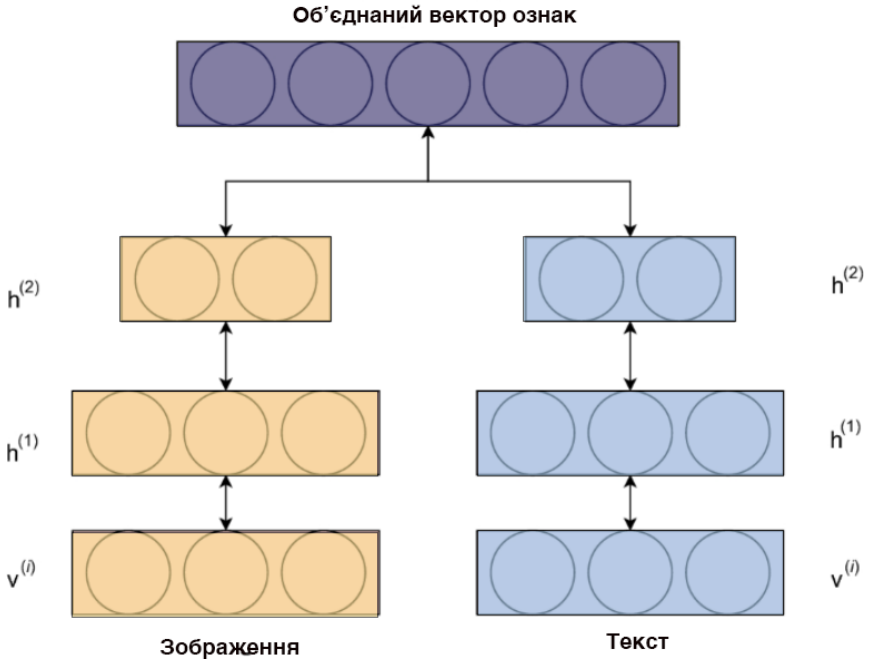


Рисунок 2.3 – Схема злиття різномодальних даних у об’єднаний вектор унікльних ознак

Тому методи і підходи для опрацювання мультимодальних наборів даних широко використовується в різних сферах, включаючи оборону, моніторинг навколишнього середовища, охорону здоров'я, фінанси тощо [53].

Злиття даних передбачає кроки й методи для інтеграції та об'єднання даних з різних джерел. Джерела даних можуть включати датчики, бази даних та інші сховища інформації [54]. Процес може змінюватися залежно від програми та типів даних, але загалом він складається з таких загальних кроків:

1. Попередня обробка даних: На цьому етапі ми готуємо і перетворюємо необроблені дані з різних джерел у стандартний формат. Це робиться шляхом виконання таких завдань, як очищення, нормалізація, масштабування та видалення пропусків або помилок. Таким чином, це може включати очищення даних, нормалізацію, масштабування та видалення пропусків.
2. Вилучення ознак: На цьому кроці ми витягуємо відповідні ознаки або атрибути з попередньо оброблених даних. Вилучення ознак має на меті визначити та зафіксувати найбільш інформативні аспекти даних, що мають відношення до завдання злиття. Тому для вилучення значущих ознак можуть використовуватися статистичні методи, алгоритми обробки сигналів або знання, специфічні для конкретної галузі.
3. Інтеграція даних: Після вилучення ознак дані з різних джерел об'єднуються в єдиний набір даних. Таким чином, методи інтеграції залежать від характеру даних. Крім того, для числових даних можна використовувати такі методи злиття, як зважене усереднення, статистичні методи або регресійні моделі.
4. Вибір алгоритму злиття: Відповідний алгоритм або метод об'єднання обирається на основі конкретного завдання об'єднання та бажаного результату. Це можуть бути статистичні методи, алгоритми машинного навчання або підходи на основі правил.

5. Прийняття рішень або аналіз: Після того, як дані з різних джерел об'єднані за допомогою обраного алгоритму злиття, отримані в результаті злиття дані використовуються для прийняття рішень або аналізу.

На рисунку 2.4., представлено двоетапний процес обробки бімодальних даних. На першому етапі відбувається кодування текстових і візуальних даних окремими модулями та визначення вагових коефіцієнтів для кожної з отриманих репрезентацій [55]. На другому етапі здійснюється інтеграція закодованих представлень у єдиний вектор унікальних ознак, що надалі використовується метамоделлю для визначення фінального результату.

Рисунок 2.4 – Узагальнена візуалізація кроків опрацювання бімодальних даних

Крім того, глобальне злиття даних особливо актуальне в таких сферах, як глобальна безпека і моніторинг навколишнього середовища. Крім того, когнітивне злиття даних має на меті моніторинг продуктивності промислових активів, оптимізацію графіків технічного обслуговування та підвищення загальної надійності активів [56]. Таким чином, об’єднання даних з різних джерел має на меті подолання обмежень і недоліків окремих джерел.

2.2. Формальний опис структури модальних даних

Важливо етам у контексті поставленої задча є формалізації структури модальних даних, що пов’язаних із трансформацією аудіоінформації, відеоінформацію в набір числових векторів, де ключову роль відіграють

послідовні дані (sequential data), що характеризуються тимчасовою впорядкованістю та залежностями між елементами в межах потоку.

Ще одним поширеним підходом до вирішення задач регресії є використання багатошарової штучної нейронної мережі, реалізованої, зокрема, у вигляді моделі MLPRegressor [57] (Multi-Layer Perceptron Regressor). Цей алгоритм базується на принципі імітації роботи біологічних нейронів і являє собою послідовність шарів повнозв'язаних штучних нейронів. На відміну від класичних моделей, таких як лінійна або поліноміальна регресія, багатошаровий перцептрон здатний апроксимувати дуже складні нелінійні залежності між вхідними ознаками X та цільовою змінною y.

Модель складається зі вхідного шару, одного або кількох прихованих шарів і вихідного шару. Кожен шар містить певну кількість нейронів, і кожен нейрон виконує обчислення лінійної комбінації своїх вхідних значень із вагами та зміщенням, після чого результат передається через нелінійну функцію активації. У процесі навчання MLPRegressor мінімізує певну функцію втрат, яка зазвичай є середньоквадратичною помилкою:

Модальні дані, що надходять на вхід моделі, є прикладом таких послідовностей, і тому обрана модель має бути здатною не лише ефективно зберігати контекст, а й адаптивно реагувати на динаміку вхідного сигналу. У цьому контексті однією з найпридатніших архітектур для роботи з такими даними є RNN (Recurent Neural Network), яка забезпечує здатність до тривалого збереження інформації та боротьби з проблемами затухання або вибуху градієнтів під час навчання [58].

1. Формальний опис структури модальних даних має відповідати низці ключових вимог, які забезпечують повноту, зрозумілість, гнучкість та програмну сумісність моделі:
2. Інтуїтивність: усі змінні, що використовуються у формальному описі, повинні бути семантично зрозумілими, логічно пов’язаними

з відповідними функціями компонентів моделі та мати однозначні позначення.

3. Повнота: рівняння та формалізми мають охоплювати як пряме проходження (forward pass), що імітує нормальну роботу моделі, так і зворотне проходження (backpropagation), що відповідає за навчання. Усі функціональні елементи архітектури повинні бути враховані без виключень.
4. Узагальненість: опис повинен базуватись на повній (ванільній) формі RNN, включно з механізмом pinhole connections, які дозволяють стану комірки впливати на вузли керування — це критично для точного захоплення довготривалих залежностей у послідовностях.
5. Ілюстративність: документація повинна містити чітко позначені схеми та діаграми, які візуалізують як внутрішню структуру RNN-комірки, так і послідовність операцій обробки даних, з мінімізацією когнітивного навантаження на дослідника або розробника.
6. Модульність: формалізм повинен дозволяти легку інтеграцію RNN-комірки як частини більш складної архітектури, з можливістю масштабування як у глибину (deep sequence), так і по рівнях представлення (deep representation).
7. Векторна нотація: всі рівняння повинні бути подані у векторно-матричній формі, що дозволяє безпосередньо реалізувати їх у програмних бібліотеках для лінійної алгебри (таких як NumPy, PyTorch чи TensorFlow) без необхідності ручного розгортання по індексах.

Для ефективної обробки послідовних модальних даних, що надходять на вхід системи (наприклад, аудіосигналів або транскрибованих фрагментів мовлення), обрано архітектуру, здатну моделювати часову динаміку та залежності між елементами послідовності. З цією метою формалізація завдання

буде базуватись на підході рекурентної нейронної мережі (Recurrent Neural Network, RNN) — класу моделей, що імітує часову еволюцію системи за допомогою ітеративного оновлення стану [59].

Почнемо з аналітичного опису RNN на основі диференціальних рівнянь, що дозволяє розглядати динаміку мережі як безперервний процес у часі. Нехай $\vec{s}(t)$ — це d-вимірний вектор стану системи у момент часу *t*. Тоді загальне формулювання еволюції стану можна описати за допомогою нелінійного неоднорідного диференціального рівняння першого порядку [60]:

$$ds(t)/dt = f(s(t), x(t), t) \quad (2.1)$$

де:

- s(t)— вектор внутрішнього стану моделі (hidden state),
- x(t) — вхідний вектор (наприклад, вектор ознак аудіосигналу або токенізований текст),
- f (s(t), x(t))— нелінійна функція, яка описує взаємодію між поточним станом, вхідними даними та, за необхідності, самим часом.

Тепер встановіть затримку $\tau 0$, що дорівнює одному часовому кроку. Це можна інтерпретувати як збереження значення сигналу зчитування в пам’яті

на кожному часовому кроці, щоб використовувати його у наведених вище рівняннях на наступному часовому кроці. Після одноразового використання пам’ять може бути перезаписана оновленим значенням сигналу зчитування для використання на наступному часовому кроці і так далі10. Таким чином, поклавши $\tau 0 = \Delta T$ і замінивши для зручності знак апроксимації на знак рівності в рівнянні 19, отримаємо

$$g_t(x)=h_1(x)+h_2(x,N(x,P)) \quad (2.2)$$

Після дискретизації всі вимірювання часу в рівнянні 2.2 стають цілими числами, кратними часу дискретизації кроку, ΔT. Тепер ΔT можна відкинути з аргументів, що залишає вісь часу безрозмірною. Таким чином, всі сигнали перетворюються на послідовності, областю визначення яких є дискретний

індекс n, а рівняння перетворюється на нелінійне неоднорідне різницеве рівняння першого порядку.

Метод зворотного Ейлера – це стійке правило дискретизації, яке використовується для чисельного розв'язування звичайних диференціальних рівнянь [70]. Простий спосіб виразити це правило полягає в тому, щоб підставити пряму формулу скінченних різниць у визначення похідної, послабити вимогу $\Delta T \to 0$ і оцінити функцію в правій частині (тобто величину, якій дорівнює похідна) в момент часу $t + \Delta T$.

Застосування правила дискретизації легко поширити на повне рівняння 11, що містить будь-які або всі члени із запізненням у часі та їхні відповідні матричні коефіцієнти, без наведених вище спрощень. Таким чином, загальність не втрачається.

Знову ж таки, додаткові члени, що містять подібним чином об'єднані вхідні сигнали із затримкою в часі (як буде показано пізніше в цій роботі) та сигнали стану, можуть бути включені в дискретизацію, послаблюючи наведені вище спрощення, якщо це необхідно для задоволення вимог конкретної задачі, що розглядається. Визначивши дві додаткові вагові матриці та вектор зсуву, перетворює наведену вище систему до канонічної форми рекурентної нейронної мережі (RNN) [61]:

$$Wr = (\Delta T)WsB \tag{2.3}$$

$$Wx = (\Delta T)WsC \tag{2.4}$$

$$\vec{\theta}s = (\Delta T)Ws\vec{\phi} \tag{2.5}$$

Це рівняння задає зміну стану мережі в часі на основі поточного стану та вхідного сигналу. Такий підхід дозволяє трактувати RNN як дискретизацію безперервного процесу. У дискретному вигляді (що використовується в практичних реалізаціях) воно набуває вигляду:

$$\vec{\theta}s = (\Delta T)Ws\vec{\phi} \tag{2.6}$$

$$\vec{\theta}s = (\Delta T)Ws\vec{\phi} \tag{2.7}$$

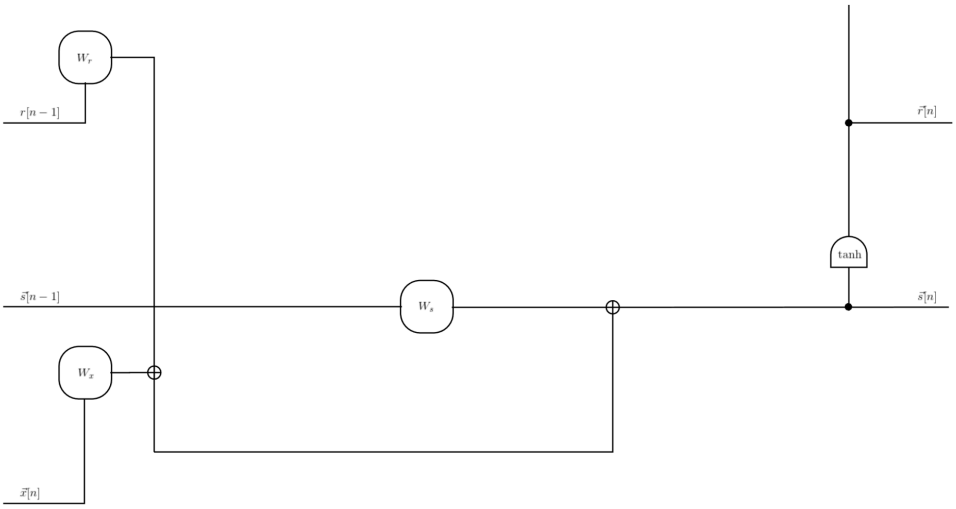


Рисунок 2.5 – Канонічна комірка RNN

Формулювання RNN у рівняннs зображеному на рисунку 2.5, пізніше буде логічно розвинуто в систему LSTM. Раніше що вигідно ввести процес «розгортання» та поняття «комірки» РНМ. Ці поняття будуть простіше описати за допомогою стандартного визначення RNN, яке далі виводиться з рівняння 30 на основі аргументів стабільності [62].

Цей підхід створює основу для подальшої деталізації — переходу до LSTM-мереж, які є вдосконаленою формою RNN і краще пристосовані до моделювання довготривалих залежностей завдяки введенню механізмів збереження та контролю інформації в пам’яті. В подальших секціях буде детально розглянуто формальну структуру ванільного LSTM-блоку [63], включно з усіма внутрішніми елементами (вхідний, вихідний, забувальний шлюзи, а також стан комірки), з векторно-матричним представленням та повною схемою інформаційного потоку.

Формалізація завдання не лише покращує точність моделювання часових патернів, але й дозволяє інтегрувати LSTM-модулі [64] в більші архітектури, як описано раніше — з подальшим ансамблюванням через стейкінг та злиттям результатів для фінального передбачення.

2.3. Аналіз існуючих моделей ансамблю машинного навчання для роботи з модальними даними

Ансамблеві методи — це потужний інструмент побудови моделей машинного навчання, який дозволяє підвищити точність передбачень, а також побудувати більш уніфіковані та стійку структури для опрацювання різнотипових даних.

Ансамблеві підходи зменшують зсув і дисперсію даних, одночасно підвищуючи точність моделі. У більшості ансамблевих систем базове навчання виконується за допомогою єдиного алгоритму, що забезпечує однорідність отриманих результатів серед усіх моделей в ансамблі. Однорідні мають схожі властивості і зазвичай опрацьовують дані одного типу.

Похідною є можливість побудови гетерогенних ансамблів на основі відповідних даних, проте специфіка застосування ансамблевих підходів саме гетерогенних та мультимодальних наборах є недостатньон дослідженою, оскільки у таких випадках ансамблювання потребує вузькоспеціалізованого застосування та важко знайти «універсальну» модель для вирішення усіх завдань. Особливо у випадках злиття мультимодальних або мультисенсорних наборів даних.

При цьому при ансамблюванні важливими є поняття слабкого та сильного учня (weak/strong learners). Слабкий учень — це базова модель (наприклад, лінійна регресія або дерево рішень), яка самостійно може мати обмежену ефективність. Сильний учень — це модель, яка об'єднує результати кількох слабких учнів, зменшуючи похибку чи варіативність [65].

У даному дослідження застосовуватимемо ансамблеві моделі за принципом strong learners, що дозволить нам знизити дисперсію для методу синхронізації даних, а також зменшити зсув для методу бустінгу та покращити загальну прогностичну ефективність запропонованого рішення.

Детальніше з найпопулярнішими методами ансамблювання можна ознайомитися на Рис. 2.6.

Рисунок 2.6 – Типологія ключових ансамблевих методів

Найпоширенішими типами ансамблевих методів є стекинг, бэггинг і бустинг.

Стекинг (stacking) — це метод, який використовує кілька різнорідних слабких учнів. Кожна модель навчається на вхідних даних, а їхні передбачення подаються мета-моделі (або мета-учню), яка формує кінцевий прогноз.

Мета-модель навчається на основі передбачень слабких учнів. Для цього набір даних розбивається на дві частини: слабкі учні навчаються на першій частині, а їхні прогнози тестуються на другій. Ці прогнози формують новий набір даних, на якому навчається мета-модель [66].

Бегинг (bagging) — це метод, який передбачає навчання кількох однакових моделей на різних підвибірках даних. Зразки створюються методом бутстрепу (bootstrap), тобто шляхом випадкового вибору з поверненням.

У випадку регресії результати моделей усереднюються, а для класифікації використовується голосування. Голосування може бути жорстким (за найбільш популярний клас) або м’яким (усереднення ймовірностей) [67].

Бустинг (boosting) — це метод, у якому моделі навчаються послідовно. Кожна наступна модель фокусується на коригуванні помилок попередньої. Найпоширенішими алгоритмами бустингу є адаптивний бустинг (AdaBoost) і градієнтний бустинг [68].

Адаптивний бустинг (AdaBoost) - спочатку навчає першу базову модель, коригуючи ваги неправильно класифікованих зразків. Наступні моделі навчаються на оновлених вагових коефіцієнтах. Кінцевий результат ансамблю визначається як зважена сума передбачень моделей [69].

Підсумовуючи до переваги ансамблевих методів

1. Підвищення точності - завдяки поєднанню кількох моделей, ансамблеві методи здатні зменшити помилки і загалом покращити продуктивність. Вони часто показують кращі результати порівняно з окремими моделями, особливо на складних задачах класифікації чи регресії.
2. Зменшення варіативності та зміщення - ансамблі можуть компенсувати недоліки окремих моделей: наприклад, зменшити високий дисперс у нестабільних моделей або компенсувати зміщення у недонавчених моделей. Це сприяє кращій узагальнюваності на нових даних.
3. Гнучкість у виборі моделей -Ансамблеві методи дозволяють комбінувати моделі різного типу (дерева, SVM, логістична регресія тощо), що забезпечує більшу адаптивність до різних типів даних та завдань.

До недоліків ансамблевих методів відносять:

1. Підвищення точності - завдяки поєднанню кількох моделей, ансамблеві методи здатні зменшити помилки і загалом покращити продуктивність. Вони часто показують кращі результати порівняно з окремими моделями, особливо на складних задачах класифікації чи регресії.
2. Зменшення варіативності та зміщення - ансамблі можуть компенсувати недоліки окремих моделей: наприклад, зменшити високий дисперс у нестабільних моделей або компенсувати

зміщення у недонавчених моделей. Це сприяє кращій узагальнюваності на нових даних.

3. Гнучкість у виборі моделей -Ансамблеві методи дозволяють комбінувати моделі різного типу (дерева, SVM, логістична регресія тощо), що забезпечує більшу адаптивність до різних типів даних та завдань.

Таким чином, ансамблеві методи є ефективним підходом для роботи з модальними даними, що поєднує простоту реалізації та високу точність передбачень.

У сфері машинного навчання стекінг є потужною технікою в арсеналі ансамблевих методів. Простими словами, стекінг (або стекування) в машинному навчанні - це як команда супергероїв, які зібралися разом, щоб вирішити складну проблему. Кожен супергерой має свої унікальні здібності, а вони працюють разом і перемагають... Так само і в стекінгу є кілька моделей, кожна з яких навчена робити прогнози по-своєму. Потім є ще одна модель, яка називається мета-модель, що навчається, яка бере прогнози від інших моделей і використовує їх, щоб зробити остаточний прогноз [70].

Я думаю, що ідеальне порівняння стекінгу - це якась викривлена система прийняття рішень у політиці. У нас є група політиків, які пропонують свої рішення щодо проблеми - у нас є президент, який підписує документи. Подібно до такої демократії в політиці, стекінг має значну перевагу в ML - використання колективного досвіду. Поєднуючи взаємодоповнюючі сильні сторони і пом'якшуючи індивідуальні слабкі сторони, стекінг знижує ризик упередженості моделі і перенавчання - в результаті ми отримуємо оптимальні результати.

На фінансових ринках стекінг знайшов застосування у прогнозуванні цін на акції та алгоритмічній торгівлі, де точне прогнозування має першорядне значення для прийняття обґрунтованих рішень. В охороні здоров'я стекінг допомагає діагностувати та прогнозувати захворювання, використовуючи дані

від історії хвороби та письмових скарг до флюорографічних знімків або даних з флюорографа.

Поєднання прогнозів декількох моделей може забезпечити кращу продуктивність, ніж будь-яка окрема модель. Стекінг розширює цей принцип, представляючи ієрархічний підхід, в якому прогнози різних базових моделей фільтруються метанавчальним пристроєм, таким чином синтезуючи більш надійний і точний остаточний прогноз.

На відміну від традиційних методів ансамблевого навчання, таких як bagging і busting, які працюють на одному рівні агрегації, стекінг представляє багаторівневу архітектуру. На базовому рівні ансамбль абсолютно різнорідних базових моделей навчається незалежно на вхідних даних, кожна з яких виявляє різні закономірності та взаємозв'язки в наборі даних. Ключове слово тут - «незалежно».

Роль мета-навчаючого пристрою подвійна: він вивчає оптимальну комбінацію прогнозів базових моделей і генерує остаточний прогноз на основі цієї комбінації. Мета-навчальник може приймати різні форми, починаючи від простих лінійних моделей, таких як лінійна регресія, і закінчуючи більш складними алгоритмами, аж до нейронів або алгоритмів градієнтного бустінгу.

Алгоритм стекування [71]:

1. Розбиття вибірки на k складок.
2. Для кожного об'єкта в k-му згині робиться прогноз за допомогою базових алгоритмів, навчених на решті k-1 згинів.
3. Створюється набір прогнозів базових алгоритмів для кожного об'єкта у вибірці.
4. Метамодель навчається на згенерованих прогнозах базових алгоритмів.

Початковий навчальний набір розбивається на k згинів для перехресної перевірки. Кожна базова модель навчається на k-1 згинах і робить прогнози на решті згинів. Це повторюється для кожної складки, і прогнози об'єднуються в

нову ознаку. Ця процедура повторюється для кожної базової моделі. Крім того, базові моделі навчаються на всьому вихідному навчальному наборі, а потім роблять прогнози на тестовому наборі для створення нового тестового набору.

До речі, існує спрощена реалізація «стекування» - блендінг. При блендингу вихідний набір даних ділиться на дві частини: навчальну та перевірочну. Базові моделі навчаються на навчальному наборі, а потім роблять прогнози на перевірочному наборі. Ці прогнози використовуються для навчання мета-моделі, яка потім використовується для прогнозування на валідаційному наборі даних

Ще одним важливим підходом до розв'язання задач регресії є Stacking (stacked generalization). Цей метод також об'єднує кілька базових моделей, однак, на відміну від VotingRegressor, робить це через побудову додаткової навченої моделі другого рівня. Суть Stacking полягає в тому, що спочатку кілька різних базових моделей f_1, f_2, f_m навчаються на вихідних даних (X,y) і для кожного об'єкта x_i створюють свої передбачення y_1, y_2, y_m. У результаті для кожного прикладу формується новий вектор ознак: $z_i = f(y_1, y_2, y_m)$, де z_i — це набір прогнозів усіх базових моделей для об'єкта і.

Далі на основі цих нових ознак навчається окрема мета-модель g, яка отримує вхідні дані z_i та справжні значення y_i, і намагається побудувати оптимальне передбачення: $y_i = g(z_i)$.

Тобто замість простого об'єднання результатів моделей, як у Voting, у Stacking будується нова навчальна задача: мета-регресор аналізує вихідні прогнози моделей і вчиться, яким чином краще їх комбінувати для підвищення точності.

Основна технічна відмінність між підходом Stacking і VotingRegressor полягає в тому, що VotingRegressor використовує просту формулу об'єднання прогнозів або її зважену версію, де ваги wmw_mwm враховують важливість кожної моделі. Щоб уникнути переобучення на тих самих даних, які використовуються для тренування базових моделей, у практиці Stacking часто

застосовується схема крос-валідації: базові моделі спочатку генерують прогнози на даних, які вони ще не бачили, і тільки потім ці прогнози використовуються для навчання мета-регресора. Завдяки цьому мета-модель отримує більш реалістичну оцінку помилок базових моделей і краще навчається компенсувати їхні слабкі сторони.

Основна різниця між VotingRegressor [72] та Stacking полягає у способі комбінування прогнозів базових моделей. У VotingRegressor фінальний прогноз розраховується як середнє (арифметичне або зважене) значення прогнозів усіх моделей, без додаткового навчання. Це простий і швидкий підхід, який дозволяє об’єднати сильні сторони різнорідних моделей, компенсуючи їхні слабкості.

Натомість Stacking формує другий рівень навчання: прогнози базових моделей використовуються як нові ознаки, на основі яких метамодель (метарегресор) вивчає закономірності і формує фінальний прогноз. Це дає змогу виявляти складні залежності між помилками окремих моделей і потенційно покращити точність.

Отже, VotingRegressor простіший та менш ресурсоємний, але Stacking забезпечує більшу гнучкість і часто вищу точність, особливо коли базові моделі мають різну природу.

У Таблиці 2.1 наведено порівняльну характеристику трьох найпоширеніших методів ансамблювання в машинному навчанні — стекінгу, бустингу та беггінгу — за ключовими критеріями ефективності та застосування.

Таблиця 2.1 – Таблиця порівняння ансамблевих методів.

Тип	**Стекінг**	**Бустинг**	**Беггінг**
Основна ідея	Комбінування різних моделей через метамодель	Послідовне навчання моделей з фокусом на помилки попередніх	Паралельне навчання моделей на різних підвибірках даних
Тип комбінації	Навчання метамоделі для об’єднання передбачень	Зважене об'єднання слабких моделей	Середнє або голосування результатів моделей
Взаємодія між моделями	Наявна (через метамодель)	Сильна (послідовна корекція)	Відсутня (моделі навчаються незалежно)
Стійкість до перенавчання	Висока за правильного налаштування	Помірна, ризик переобучення за великої кількості ітерацій	Висока, особливо з великими деревами або базовими моделями
Паралелізація навчання	Обмежена	Важко паралелізувати через послідовність виконання	Добре паралелізується
Приклади методів	Stacked Generalization	AdaBoost, XGBoost, LightGBM	Random Forest
Переваги	Висока точність, хороша робота на складних даних	Висока точність, хороша робота на складних даних	Стабільність, зменшення варіації даних
Недоліки	Вища складність побудови і налаштування	Чутливість до шумів і викидів	Не гарантує покращення результатів

2.4. Розроблення моделі машинного навчання

Після детального опрацювання можливих напрямів застосування технологій перетворення аудіо в текст у медичній сфері, а також формування чітких вимог до системи та структури відповідних датасетів, настає етап безпосередньої розробки моделі машинного навчання. Основна мета цього етапу — побудова універсального та гнучкого рішення, яке зможе забезпечити високу точність розпізнавання в широкому спектрі клінічних задач.

Для досягнення високої ефективності та узагальнювальної здатності моделі обрано підхід ансамблювання моделей за допомогою техніки стейкінгу (stacking). Цей підхід дозволяє поєднувати вихідні передбачення кількох базових моделей (наприклад, LSTM, GRU [73], трансформери або навіть класичні ML-алгоритми), які навчаються паралельно на спільному наборі даних. Кожна з моделей може спеціалізуватися на різних аспектах вхідного сигналу — одна ефективно обробляє короткі та чіткі фрази, інша — краще розпізнає медичну термінологію або працює з шумними аудіозаписами.

Отримані результати окремих моделей поєднуються за допомогою м’якого голосування [74] (soft voting) — методу, при якому враховуються не тільки остаточні рішення моделей, а й їх ймовірнісні передбачення. Такий підхід забезпечує більшу гнучкість, дозволяючи отримати зважене об’єднане передбачення, яке відображає загальний консенсус ансамблю, а не просто більшість голосів.

Після етапу софтвоутингу результуюче зважене передбачення проходить через функцію активації ReLU [75] (Rectified Linear Unit). Використання ReLU дозволяє моделі краще опрацьовувати нелінійні залежності в даних, фільтруючи негативні значення та підвищуючи стійкість до вибуху градієнтів. Такий підхід є особливо ефективним у багатошарових нейронних мережах, де складність простору рішень вимагає глибокої та адаптивної активації.

Таким чином, побудова моделі відбувається у кілька логічно структурованих етапів:

1. Генерація попередніх передбачень базовими моделями;
2. Злиття результатів через механізм soft voting;
3. Обробка через функцію активації ReLU;

Отримання фінального результату, що відображає як індивідуальні сильні сторони окремих моделей, так і їхню колективну узгодженість.

У подальших підрозділах буде розглянуто конкретну реалізацію цієї архітектури, вибір моделей для стейкінгу, параметри навчання, а також способи оцінки ефективності побудованої системи. Особлива увага приділятиметься способам декодування, таким як greedy search, beam search та стохастичні методи, що також можуть бути інтегровані як частина постобробки результатів для формування точних та інформативних рефератів медичних записів.

Процес повторюється до отримання спеціального токена завершення. Хоча жадібний пошук є простим та ефективним з обчислювальної точки зору, він може призводити до субоптимальних рішень через свою жадібну природу.

Більш складним, але часто більш ефективним методом є декодування з пучком [76] (beam search decoding). Цей підхід розглядає не один найбільш імовірний токен на кожному кроці, а зберігає безліч k найкращих часткових гіпотез (пучок). На наступному кроці для кожної гіпотези з пучка розраховуються ймовірності всіх можливих наступних токенів, і топ-k нових гіпотез додаються до пучка. Процес повторюється до повного завершення генерації рефератів для всіх гіпотез з пучка. Кінцевий реферат вибирається як гіпотеза з найвищою загальною ймовірністю. Beam search дозволяє здійснювати більш ретельний пошук у просторі можливих рішень, хоча і вимагає більше обчислювальних ресурсів.

Стохастичне декодування [77] (sampling-based decoding) є альтернативним підходом, який часто використовується для збільшення різноманітності згенерованих рефератів. Замість детермінованого вибору найбільш імовірного токена, при стохастичному декодуванні токени вибираються випадковим чином відповідно до розподілу ймовірностей,

передбаченого моделлю. Існують різні варіанти цього методу, такі як температурне вибіркове кодування (temperature sampling) та ядерне вибіркове кодування (nucleus sampling).

Температурне вибіркове кодування використовує параметр "температури" [78] (універсальні ваги) для керування ступенем різноманітності вибірки. При високих температурах розподіл ймовірностей стає більш рівномірним, що призводить до генерації більш різноманітних, але потенційно менш якісних рефератів. При низьких температурах модель поводиться більш передбачувано, обираючи найбільш імовірні токени. Ядерне вибіркове кодування обмежує вибірку Top-P parameter відсоткахми найбільш імовірних токенів, усуваючи малоймовірні та потенційно шумні токени [79].

Крім методів декодування, можуть застосовуватися додаткові стратегії для покращення якості та характеристик згенерованих рефератів. Наприклад, кількісне пенапенальне кодування [80] (length penalty) може використовуватися для заохочення генерації рефератів певної цільової довжини. Покрокове штрафне кодування (coverage penalty) допомагає моделі краще охоплювати різні аспекти вхідного тексту в рефераті, уникаючи надмірного повтору чи пропуску важливої інформації.

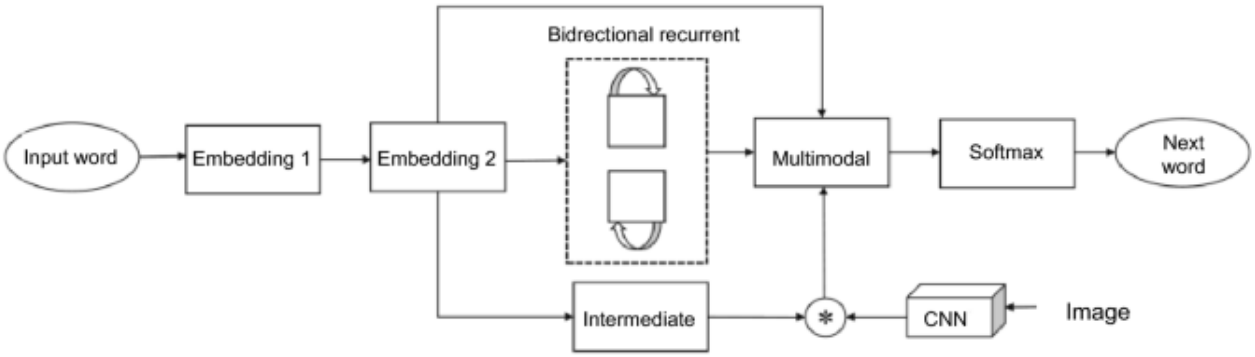


Рисунок 2.7 – Архітектурне вирішення на основі подвійного шару кодування з використанням двонаправленої рекуретної нейромережі

Окрім того, можуть застосовуватися фільтри або перевірки для гарантування відповідності згенерованих рефератів певним вимогам або обмеженням. Наприклад, фактчекінг для перевірки достовірності наведених у рефераті тверджень, фільтрація небажаного контенту (токсичного, образливого тощо) або перевірка граматичної та стилістичної коректності [81].

Для покращення продуктивності та надійності моделі ми застосували низку методів доповнення даних, спрямованих на збільшення різноманітності навчальних даних та підвищення точності виявлення в умовах низької освітленості [82]. Основними методами є наступні:

1. Зображення перевертаються горизонтально або вертикально, щоб врахувати асиметрію зображень і варітивність сценаріїв освітлення.
2. Випадкові обертання застосовуються до зображень, що сприяє здатності моделі узагальнювати під різними кутами.
3. Фільтри на основі ядра використовуються для імітації руху та підкреслення важливих особливостей.
4. Збільшення різкості покращує зображення, виділяючи деталі вибоїн, допомагаючи розпізнавати краї та текстуру.
5. Фільтри розмиття, такі як гаусівське розмиття, імітують розмиття при русі та шум при слабкому освітленні, допомагаючи моделі розпізнавати ключових ознак в умовах злегка розмитого зображення.
6. Щоб навчити модель виявляти ключові ознаки, незважаючи на часткову оклюзію, ми використовували методи, які видаляють або маскують ділянки зображення.

Кожна з цих методик доповнення підвищує надійність моделі, дозволяючи їй більш ефективно узагальнювати реальні ситуації та керувати складними умовами, такими як низьке освітлення, розмиття руху та часткова оклюзія [83]. Така комбінація доповнень призводить до покращення продуктивності моделі, що призводить до значного підвищення точності та надійності. На рисунку 2.8, можна детальніше з запропононованим принцип

аггрегації мультимодальних даних на основі змішування з застосування глибинних нейронних мереж.

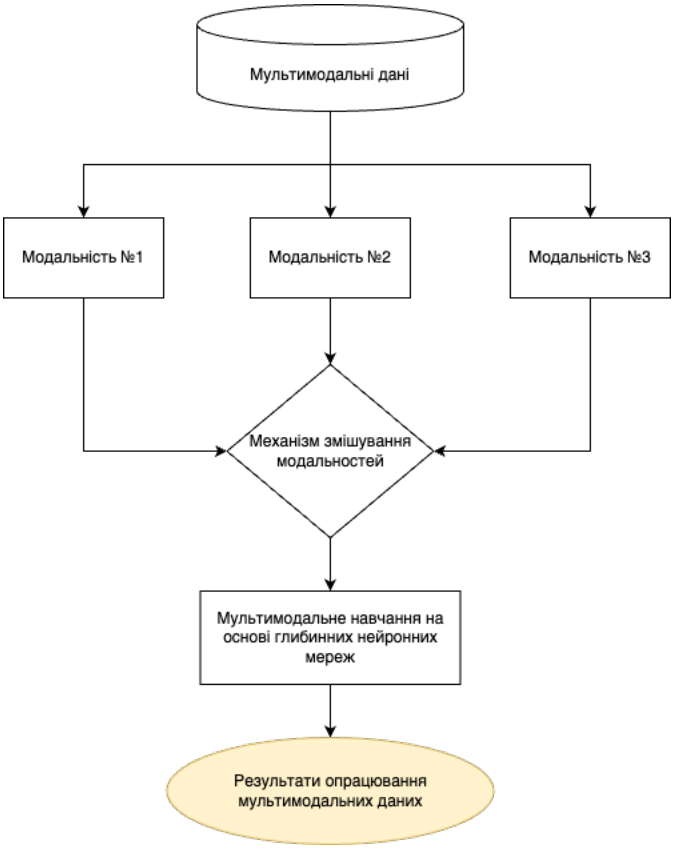


Рис. 2.8 – Принцип аггрегації мультимодальних даних на основі змішування та глибинних нейронних мереж

2.5. Висновки

У даному розділі проведений детальний аналіз методів комплексування даних в мультимодальних системах, описано та проаналізовано алгоритми приведення даних до єдиного вигляду, необхідних для роботи з мельтимодальними наборами.

Розглянуто формальну структуру представлення модальних даних, що забезпечує узгоджене й уніфіковане кодування інформації для подальшої обробки. Це дало змогу забезпечити сумісність даних між різними моделями та підмоделями, які входять до складу ансамблю.

Проведено порівняльний аналіз традиційних ансамблевих підходів, таких як беггінг, бустинг, стекінг, тощо. Переваги та обмеження кожного з методів

було проаналізовано з урахуванням завдань класифікації на основі мультимодальних даних. На основі цього аналізу зроблено вибір на користь гібридної архітектури, яка поєднує сильні сторони зазначених підходів.

У результаті дослідження було запропоновано архітектуру ансамблю, що включає базові моделі, спеціалізовані для кожної модальності, а також метамодель або механізм гейтера для злиття результатів. Запропонована модель забезпечує ефективне об'єднання репрезентацій з різних модальностей, покращуючи точність класифікації та стійкість до варіативності даних.

Таким чином, у розділі сформовано концептуальну та архітектурну основу розроблення ансамблю моделей машинного навчання, що буде використаний у наступних етапах експериментального дослідження..

Результати розділу опубліковано у працях автора [84, 85, 86, 87].

РОЗДІЛ 3. РОЗРОБЛЕННЯ МЕТОДІВ АНАЛІЗУ МОДАЛЬНИХ ДАНИХ

У розділі розроблено алгоритм структуру мультимодальних даних, концепцію аналізу і попередньої обробки кожної з вхідних модальностей, алгоритм кодування-декодування даних, алгоритм зведення модальних даних до єдиного вигляду, а також метод синхронізації даних за часовою ознакою з можливістю перевірки якості внесених даних.

Матеріали розділу опубліковані у роботах автора [121, 122, 123].

3.1. Побудова структури мультимодальних даних

У сучасних умовах стрімкого розвитку цифрових технологій мультимодальні дані стали невід'ємною складовою багатьох дослідницьких та прикладних задач. Інформація, що надходить із різних джерел — відео, аудіо, тексту, сенсорних систем тощо — містить цінні, але різнорідні сигнали, які потребують узгодженого опрацювання. Ефективна робота з такими даними вимагає побудови структурованого підходу до їх представлення, що дозволяє поєднувати та аналізувати інформацію з різних модальностей як єдине ціле.

Побудова структури модальних даних є критично важливою для забезпечення якості подальших обчислювальних процедур — зокрема, для застосування глибинного навчання, машинного зору, обробки природної мови чи сенсорного аналізу. Вона дозволяє не лише узгоджувати різні типи сигналів у спільному інформаційному просторі, але й створює підґрунтя для виявлення складних взаємозв'язків між ними. У цьому підрозділі розглядаються основні підходи до формування такої структури, а також виклики, що постають при інтеграції різних типів даних у межах мультимодального аналізу.

Для комплексного аналізу стану виробничих процесів та обладнання пропонується інтегроване використання трьох модальностей даних — зображень, звуку та тексту, кожна з яких потребує специфічних методів попереднього опрацювання та кодування.

Зображення (фото та відео з камер) застосовуються для виявлення поверхневих дефектів, таких як тріщини, вм’ятини або неправильні з'єднання деталей. Основними методами кодування є згорткові нейронні мережі (CNN) для екстракції візуальних ознак, моделі сегментації [88] (U-Net) для виділення дефектних зон та спеціалізовані підходи візуального аналізу аномалій, як-от MVTec AD [89].

Аудіодані (сигнали з вібраційних сенсорів та мікрофонів) використовуються для виявлення аномальних шумів або збійних режимів роботи обладнання. Для кодування аудіосигналів застосовуються спектральні методи, зокрема Mel-спектрограми та MFCC, а також глибокі моделі, такі як WaveNet [90] або рекурентні мережі LSTM, які дозволяють виявляти часові патерни у звуках.

Текстові дані (звіти, технічна документація, записи у лог-файлах та інша суміжна і похідна інформація) надають інформацію про помилки, події та суб’єктивні оцінки стану обладнання. Для кодування текстової інформації використовуються сучасні трансформерні архітектури (наприклад, BERT, T5), методи аналізу тональності, а також інструменти обробки природної мови (NLP) [91] для витягування ключових фактів та контексту.

Кожна модальність проходить свій етап кодування ознак, після чого їх репрезентації інтегруються для побудови єдиної моделі, що дозволяє врахувати всі аспекти поведінки виробничого обладнання. Комбінація даних трьох модальностей – зображення, текст та аудіо, дозволяє отримати повнішу картину стану обладнання та процесів, підвищуючи точність прогнозування несправностей та ефективність різних сфер людської життєдіяльності.

Використання технологій штучного інтелекту дозволяє зменшити залежність від людського фактору та забезпечити швидке реагування на потенційні несправності. Основними підходами у промисловій сфері є мультимодальне навчання, аналіз аномалій, прогнозування несправностей

(використання алгоритмів глибокого навчання для виявлення відхилень від нормальних параметрів, прогнозування поломок на основі історичних даних) та інтелектуальна автоматизація.

Запропонована система здатна не лише виявляти потенційні відхилення або аномалії в роботі об'єктів моніторингу, а й автоматично формувати рекомендації для користувача щодо подальших дій — наприклад, щодо корекції параметрів чи запобіжних заходів. Задачі адаптивного управління й оптимізації складних процесів можуть ефективно розв’язуватись шляхом застосування методів навчання з підкріпленням, що дозволяє системі самостійно знаходити найкращі стратегії дій у змінному середовищі.

Автоматизація прийняття рішень допоможе суттєво зменшити час реакції на проблеми та покращити загальну ефективність підприємства [92].

Щоб довести ефективність запропонованого підходу, необхідно розробити нейромережеву архітектуру, здатну обробляти мультимодальні дані. Така архітектура має містити окремі спеціалізовані блоки для обробки різних типів даних: згорткові нейронні мережі (CNN) для виділення ознак із зображень, рекурентні нейронні мережі або їх модифікації (RNN/LSTM) для обробки текстових даних і часових послідовностей, а також аудіоаналітичні модулі на основі спектрограм для роботи зі звуковими даними. Для ефективної інтеграції ознак з різних модальностей мають використовуватися сучасні механізми об'єднання репрезентацій [93] (fusion models), наприклад, Cross-Attention Transformers [94] або мультимодальні варіації моделей BERT, що забезпечують узгодження різних типів ознак на рівні глибинних залежностей.

Для навчання такої моделі потрібно використовувати реальні датасети, які містять реальні сценарії у роботі з тої чи іншої сфери опрацювання. Додатково пропонується застосувати генерацію синтетичних даних для розширення варіантів можливих недорепрезентованих класів, що дає змогу покрити ширший спектр потенційних ситуацій у кожній з досліджуваних сфер [95].

Оцінка якості роботи моделі повинна базуватися на таких метриках, як точність, влучінсть, повнота, F1-міра. Для перевірки практичної ефективності пропонується провести реальне тестування моделі у реальноусу середовищі, інтегрувавши її до системи визначення психоемоційного стану пацієнта перед проведенням оперативного втручання в мережі приватних клінік «Родина».

Математична модель базується на припущенні про наявність множини вхідних даних, які кодуються векторними представленнями у відповідних просторах ознак, після чого інтегруються в єдиний вектор унікальних ознак для подальшого прогнозування технічного стану або виявлення дефектів за допомогою мета-моделі машинного навчання. Детальніше ознайомитися з конвеєром опрацювання унімодальних даних на рисунку 3.1.

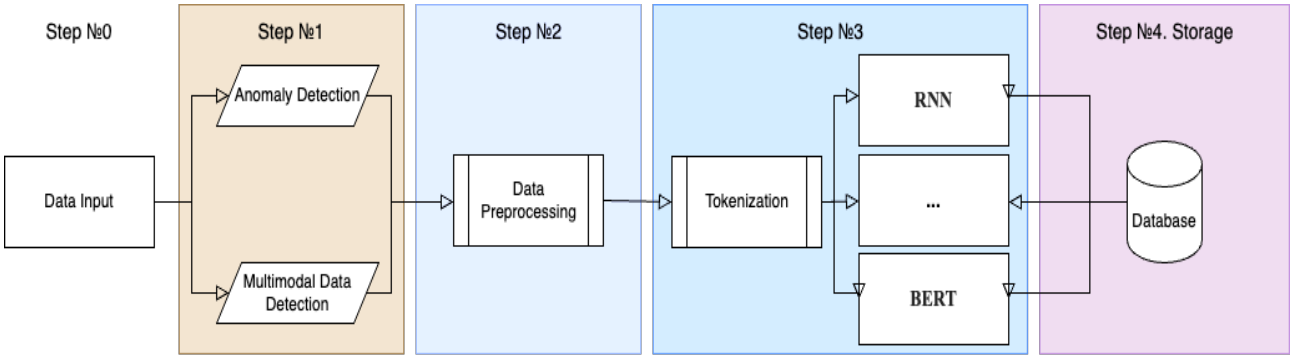


Рисунок 3.1. Приклад роботи структури унімодальних даних

3.2 Розроблення концептуальної моделі аналізу мультимодальних даних

Під час роботи з аудіоданими в задачах виявлення депресії важливо правильно підготувати дані, щоб забезпечити точність моделі. Процес попередньої обробки аудіо включає кілька важливих кроків, кожен з яких спрямований на покращення якості даних і підвищення ефективності подальшого аналізу ознак (рис. 3.2).

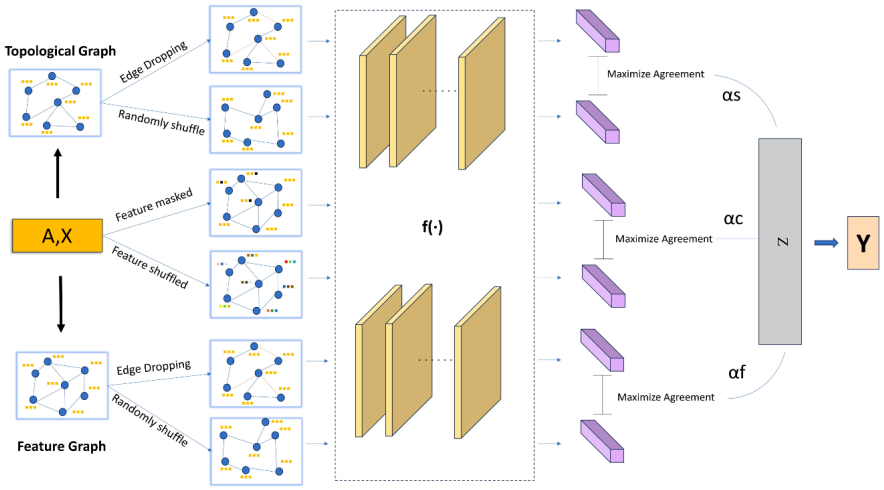


Рисунок 3.2. Схема попередньої обробки модальних даних [96]

Три архітектурні рамки виникли з найсучаснішого стану для розробки єдиної багатокатегорійної методи. Перший підхід включає багатозадачні методики навчання, запропоновані U2Fusion Xu et al. (2020a) [97]. Мета - розробити більш комплексну архітектуру, здатну працювати з класичними завдання об'єднання зображень, одночасно інтегруючи загальні компоненти для вирішення проблем, пов'язаних із кількома видами. Через ітераційне навчання для визначених завдань можна розробити єдиний набір ваг, що дозволить архітектурі керувати різними завданнями ефективно. Цей тип архітектури потребує механізмів для контролю передачі інформації між навчаннями завдання та запобігання забуванням, подібні до тих, що пропонуються в існуючих методах [98]. Однак цей підхід керує різними виконує завдання окремо під час логічного висновку і не поєднує інформацію між різними категоріями.

Такі архітектури, як MVMM-NET Song та ін. (2021) [99] призначені для повного керування об'єднанням зображень у кількох категорій, включаючи злиття з кількома видами. Ця друга категорія методів реалізує модель, здатну до кодування зображення з різних категорій і використання їх у складній системі злиття. Однак обмеження запропонований метод полягає в його

полегшеній внутрішній архітектурі, що може знизити його ефективність. Архітектури майбутнього для уніфікованого багатокатегорійного злиття слід застосовувати більш надійні конструкції, які можуть точно представляти та об’єднувати інформацію,

потенційно включають більш складні методи, такі як глибоке навчання.

Останній тип структури — це спеціалізована архітектура, яка забезпечує об’єднання зображень через уніфіковану структуру, як Сучасний стан демонструє, що методи зіставлення функцій можуть бути ефективними для мультимодального злиття зображень, як це видно в Rift (2019) [100] та RDFM (2023) [101]. Крім того, методи зіставлення функцій, такі як OmniGlue (2024) [102] і SuperGlue (2020) [103] довели бути високоефективним при обробці даних з кількох переглядів.

Ці висновки свідчать про те, що підходи до зіставлення ознак справедливі значний потенціал для розробки спільної архітектури, застосовної до всіх цих категорій. Цей тип єдиної моделі архітектура представляє багатообіцяючий напрямок, як і архітектури на основі трансформаторів, які привертають увагу багатьох.

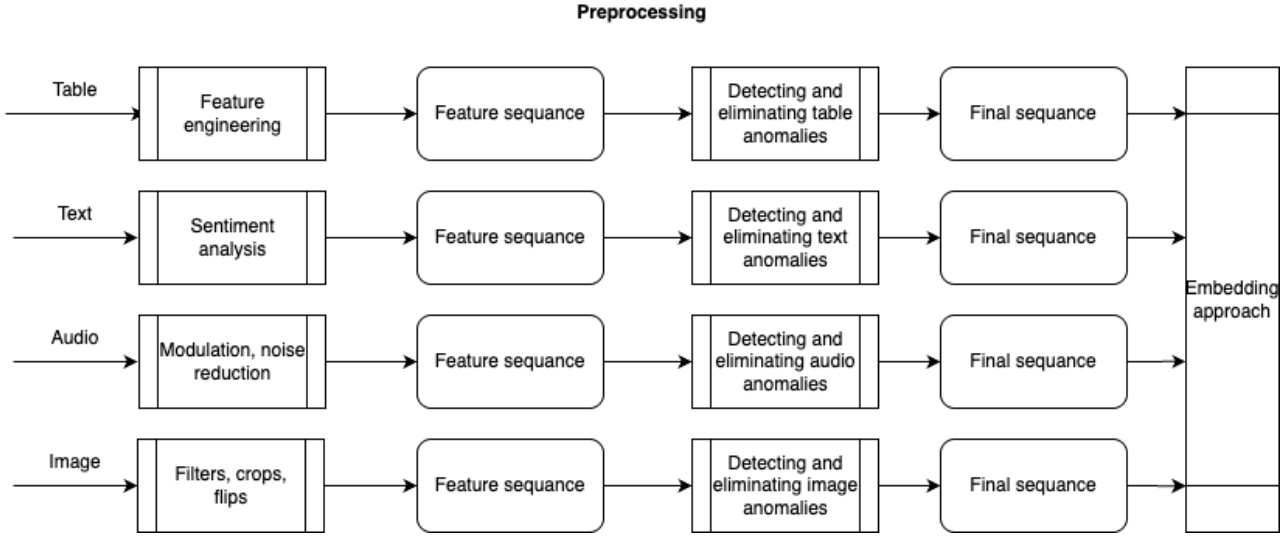


Рисунок 3.3. Запропонована схема моделі аналізу модальних даних, на основі попереднього опрацювання, пошуку анамалій з подальшим злиттям в єдиний вектор унікальних ознак.

У цьому розділі розглянуто оцінку запропонованого методу перекладу у двох багатомовних сценаріях — «один до багатьох» та «багато до багатьох». Метою є продемонструвати, як запропонований підхід, заснований на каскадних обчислювальних графах, покращує якість перекладу шляхом ефективного розширення мовних підвузлів і дублювання вузлів зі спільними знаннями.

У першому сценарії — «один до багатьох» — аналізується здатність моделі здійснювати переклад з однієї мови на кілька інших. Особливу увагу приділено випадкам, коли мови-реципієнти є спорідненими або мають схожі риси з мовою-джерелом. Результати демонструють, що запропонований метод перевершує базові моделі на більшості мовних пар у трьох різних наборах даних: TED2013 [104], TED2020 [105] та BIBEL [106]. Це покращення пов’язується з можливістю моделі ефективно працювати навіть із малоресурсними мовами, не покладаючись на мовну спорідненість. Базова модель MNMT [107], натомість, демонструє нижчі результати через негативну мовну інтерференцію, спричинену спільним параметричним простором.

У другому сценарії — «багато до багатьох» — аналізується здатність моделі здійснювати переклад між усіма можливими комбінаціями мов. Цей підхід дозволяє оцінити взаємозв’язки між різними мовами в обох напрямках, а також загальну стабільність методу в умовах багатомовності. У цьому випадку модель також демонструє значно кращі результати порівняно з базовими системами, що пояснюється використанням позитивно споріднених мов і гнучким дублюванням вузлів, які зміцнюють структуру перекладу. Особливо помітно це у перекладах між англійською та перською мовами, де модель стабільно випереджає інші підходи у всіх наборах даних [108].

Таким чином, результати оцінювання підтверджують ефективність запропонованого підходу як у спрямованих перекладах з однієї мови на багато, так і в умовах повної багатомовної взаємодії.

3.3 Розроблення методу обробки різномодальних даних

При опрацюванні аудіоданих в завданнях класифікації надзвичайно важливо провести ефективну попередньо підготовку аудіозаписів. Якісий препроцесигн дозволить забезпечити високу точність моделі кодування даних, а також ефективне подальше опрацювання унікальних векторів ознак структурами повнjзв'язних шарів нейронної мережі. Попереднє опрацювання аудіоданих складається з декількох важливих етапів, на кожному з яких фокусуємося на підвищенні якості і ефективності екстракції подальшого вектору унікальних ознак (рис. 3.4).

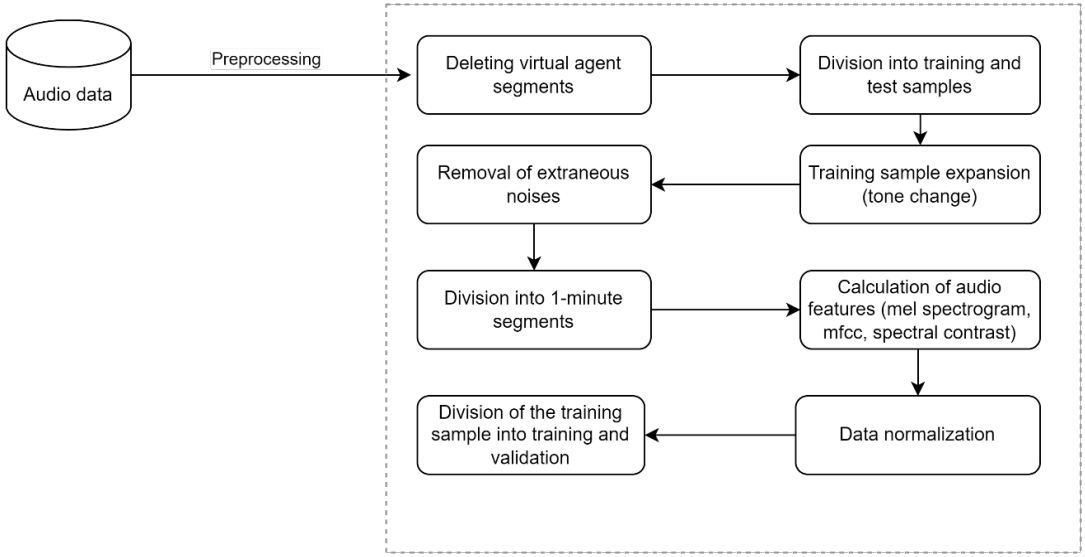


Рисунок 3.4. Алгоритм попередньої обробки аудіоданих.

Згідно з рисунком 3.4, етапи попередньої обробки аудіоданих складаються з наступних кроків:

На першому етапі опрацювання даних, нам важливо видалити аномальні та інші штучні сегменти запису, для прикладу питання від інтерв’ювера або штучнозгенеровані аудіоповідомлення опитувальника.

Подальший кроком, це підготована розподілу тренувальної та тестувальної вибірки. Тренувальну підмножину додатково розділяють на навчальну та валідаційну частини у пропонції 80% вибірки - це тренувальні дані, а решта 20% - це дані для тестування ефективності роботи натренованої

моделі. Саме цей підхід, дозволяє в процесі навчання ефективно оцінювати результативність моделі на нових даних і вчасно коригувати її параметри та гіперпараметри.

Методи аугментація даних - це ефективний інструмент для збільшення обсягу тренувальних даних. Окрім цього, даний підхід дозвляє збалансувати недорепрезентовані класи нашої вибірки. При опрацювання аудіоповідомлень, простим і ефективним методом опрацювання даних є зміна толнальності запису. Для прикладу, застосування методів для підвищення або пониження тональності повідомлення, дозволяє наповнювати вибірку новими записами, без втрати змісту на сенсів закладених у ньому. Важливо, що така незначна зміна тональності зберігає природність звучання голосу й водночас розширює звукове представлення даних.

Наступним етапом є очищення від шумових артефактів та розбиття на сегменти. Саме на цьому етапі відбувається видалення фонового шуму а саме, гудіння, сторонні голоси, інші аудіоперешкоди, тощо. Виконання частотної модуляції аудоповідомлення дозволяє покразити чистоту записів, що сприє подальшому ефективному витягуванню унікальних ознак з похідних спектрограм.

У подальшому, кожен очищена аудіодоріжка розділяється на уривки тривалістю одна хвилина, у межах нашого дослідження, встановлено, що фрагменти довжиною 5-10 секунд, не є достатньо інформаційно наповненими. Це в свою чергу, не дає можливості їх ефективно використовувати для подальшого аналізу в рамка. Розподіл на фрагменти довжиною в одну астрономічну хвилиниу дозволяє краще розуміти контекст записів, та відстежувати зміни у манері мовлення різних користувачів.

Для кожного аудіоповідомлення виконується обчислення аудіохарактеристик, а саме відбувається поділ на окремі сегменти для кожного з яких розраховуються параметри, мел-спектрограми, мел-частотні кепстральні коефіцієнти (MFCC) і спектральний контраст. Мел-спектрограма відображає

розподіл енергії по частотах протягом часу і дозволяє відстежувати зміни голосу та інтонації. MFCC представляють числову характеристику спектру мовлення, особливо корисну для опису артикуляційних особливостей мовця, таких як ритм, тон і темп. Спектральний контраст фіксує відмінності між піковими та мінімальними амплітудами у різних діапазонах, що допомагає виявляти монотонність або знижену експресивність аудіоповідомлень.

Усі ці характеристики використовуються як вхідні ознаки для подальшого навчання моделі.

Для стабільної роботи моделей машинного навчання, ефективно застосувати нормалізацію даних, а саме зводити дані до однаковий масштаб, через операцію віднімання середнього та поділу на стандартне відхилення, саме цей етап є важливим для забезпечення точності багатьох методів машинного навчання. Особливо у випадках, коли передбачається застосування ансамблів.

Отже, послідовне виконання всіх етапів забезпечує якісну підготовку аудіоданих для моделювання і успішного виявлення ключових ознак для подальшого класифікації даних на рівні мета-моделей з запропонованого підходу суміші експертів.

Окрім деталізація методу опрацювання аудіоповідомлення в рамках апробації отриманих результатів, вважаю за доцільне детальніше розгляднути важливі кроки у попередні обробці текстової інофрмації (див. рисунок 3.6).

Вона дозволяє сформувати текстові підказки для подальшого виділення релевантних ознак і очищає дані від надлишкової або неінформативних даних. Це допомагає моделі фокусуватися лише на суттєвих даних і важливих знаннях, що сприяє підвищенню ефективності навчання моделей машинного навчання й забезпечує зростання точності фінальних прогнозів.

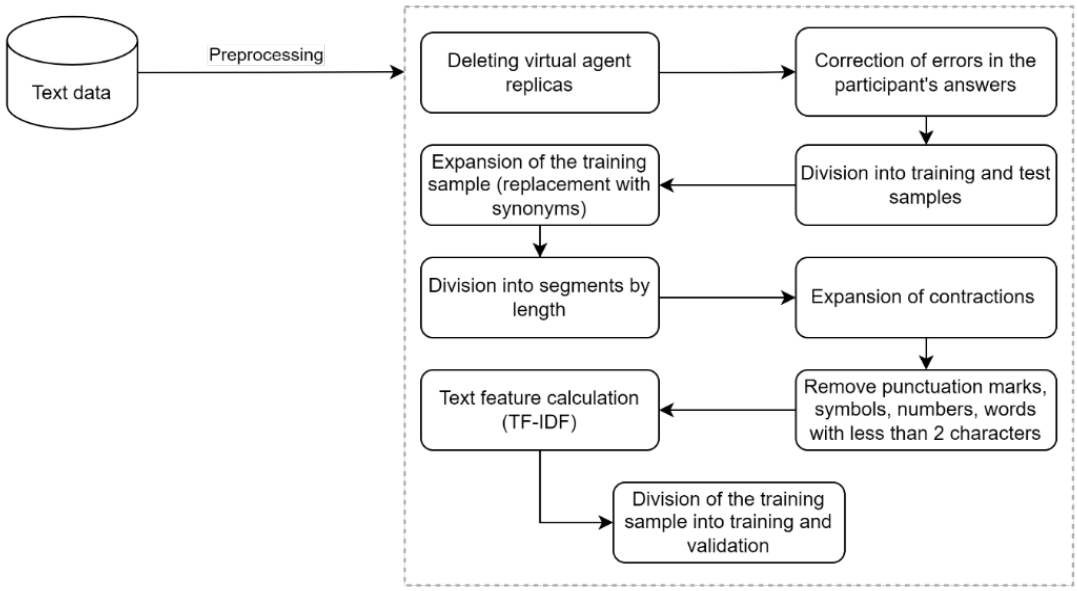


Рисунок 3.5. Алгоритм попередньої обробки текстових даних.

Окрім деталізації методу опрацювання аудіоповідомлень, важливо детальніше розглянути кроки попередньої обробки текстової інформації (див. рисунок 3.5). Попередня обробка текстів дозволяє сформувати чіткі та релевантні підказки для моделі, очищаючи дані від зайвих або неінформативних елементів. Це допомагає підвищити точність навчання моделей машинного навчання і сприяє покращенню якості фінальних прогнозів.

На першому етапі попередньої обробки здійснюється видалення неключових даних, щоб зосередити аналіз виключно на опрацюванні важливих частин інофрипції. У випадках, коли анотації неточні, ми можемо застосовувати метод ручного видання даних, який дозволяє нам здійснити видалення нерелевантого тексту досягнення вищого рівння узгодженості між текстом і аудіозаписом.

Наступним кроком є виправлення помилок у відповідях учасників. Транскрипти перевіряються вручну для виправлення розбіжностей між фактичними висловлюваннями та текстовим представленням. Це забезпечує підвищення точності аналізу та мінімізує ризик викривлення даних у процесі моделювання.

Подальший етап передбачає поділ текстових даних на навчальну та тестову вибірки у співвідношенні 80% навчальних даних, та 20% тестувальних. Додатково тренувальну вибірку ділять на навчальну та валідаційну частини, що дає змогу об'єктивно оцінювати ефективність моделі та своєчасно коригувати її гіперпараметри у процесі навчання.

Аугментація навчальної вибірки є важливою складовою попередньої обробки. Для збільшення обсягу тренувальних даних використовується заміна слів синонімами за допомогою лексичної бази WordNet [109]. Ймовірність заміни складає 15%, що є оптимальним балансом між збереженням змісту тексту та забезпеченням варіативності. При цьому виключаються критично важливі слова та стоп-слова, аби не змінити семантичну структуру, важливу для аналізу психічного стану.

Після аугментації здійснюється розбиття тексту на сегменти відповідно до кількості сегментів в аудіоданих. Такий підхід дозволяє синхронізувати текстову та аудіоінформацію, що особливо важливо для мультиформатного аналізу даних.

В рамках роботи з текстовими даними, також можуть бути розширені окремі фрагменти шляхом додавання додаткових запитань або уточнень. Це сприяє отриманню більш повної картини й дозволяє моделі краще вловлювати важливі деталі у висловлюваннях учасників.

Очищення тексту є обов’язковим кроком для підвищення якості даних. Видаляються розділові знаки, спеціальні символи, цифри, а також слова, що складаються менше ніж із двох символів. Це дозволяє усунути шум та зосередитися на змістовній частині текстів.

Далі здійснюється розрахунок текстових ознак шляхом використання методу TF-IDF. Для підвищення релевантності текстових представлень додатково збільшується вага слів, пов'язаних із депресивним станом. Формується список ключових слів, який розширюється семантично близькими

словами за допомогою попередньо навчених векторів GloVe, що забезпечує більш глибоке розуміння змісту тексту.

На завершення, після всіх етапів попередньої обробки, навчальні дані остаточно поділяються на навчальну і валідаційну підвибірки у співвідношенні 80% тестових даних, та 20% валідаційних. Це дозволяє ще на етапі навчання відстежувати якість роботи моделі і вчасно коригувати її для досягнення максимальної точності прогнозів.

На рисунку 3.6 схематично наведено три підходи до об'єднання даних різної модальності:

a) Раннє злиття (Early Fusion) передбачає об'єднання сирих або низькорівневих ознак ще до основного навчання моделі.
b) Пізнє злиття (Late Fusion) виконується після того, як окремі модальності оброблені власними моделями — результати інтегруються на рівні прийняття рішення.
c) Гібридне злиття (Hybrid Fusion) комбінує переваги обох підходів, дозволяючи інтеграцію як на рівні ознак, так і на рівні рішень.

Ця схема ілюструє гнучкість системи у виборі оптимальної стратегії злиття залежно від поставленої задачі та характеристик даних.

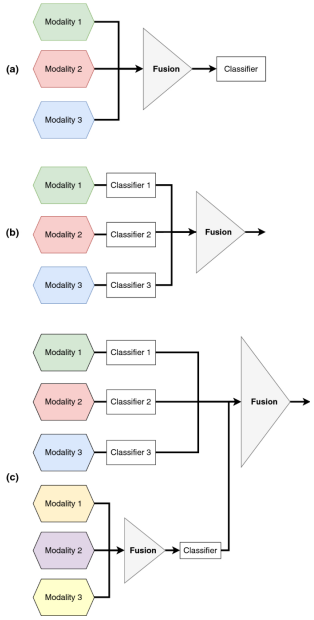


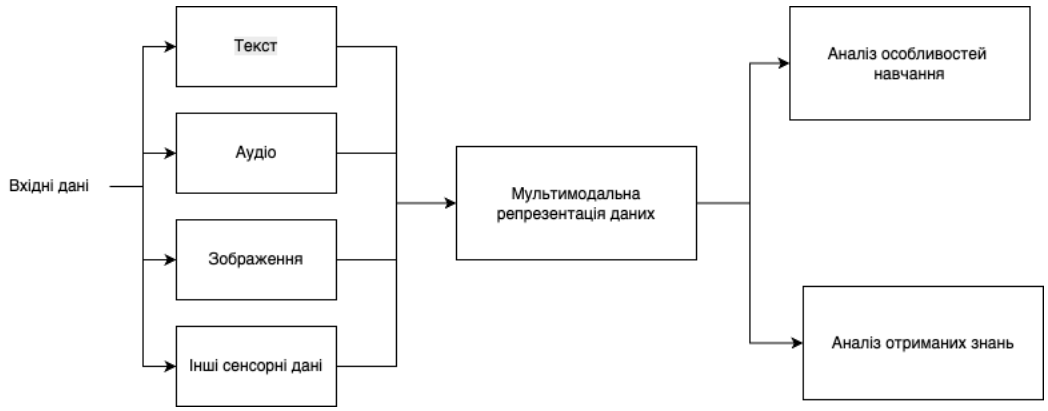
Рисунок 3.6. – Схема роботи різних типів змішувань [110]

3.4 Алгоритм зведення мультимодальних даних до єдиного вигляду

Якість даних – узагальнене поняття, що відображає ступінь їх придатності для вирішення певного завдання. Відповідно до стандартів основними критеріями якості є повнота, достовірність, точність, узгодженість, доступність та своєчасність.

Оцінка якості даних та дії щодо його підвищення є необхідним етапом будь-якого аналітичного проекту, оскільки аналітичні алгоритми або не зможуть працювати з неякісними даними або будуть давати некоректні результати.

У мультимодальних системах штучного інтелекту, де обробка даних здійснюється з різних джерел — таких як аудіо, текст, зображення тощо — ключовим завданням є зведення цих гетерогенних модальностей до єдиного вигляду для подальшої інтеграції та аналізу. Існує кілька стратегій об'єднання модальних ознак, серед яких раннє злиття (early fusion) та пізнє злиття (late fusion) є найпоширенішими.



Рисинок 3.7 – Артектурне вирішення моделі злиття даних

Стратегії Fusion (злиття, або змішування) об’єднують інформацію з різних кодувальників, гарантуючи, що багатомодальна модель використовує сильні сторони кожної модальності:

- Раннє злиття: інтегрує необроблені дані або початкові функції з кожної модальності перед тим, як пропустити їх через модель, сприяючи ранній взаємодії між модальностями.

- Пізнє злиття: об’єднує результати індивідуальних модально-специфічних моделей, дозволяючи кожну модальність обробляти незалежно перед інтеграцією.
- Гібридне злиття: поєднує аспекти як раннього, так і пізнього злиття, щоб збалансувати інтеграцію модальностей на різних етапах обробки.
- Крос-модальне злиття: виходить за рамки простого поєднання, часто використовуючи увагу або інші механізми для динамічного зв’язку характеристик із різних модальностей, підвищуючи здатність моделі фіксувати міжмодальні зв’язки.

Архітектура моделі раннього змішування, показана на рисунку 3.8, поєднує унікальні вектори аудіо та тексту на ранніх стадіях опрацювання, що дозволяє ефективно об’єднувати бімодальні дані для підвищення точності результатів фінального класифікатора.

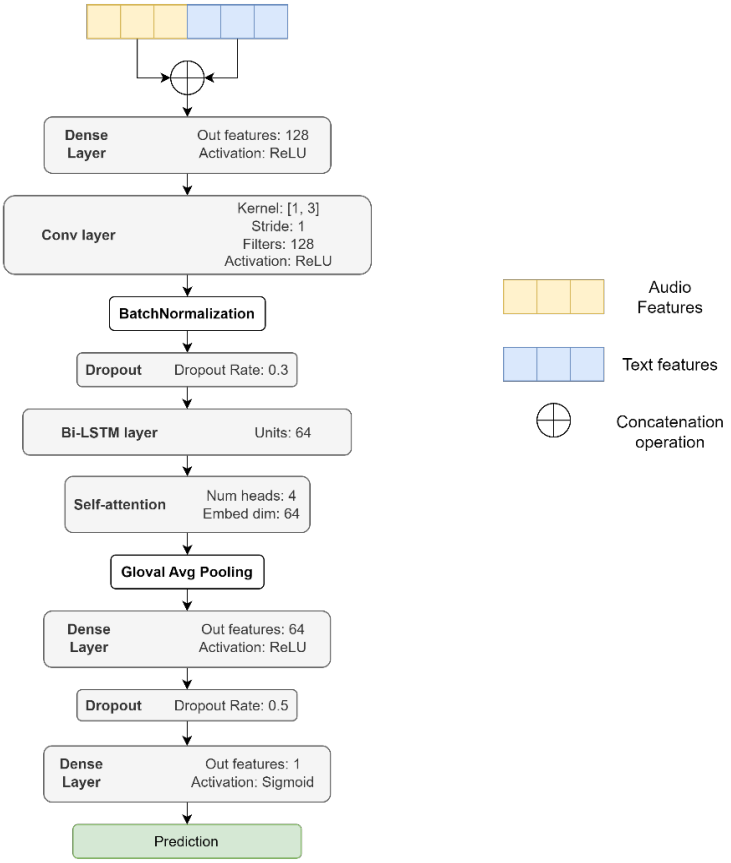


Рис. 3.9. Архітектура мультимодальної мережі раннього злиття.

Вхідні дані, які застосовувалися для валідації роботи системи, складаються з аудіо та текстових даних. Усі аудіодані наведні у форматі записів .mp3, що в подальшому буде перетворено методами препроцесингу у різнотипові спектрограми та карти спектральних контрастів, які узагальнюватимуть в собі акустичні особливості сигналів.

Текстові набори даних, це звичні для усіх послідовності символів, які опісля попереднього опрацювання набувають вигляду векторів унікальних ознак, отриманих за допомогою різних методів кодувальників, зокрема TF-IDF [111], Bag of Words [112], Word2Vec [113] тощо.

Вектори унікальних ознак кожної з опрацьованхи модаьностей об'єднуються через шар злиття (fusion layer), що забезпечує їх спільну подальшу обробку фінальним класифікатором.

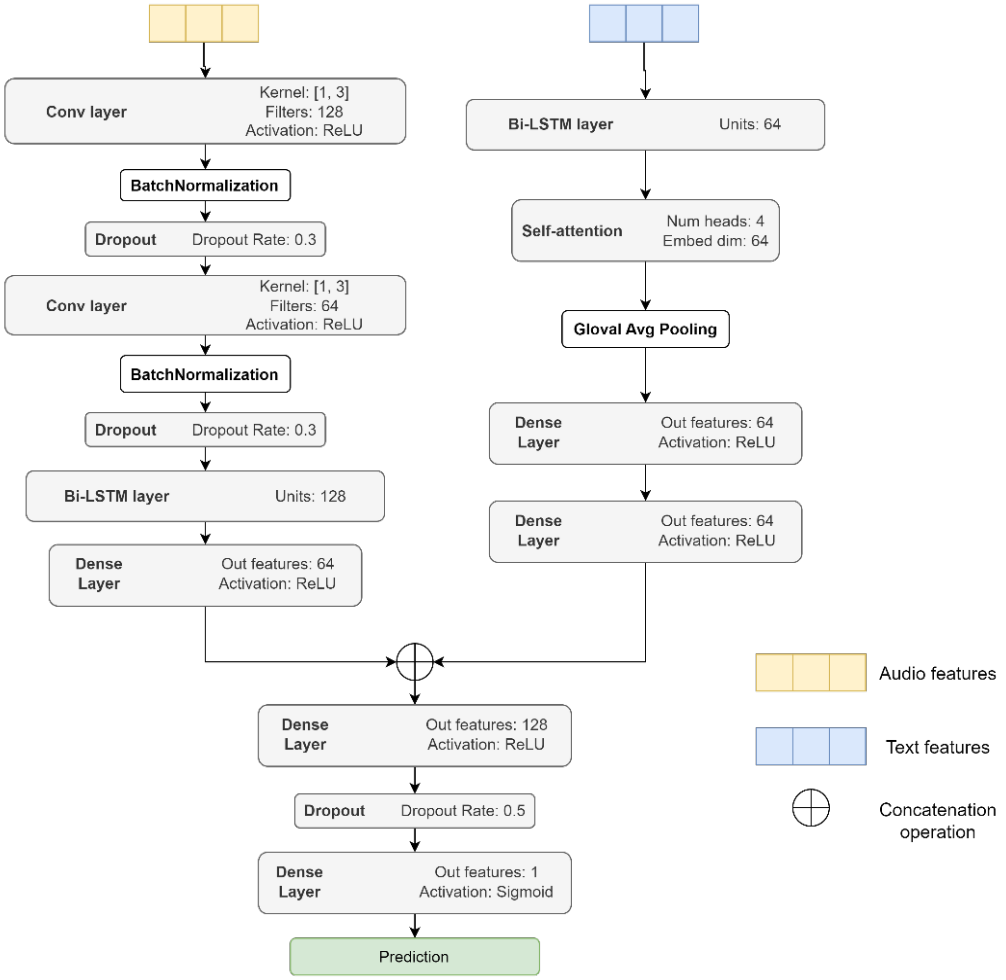


Рис 3.9 Архітектура мультимодальної мережі пізнього злиття.

Архітектура подальшого класифікатора складається з повнозв’язного шару розмірністю 128 нейронів, з подальшим застосуванням методу активацією ReLU. Даний метод активації, дозволяє зменшити складність даних, та нівелювати унікальні ознаки з від’ємним знаком, перетворивши їх у нульові знчення. Вихідний шар із функцією активації Sigmoid [114] визначає ймовірність наявності депресії. Архітектура моделі реалізована за принципом пізнього злиття, де обробка аудіо- та текстових ознак спочатку відбувається окремо, а об'єднання результатів здійснюється на пізньому етапі. Такий підхід дає змогу якісно виділяти специфічні ознаки кожної модальності, що в результаті покращує точність спільного аналізу.

Аудіодані опрацьовуються використовуючи підходи згорткових нейронних мереж, для вилучення локальних унікальних ознак, а потім Bi-LSTM [115] для фіксації довготривалих залежностей у послідовностях даних. Початкові згорточні шари застосовують фільтри для виявлення закономірностей в аудіо, а шари пакетної нормалізації та відсіювання підтримують стабільність і запобігають надмірному накладанню. Шар Bi-LSTM, що складається з 128 одиниць, обробляє послідовну природу аудіоданих, виділяючи часові зв'язки з різних сегментів. Останній щільний шар ще більше зменшує розмірність вилучених ознак, готуючи їх до поєднання з текстовими ознаками.

Для текстових даних використовується Bi-LSTM-шар з 64 одиницями для захоплення контексту в прямому і зворотному напрямках. Механізм самоуваги фокусується на найбільш релевантних частинах тексту, полегшуючи моделі визначення ключових ознак депресії. Глобальне середнє об'єднання зменшує кількість ознак, зберігаючи важливу інформацію для класифікації.

Після окремої обробки аудіо та текстові ознаки об'єднуються за допомогою шару конкатенації на пізній стадії злиття. Отриманий набір ознак потім пропускається через щільні шари з активацією ReLU для уточнення об'єднаних ознак. Остаточний вихідний шар з функцією активації Sigmoid

використовується для класифікації комбінованого представлення як такого, що вказує на наявність або відсутність депресії.

Запропоноване архітектурне вирішення, містить у собі такі особливості, спершу ми концентруємося на екстракції унікальних ознак у різних модальностях даних, для прикладі у текстових та аудіо даних, як це було наведено на прикладі вище, а опісля з застосування механізмів злиття та уваги, ми намагаємося ефективно об'єднувати і узагальнювати отриманні знання, що сприяти кращій продуктивності фінальної мета моделі.

На рисунку 3.10 зображено типовий конвеєр обробки мультимодальних даних, де на вхід системи надходять різні типи сигналів (текст, зображення, аудіо тощо). У процесі підготовки дані можуть проходити через генератори з додаванням шуму для підвищення стійкості моделі та імітації реальних сценаріїв. Далі кожен тип даних обробляється відповідним дескриптором — модулем, який виділяє ключові ознаки. Згодом ці ознаки подаються на предиктори, що виконують основне завдання прогнозування або класифікації. Такий підхід забезпечує адаптивність і підвищену точність у мультимодальних умовах.

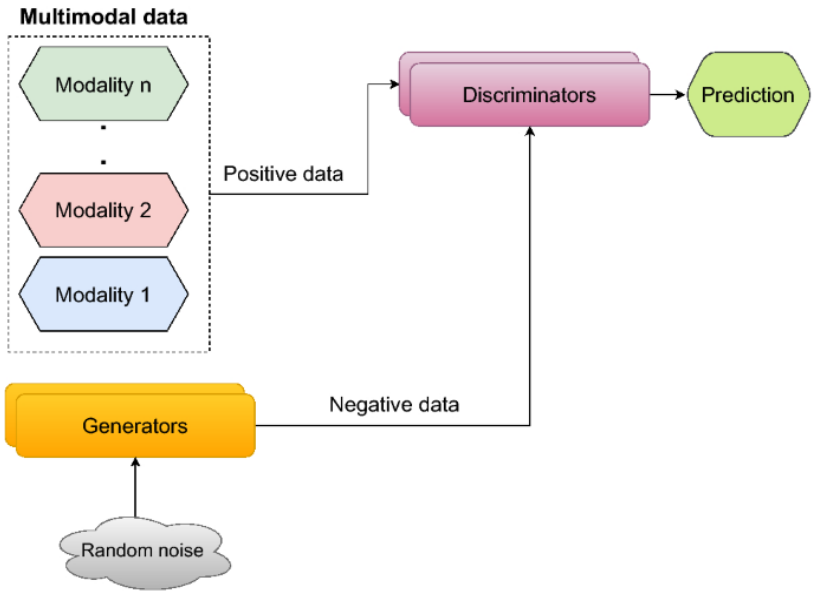


Рисунок 3.10 – Принцип роботи класифікатора на основі мультимоадлних даних [116]

3.5. Висновки

У даному розділі розробленено модель опрацювання мультимодальних українських наборів даних. Визначено основні операції над даними. Проаналізовано методи аналізу уніфікованих наборів даних. Розроблено метод синхронізації в часових наборах даних на основі аналізу часових міток та марковської моделі. Це дає змогу забезпечити цілісність даних та усунути надмірність у опрацьовуваних наборах.

Запропоновано ансамблевий підхід на основі методів машинного навчання для опрацювання мультимодальних українських наборів даних, що дозволило підвищити метрики точність опрацювання наборів даних.

Розроблений алгоритм перевірки якості опрацьованих даних, який базується на перевірці наявності аномалій, виявлення суперечливих даних та розсихнорнізованхи даних. Це дало змогу зменшити кількість запитів до системи кодувальника, а також знаходити наефективнішу модель кодування з переліку наявних і інтегрованих в рамках розробленої системи.

Результати розділу опубліковано у працях автора [117, 118, 119].

РОЗДІЛ 4. РОЗРОБЛЕННЯ АРХІТЕКТУРИ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ТА АПРОБАЦІЯ РЕЗУЛЬТАТІВ

У розділі розроблено архітектуру системи та подано її у вигляді діаграми класів. Розроблено протокол обміну даними. Здійснено апробацію результатів та визначено кількісних параметри розробленої інформаційної системи. Розроблено архітектуру опрацювання мультимодальних даних для медицини, яка використана в інформаційній системі визначення психоемоційного стану пацієнта, який очікує на оперативні втручання, впровадженій в приватному підприємстві "Мережа Медичних Центрів Родина".

Матеріали розділу опубліковані у роботах автора [120, 125, 129].

4.1 Розроблення архітектурного вирішення системи

Архітектура системи обробки мультимодальних даних побудована з урахуванням специфіки різнорідної інформації (текст, зображення, метадані тощо) та потреби в точному виявленні аномалій, покращенні якості даних і забезпеченні ефективної взаємодії між різними форматами інформації. З метою оптимізації побудови та реалізації проєкту архітектуру було поділено на шість функціонально взаємопов’язаних модулів, кожен з яких виконує окремі завдання в загальному конвеєрі обробки:

1. Модуль збору вхідних мультимодальних даних виконує роль інтерфейсу між системою та зовнішнім світом. Він відповідає за прийом даних з різних джерел: бази даних, API, файлових систем чи потокових сервісів. Підтримується широкий спектр форматів – текстові повідомлення, метадані, зображення, відео тощо. Для кожного типу даних реалізовані спеціалізовані адаптери, що забезпечують уніфікацію формату при первинному зборі.
2. Модуль виявлення аномалій та класифікації типу даних, після збору інформації дані передаються в модуль аналізу на наявність аномалій.

Використовуються алгоритми на зразок Isolation Forest або Isolation Scaffolding для виявлення викидів та структурних відхилень. Також здійснюється аналіз мультиформатності даних – система автоматично класифікує, які типи даних присутні, та ідентифікує міжмодальні взаємозв'язки. Це забезпечує коректну підготовку даних до наступних етапів обробки.

3. Модуль попередньої обробки та стандартизації, на цьому етапі дані очищуються, нормалізуються та перетворюються у зручний для моделювання формат. Числові значення масштабуються, текст очищується від шумів і токенізується, а зображення проходять фільтрацію та стандартизацію розміру. Система також передбачає механізми для заповнення пропущених значень та узгодження даних між різними модальностями.
4. Модуль векторизації (embeding) та модального злиття (fusion), ключовий аналітичний модуль, який перетворює стандартизовані дані у числові вектори для подальшої обробки. Здійснюється токенізація з урахуванням типу даних, після чого запускається серія попередньо натренованих моделей для кожної модальності (наприклад, BERT для тексту, ResNet або EfficientNet для зображень). Система підтримує три методи злиття результатів:
    a. Early Fusion – об'єднання даних ще до подачі у модель,
    b. Late Fusion – комбінування результатів окремих моделей після обробки,
    c. Hybrid Fusion – гібридний підхід з поетапною інтеграцією ознак.

5. Модуль збереження результатів і записування у базу даних, після завершення обробки результати (класифікації, детекції аномалій, вектори ознак) зберігаються у структурованому вигляді у базі даних або дата-лейку. Формується аналітичний датасет, який постійно оновлюється з кожним новим

запуском системи. Це дозволяє системі самонавчатись, адаптуватись до нових сценаріїв та забезпечує подальшу інтеграцію в BI або аналітичні платформи.

6. Модуль управління потоками та синхронізацією, центральний координаційний модуль, який забезпечує узгоджену роботу всіх компонентів системи. Виконує роль розподільника навантаження, контролює черговість обробки, обробку помилок та логування. Також відповідає за динамічну масштабованість та моніторинг ефективності модулів.

Таблиця 4.1 Порівняння результатів моделей злиття мультимодальних даних [120]

Модальності	**Пізнє злиття**	**Раннє злиття**	**Гібридне злиття**
Текст	0,891	0,846	0.812
Метадані	0.758	0.702	0.734
Зображення	0.769	0.846	0.747
Текст + Метадані (Бімодальні)	0.947	0.941	0.798
Метадані + Зображення (Бімодальні)	0.892	0.883	0.802
Текст + Зображення (Бімодальні)	0.911	0.906	0.846
Мультимодальні дані (Текст + Зображення + Метадані)	0.923	0.910	0.899

Запропонована мультимодальна система організована за принципами модульності, масштабованості та підтримки мультимодальної обробки даних. Архітектура побудована таким чином, що кожен функціональний блок є незалежним сервісом, інтегрованим у загальну систему за допомогою центрального керуючого ядра суміші експертів (подіхний підхід від стекінгу).

Конвеєр обробки даних починається зі збору вхідних потоків різної природи: зображень, аудіо-даних і текстових повідомлень. Кожен тип даних передається до відповідного модуля попередньої обробки та кодування:

- зображення — до блоку комп'ютерного зору (на основі CNN або сегментаційних моделей),
- аудіо — до модуля акустичного аналізу (з використанням спектрограм і нейронних мереж),
- текст — до NLP-блоку на базі трансформерних моделей.

На наступному етапі кодування результати обробки представлені у вигляді векторних репрезентацій. Після цього механізми злиття (fusion models) об’єднують отримані вектори у спільний простір ознак. Залежно від задачі, може застосовуватися різна стратегія злиття: середнє зважування, увага (attention) або складніші мета-моделі.

Підхід суміші експертів, за принципом оркестатора координує весь процес: маршрутизує потоки даних між модулями, обирає відповідний механізм злиття залежно від ситуації, контролює актуальність даних через автоматичні механізми оновлення бази знань. Після об'єднання даних фінальна метамодель виконує класифікацію, прогнозування або виявлення аномалій.

Результати аналізу передаються до інтерфейсного шару, де вони можуть бути візуалізовані, збережені у базі або надіслані до зовнішніх інформаційних систем. Крім того, за необхідності система може формувати автоматичні рекомендації для користувача.

Таким чином, запропонований конвеєр забезпечує безперервний цикл збору, обробки, інтеграції та аналізу мультимодальних даних із можливістю гнучкого налаштування під конкретні вимоги застосування.

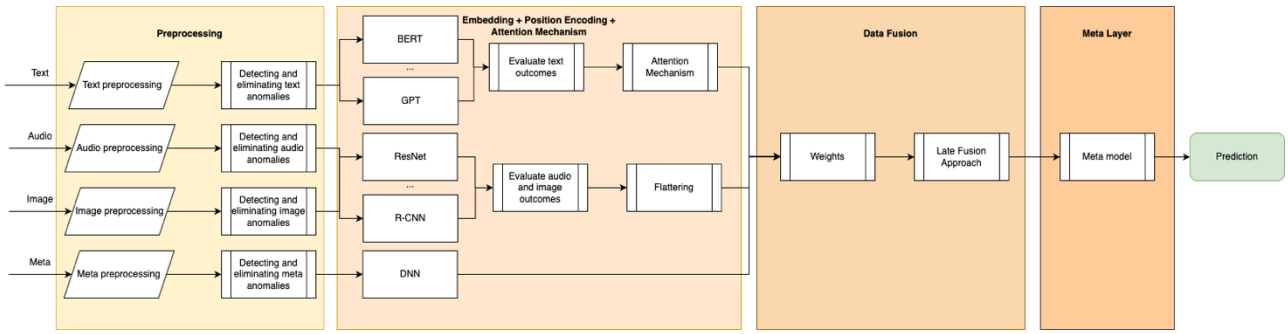


Рисунок 4.1 – Узагальнений конвеєр роботи запропонованої системи

Перший модуль, відомий як програма-агент, використовується для попередньої класифікації певних типів мономодальних даних. Він виконує такі завдання, як розділення даних і попередній аналіз. Тоді агент надсилає оброблені дані до модуля виявлення аномалій. У разі порушення зв'язку за допомогою наступного модуля можна тимчасово зберігати дані локально, що надає додаткові дані захист і запобігає втратам.

Другий модуль, оцінка аномалії даних, отримує та обробляє дані, зібрані агентом програми. Він структурує дані, запускає алгоритми виявлення аномалій і зберігає результати в базі даних для подальшого аналізу. Програма виконує додаткові оцінки будь-яких виявлених аномалій визначати їх критичність і повідомляти системних адміністраторів про несправності. Крім того, варто згадуючи про додаткові підсистеми для аналізу збоїв і зниження продуктивності, наприклад втрати точності певних наборів даних, важливі для такого модуля, що може призвести до каскадування продуктивність.

Для цілей цього дослідження ми оцінюємо три підходи до змішування:

- ранній змішування, коли дані з різних джерел поєднуються на початкових етапах обробка;
- пізнє змішування, коли дані обробляються окремо перед інтеграцією для прийняття рішень;
- гібридний змішування, що поєднує елементи як раннього, так і пізнього змішування.

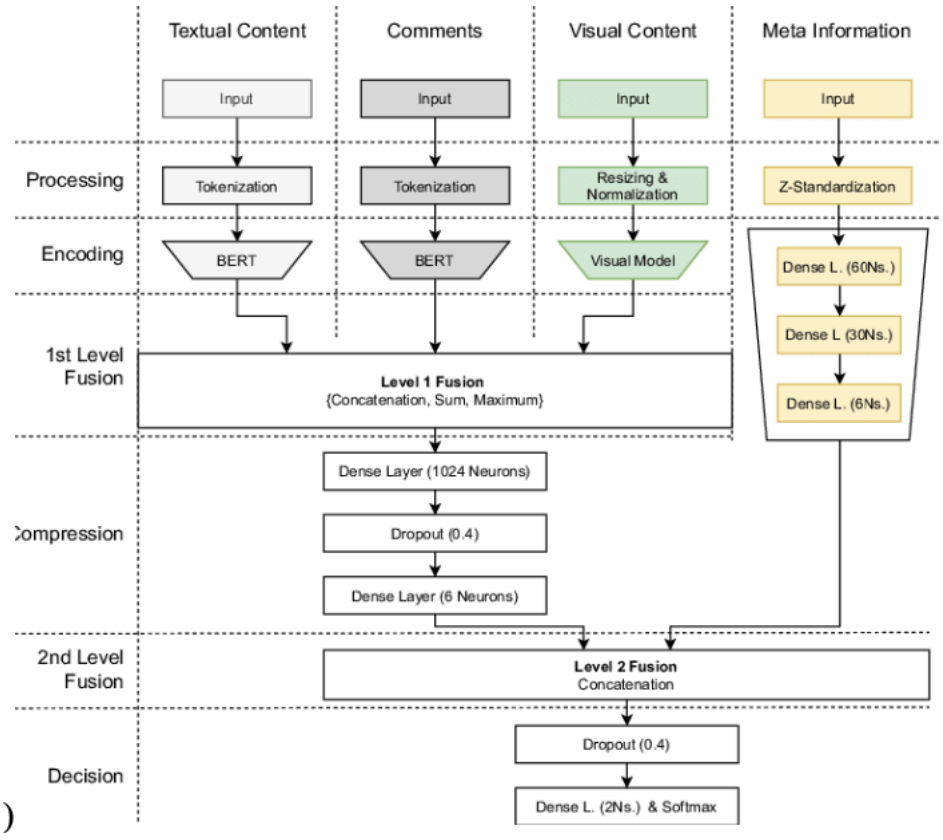


Рисунок 4.2 – Візуалізації підходу гібридного змішування (наявний ранній 1 рівень, та пізній 2 рівень змішувань)

Ці методи змішування оцінюються для підвищення точності та ефективності мультимодальні алгоритми виявлення даних інтелектуальної системи моніторингу. Більш детально ви можете переглянути результати в таблиці 4.1.

4.2. Проектування інтерфейсних рішень

Для обміну даних базовий протокол буде використовувати базовий протокол прикладного рівня HTTPS [120]. Для отримання даних буде використовуватись REST API через метод GET з випадковим ключем для забезпечення конфіденційності передачі даних [121] Структура URL адреса буде наступною https://multimodalsystem.io/input – для прийому і віддачі даних з ансамблю моделей машинного навчання, а також: https://multimodalsystem.io

/result для отримання підсумкових результатів після опрацювання даних усіх модальностей.

Таблиця 4.2 – Схема підсистем проєкту

Клієнт	Інтерфейс, який передає дані на сервер
Запит	REST API запити
Відповідь	REST API відповіді
Сервер	Сервер для опрацювання всіх даних
Модуль обробки мультимодальних даних	Модуль для опрацювання мультимодальних даних
Ансамбль моделей машинного навчання	Об’єднані моделі на основі підходу стекінгу, яка використовується для отримання узагальнених результатів

Система розроблена на мові програмування Python [122] з використанням бази даних MySQL [123] для запису тимчасових даних та моделлю машинного навчання на основі алгоритму ансамблю моделей для опрацювання мультимодальних даних для злиття і узгалальнення результатів.

4.3. Апробація методу аналізу модальних даних на основі ансамблю моделей машинного навчання

У рамках дослідження було проведено кілька експериментів для перевірки продуктивності опрацювання мультимодальних системою в різних сценаріях. Розглядатимуться сценарії перетворення тексту в мовлення, мовлення в текст, обробки природної мови.

Метод опрацювання мультимодальних даних передбачає чотири етапи.

Етап 1. Очищення та попереднього опрацювання даних, вхідні дані сортуються та автоматично очищуються від повторів.

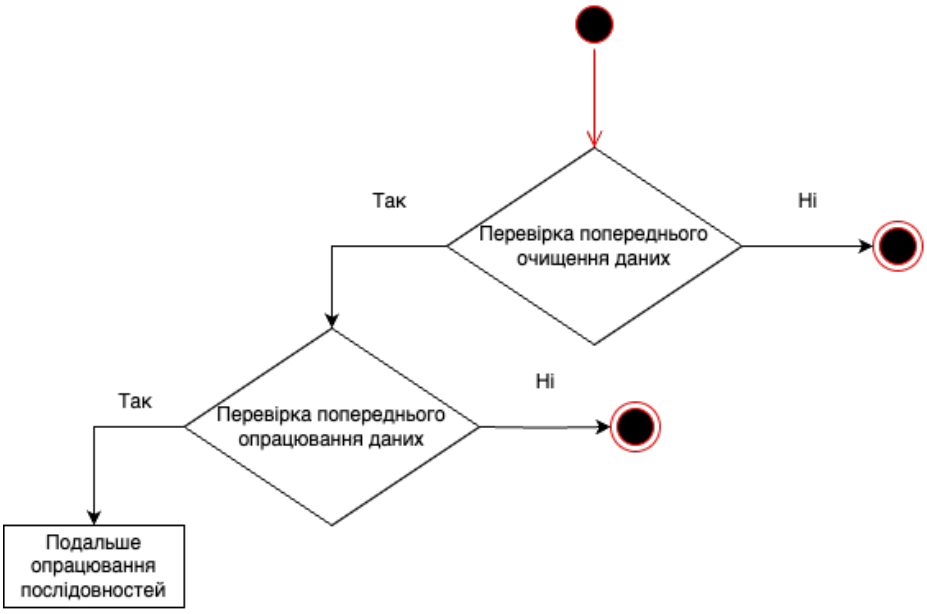


Рисунок 4.3 – Послідовності очищення та попереднього опрацювання даних

Етап 2. Групування за змістом та довжиною даних. Далі здійснюється синхронізації даних та узгодження змістів різномодальних наборів.

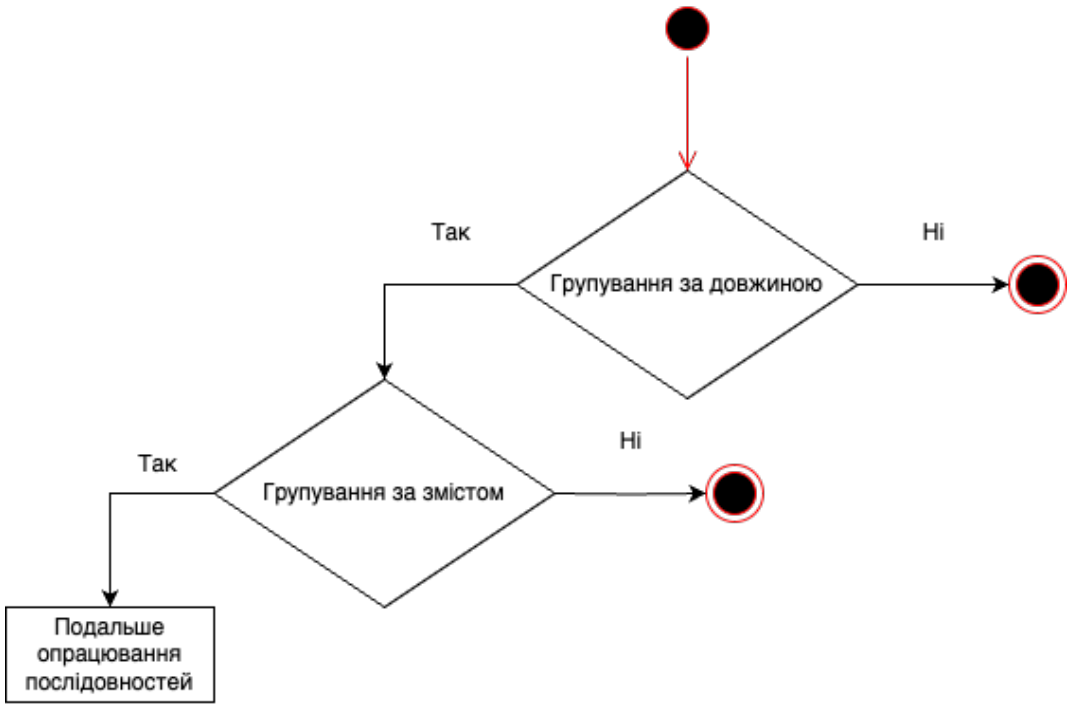


Рисунок 4.4 – Послідовність синхронізації даних та узгодження змістів різномодальних наборів

Етап 3. Виконується перевірка на існування ідентичних записів. Якщо такий запис вже існує, то він не записується повторно, а повертається відповідний запис, який міститься у базі даних системи.

Рисунок 4.5 – Послідовність перевірки на існування ідентичних записів

Етап 4. Здійснюється опрацювання за допомогою ансамблевого підходу, відбувається кодування даних в цифровий масив даних, з наступним перетворенням масиву даних у потрібний формат за домопогою функції декодувальника, зі збереженням результатів в базі даних.

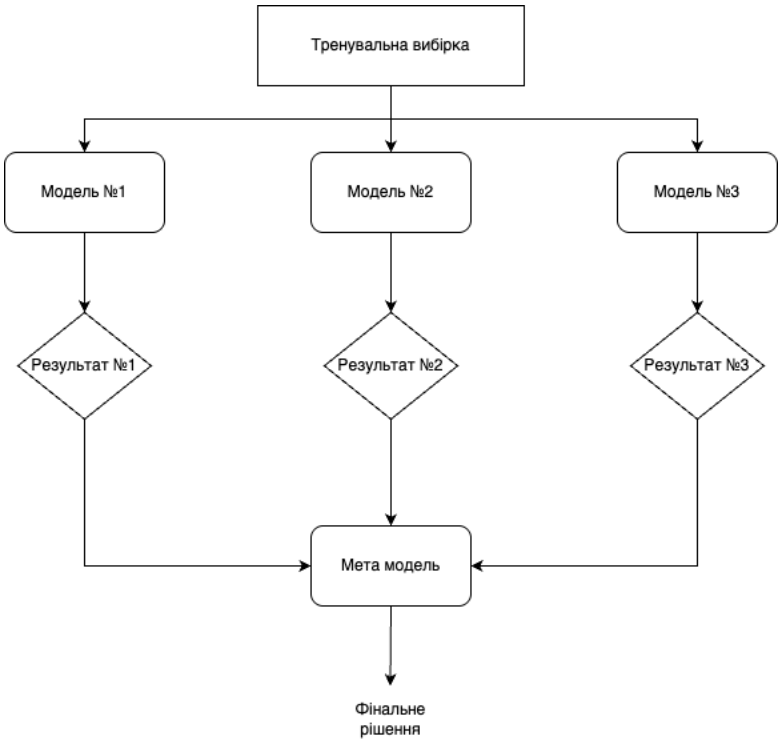


Рисунок 4.6 – Послідовність опрацювання за допомогою ансамблевого підходу

Отже, остаточна архітектура підходу буде представлена таким чином:

1. Інтерфейс додатку: це зручний інтерфейс даних для інтерфейсу обробки мультимодальних перевезень на основі Google API [124], який забезпечує платформу для обробки та аналізу мультимодальних даних.
2. Інтерфейс обробки даних: Модуль інтерфейсу мультимодальних даних також полегшує вирівнювання та синхронізацію даних між різними видами транспорту.
3. Підключення до бази даних: Модуль підключення до бази даних включає кілька підмодулів, які дозволяють користувачам взаємодіяти з базою даних, наприклад, конструктори запитів, мапери даних і конструктори схем. Ці підмодулі спрощують процес створення та управління таблицями і записами бази даних, а також дозволяють користувачам виконувати складні запити та завдання з аналізу даних. Модуль підключення до бази даних також містить надійні засоби безпеки для захисту даних, що зберігаються в базі даних. Загалом, модуль підключення до бази даних є найважливішим компонентом інтерфейсу обробки мультимодальних даних на основі Google API, що дозволяє користувачам зберігати, отримувати та управляти складними мультимодальними даними ефективно і безпечно.
4. Інтерфейс Google API: Інтерфейс Google API розроблений таким чином, щоб бути простим у використанні, з детальною документацією та зразками коду, доступними для розробників. Він також забезпечує високий рівень безпеки та надійності завдяки надійній автентифікації та механізмам обробки помилок. Загалом, інтерфейс Google API є потужним інструментом, який дозволяє розробникам інтегрувати низку сервісів Google у свої додатки, надаючи користувачам безперебійний та інтегрований досвід.
5. Функція Google Cloud [125]: цей інтерфейс є критично важливим компонентом інтерфейсу обробки мультимодальних перевезень,

оскільки він надає користувачам комплексну платформу для обробки та аналізу складних мультимодальних даних. Це дозволяє дослідникам і практикам отримати більш глибоке розуміння складних явищ, поєднуючи дані з різних джерел і модальностей.

6. Text-to-Speech API [126]: компонент Google Cloud API, який перетворює текст на природну мову. Цей API призначений для забезпечення високоякісного синтезу мовлення різними мовами та голосами.
7. Google AutoML [127]: компонент Google Cloud, який надає глибокі нейронні мережі (DNN) [128] для генерації мовлення, що забезпечить можливість публікації валідованої відповіді API сервісу Google Cloud Text-to-Speech. Модель нейронної мережі навчається на великому наборі даних записаної людської мови, щоб вивчити закономірності мовлення та генерувати точну і природну мову. Цей модуль допоможе нам отримувати якісні результати, а також видаляти неякісні результати та ігнорувати їх у майбутньому, або посилити цей напрямок, додавши коректні дані.

Вхідними даними для моделі є аудіо та текст. Для аудіо використовуються характеристики, отримані з аудіосигналу, зокрема мел-спектрограми, мел-частотні кепстральні коефіцієнти (MFCC), а також спектральний контраст. Ці параметри відображають акустичні властивості звукового сигналу. Текстові дані перетворюються на вектори ознак за допомогою методу TF-IDF. (продемонстрований на рисунку 4.12.).

У рамках даного дослідження, ключову роль відіграло апаратне забезпечення для проведення досліджень, а ключовим компонентом був графічний процесор, на якому і відбувалося тренування моделей машинного навчання. Окрім апаратної частини, в реалізації системи були також задіяні наступні бібліотеки: NumPy, Pandas, nltk (Natural Language Toolkit), nlpaug, noisereduce, pydub, librosa, matplotlib, seaborn, scikit-learn, tensorflow, keras.

4.4. Аналіз основних результатів роботи

Мультимодальна модель раннього злиття була навчена з використанням функції втрат binary_crossentropy, яка зазвичай застосовується для бінарних завдань, та метрики збалансованої точності для оцінки якості роботи моделі. Навчання тривало 25 епох. Для стабільності процесу оптимізації було використано оптимізатор Adam із фіксованою швидкістю навчання 0,001 [129].

На Рисунку 4.8 наведено графіки зміни точності та втрат моделі на навчальних і валідаційних даних упродовж епох:

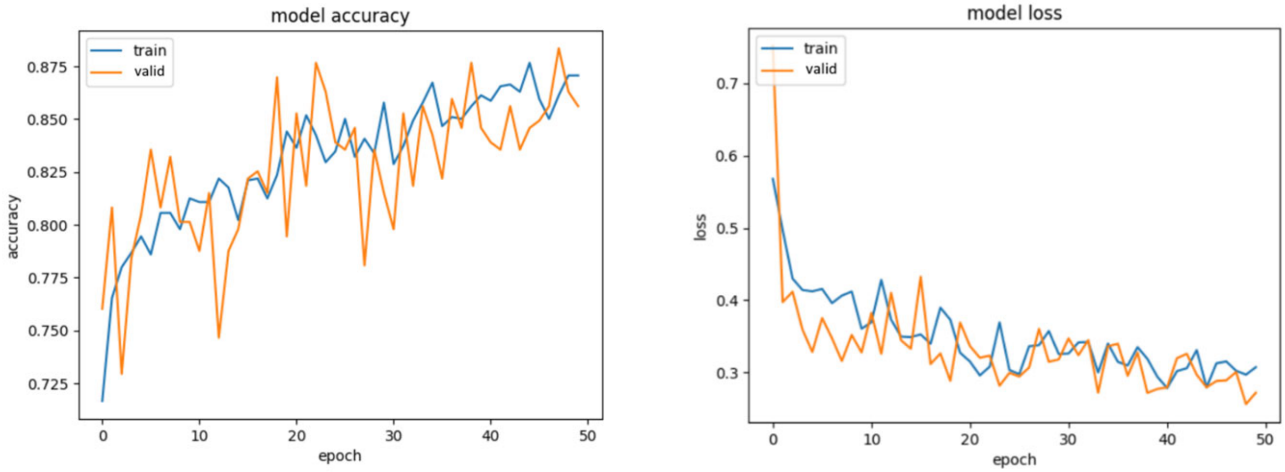
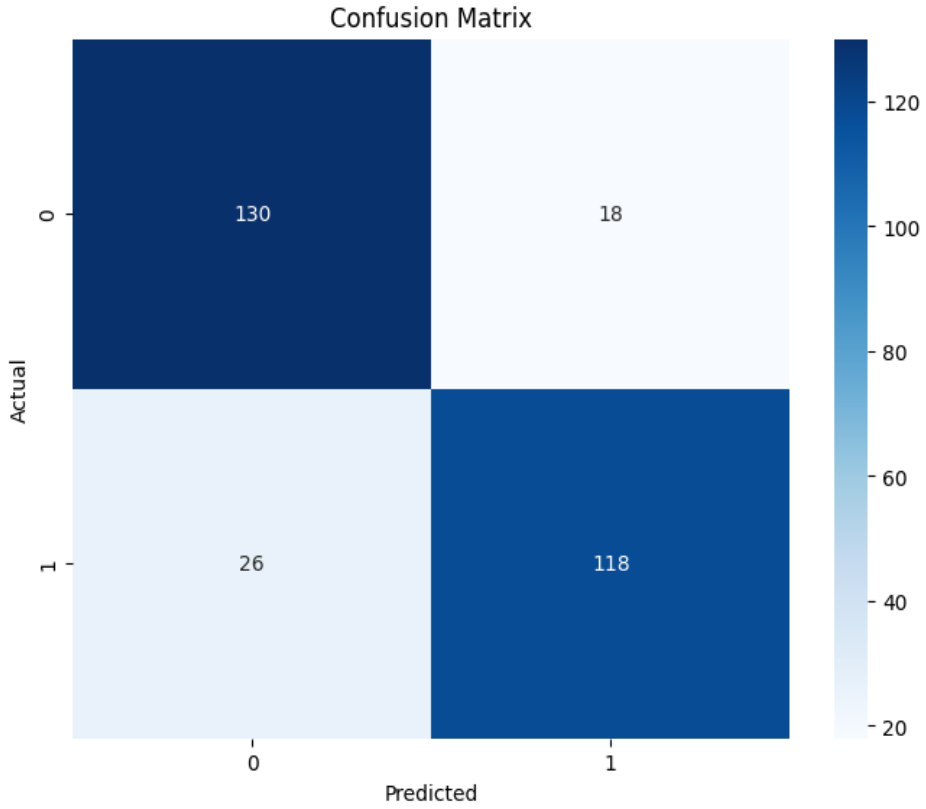


Рисунок 4.8. Графіки кривих функції точності і функції втрат для мультимодальної моделі раннього злиття.

1. Графік точності демонструє зростання загальної точності моделі під час навчання. На тренувальній вибірці точність постійно зростала й перевищила 95%. Валідаційна вибірка також показала високий рівень точності — понад 90%, що свідчить про ефективне вирішення завдання класифікації запропоновою моделлю.
2. Графік втрат ілюструє зниження начень функції втрат на навчальних і валідаційних вибірках. Спостерігається помітне зниження втрат на перших етапи навчання, що свідчить про правильний процес оптимізації. При співставленні результатів з валідаційної вибірки, помітно, що метрика функції втрат також знижується, що сввідчить про відсутність перенавчання.

Аналіз матриці невідповідностей для валідаційних даних (Рисунок 4.9) демонструє, що 273 записів з тестової вибірки прокласифіковані коректно, а допушено 10 екземлярів з помилкою типу False Positive, а 9 з помилкою типу False Negative.



Рисю 4.9. Матриця невідповідностей для валідаційних даних, для мультимодальної моделі раннього злиття.

У таблиці 4.3 представлені результати оцінки ефективності роботи моделі раннього злиття на валідаційній вибірці.

Аналізуючи звіт з метрик (табл. 4.4), можна відзначити наступне, для класу 0 значення точності позитивних передбачень становить 0,83, що свідчить про дещо більшу кількість хибнопозитивних прогнозів. Для класу 1 значення точності позитивних передбачень вище - 0,87, тобто модель дещо краще прогнозує депресію без надто великої кількості хибнопозитивних прогнозів. Для класу 0 значення повноти становить 0,88, що означає, що модель досить добре розпізнає більшість пацієнтів без депресії. Для класу 1 значення повноти

трохи нижче - 0,82, що вказує на те, що є деякі аномальні випадки, які модель не змогла виявити. Для класу 0 значення f1-показника становить 0,86, а для класу 1 - 0,84. Це означає, що модель дещо краще розпізнає відсутність депресії, ніж її наявність. Загальна точність моделі становить 0,85, що є досить хорошим результатом, враховуючи складність завдання класифікації депресії. Показник площі під кривою ROC-AUC 0,78 свідчить про помірну дискримінативну силу та сильну кореляцію на основі коефіцієнта кореляції Метьюса 0,70.

Таблиця 4.3. Звіт про метрики продуктивності моделі на тестової вибірки раннього злиття на основі даних валідації.

Метрики	**Клас**	
	0	**1**
Точність позитивних передбачень	0.83	0.87
Повнота.	0.88	0.82
F1-показникк	0.87	0.84
Загальна точність	0.85	
AUC	0.78	
MCC	0.70	

Загалом модель демонструє високу точність класифікації для обох класів, що свідчить про її ефективність у вирішенні цього завдання.

Модель пізнього злиття була навчена протягом 50 епох. Для оптимізації процесу навчання використовувався оптимізатор Adam зі швидкістю навчання 0,001, що дозволило здійснювати навчання швидше порівняно з моделлю раннього злиття. Детальніше з результатами навчання можна ознайомитися на рисунку 4.10, на якому візуалізовано вже відомі з попереднього прикладу нам метрики узагальненої точності та функції втрат.

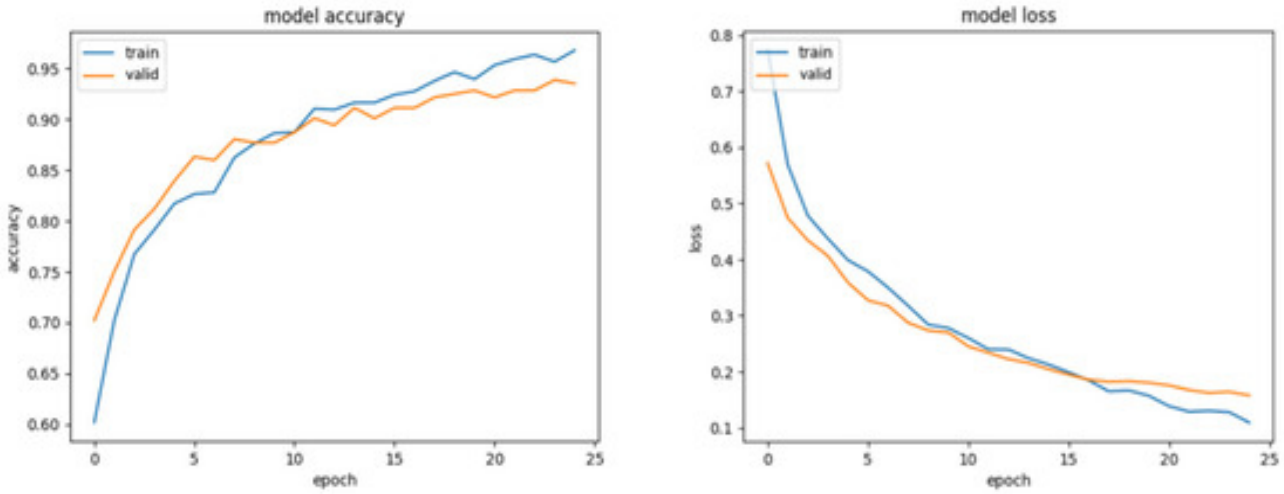


Рис. 4.10. Графіки прогресу навчання для мультимодальної моделі пізнього злиття.

Таким чином, обидва графіки показують, що модель поступово навчилася розрізняти депресію, і хоча є деякі коливання, загальна тенденція свідчить про хороший прогрес і задовільну точність.

Водночас, результати навчання моделі пізнього злиття дещо гірші порівняно з моделлю раннього злиття. Хоча точність валідації моделі пізнього злиття сягає близько 85%, вона поступається точності моделі раннього злиття, яка показала набагато стабільніші результати [129].

Матриця невідповідностей (Рисунок 4.11) показує розподіл результатів моделі на валідаційній вибірці. Модель правильно віднесла 130 зразків до класу 0 і 118 зразків до класу 1. Однак були й помилки: 18 зразків у класі 0 були помилково віднесені до класу 1, а 26 зразків у класі 1 були помилково віднесені до класу 0.

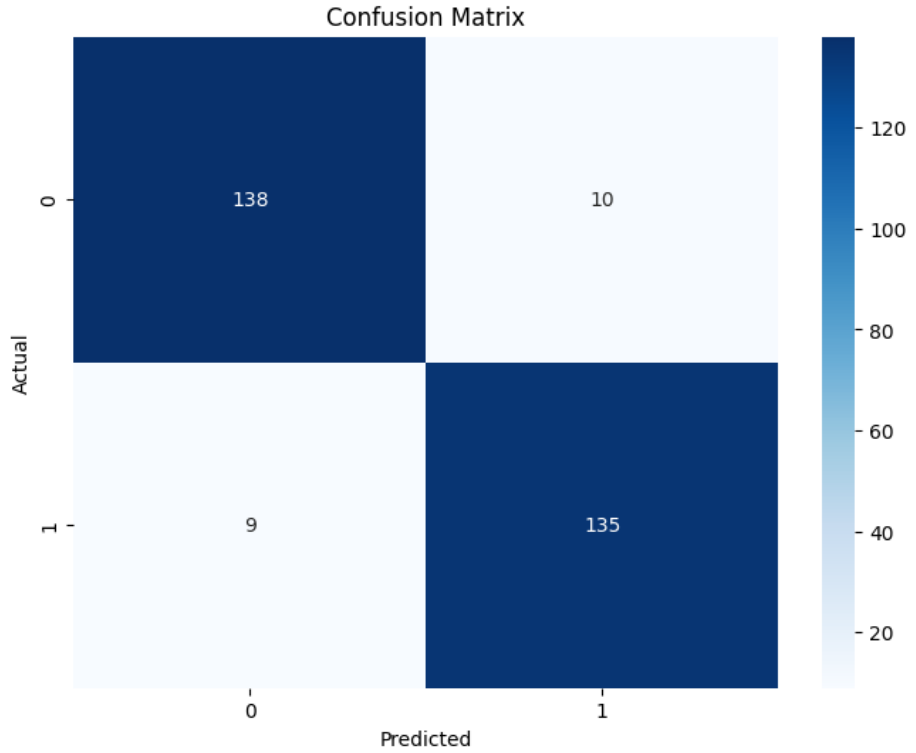


Рис. 4.9. Матриця невідповідностей для валідаційних даних, для мультимодальної моделі раннього злиття.

Ці результати свідчать про певні труднощі з правильною класифікацією депресії, але в цілому модель показала хорошу точність у виявленні обох класів [106].

Модель досягла загальної точності (accuracy) 0,94 для класу 0 та 0,93 для класу 1, що свідчить про низький рівень помилок при прогнозуванні.

Для обох класів модель показала високі значення повноти (recall) 0,93 і 0,94), що вказує на успішне розпізнавання більшості зразків у кожній категорії.

Високі значення F1-міри на рівні 0,94 для класу 1 і 0,93 для класу 0, свідчать про збалансованість запропонованих моделей, яка однаково добре справляється як із точністю, так і з повнотою отриманих передбачень. А також вдало узагальнюють інформацію з різних модальностей [129].

Оцінюючи метрику ROC-AUC кривої, площа фігури становить 0.9, це свідчить про помірно високу здатність моделі розрізняти класи, а коефіцієнт

кореляції Метьюса на рівні 0,87 підтверджує надійну узгодженість між прогнозами та фактичними значеннями даними валідаційної вибірки.

Таблиця 4.4. Звіт про показники ефективності моделі пізнього злиття на основі даних валідації.

Метрики	**Клас**	
	0	**1**
Точність позитивних передбачень	0.94	0.93
Повнота.	0.93	0.94
F1-показникк	0.94	0.93
Загальна точність	0.93	
AUC	0.90	
MCC	0.87	

Загалом модель мультимодальної класифікації працює стабільно, хоча спостерігається невелика різниця в точності між двома класами.

Порівнюючи результати роботи двох розроблених підходів — моделі раннього та пізнього злиття — можна зробити висновок, що модель пізнього злиття забезпечує вищу ефективність за ключовими метриками. Вона демонструє кращі результати за точністю позитивних передбачень, повнотою та F1-мірою, що вказує на її перевагу для задач мультимодальної класифікації [107].

Для інтерпретації результатів найкращої моделі — мережі пізнього злиття — було проведено тестування на окремій тестовій вибірці. Оскільки дані кожного учасника були розділені на сегменти, для отримання фінального рішення використовувався метод агрегування результатів на рівні учасників.

Процедура агрегування полягала в обчисленні середнього значення ймовірностей прогнозів сегментів для кожного учасника [129]. Остаточне

рішення формувалося на основі порогового значення: якщо середня ймовірність перевищувала 0,5, учасник зараховувався до одного з класів.

Для реалізації цього підходу використовувалися такі функції:

1. Функція агрегованих результатів: обчислює середнє значення ймовірностей для кожного учасника на основі прогнозів його сегментів.
2. Функція оцінки моделі: агрегує результати на рівні учасників і обчислює метрики ефективності, такі як точність, матрицю невідповідностей та звіт про класифікацію.

Запропонований метод надає можливість отримувати узагальнені результати і коректніше оцінювати загальну ефективність навченої моделі н на рівні сегментів, а в рамках всієї вибірки опрацьовуваних даних.

Матриця плутанини (Рисунок 4.12), отримана в результаті тестування, показала, що модель коректно визначила 37 записів коректно, а у чотирьох було допущено помилку типу False Positive, а у інших двох помилку типу False Negative.

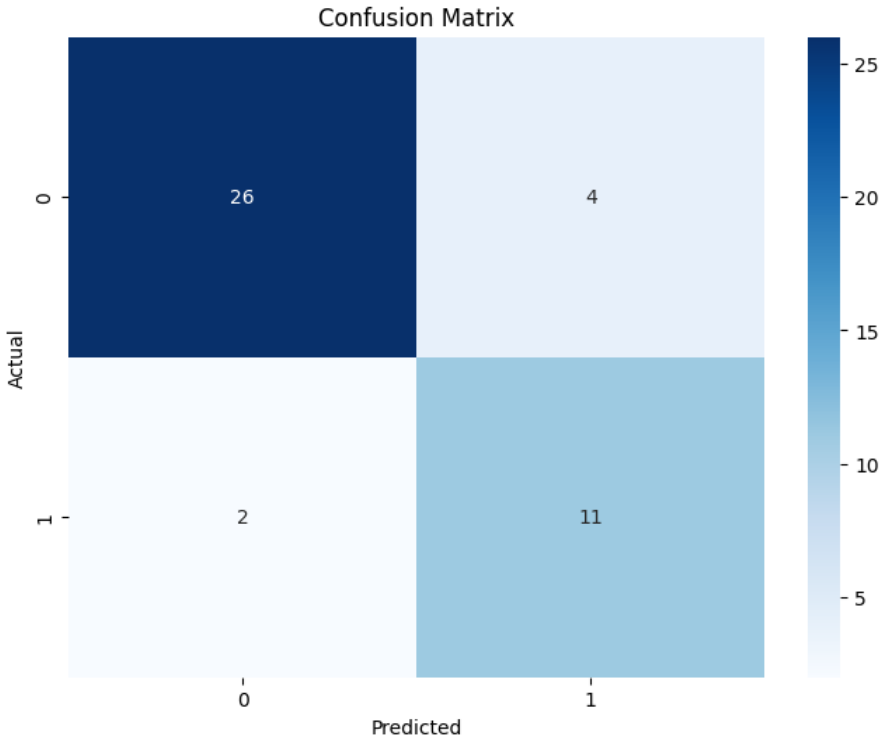


Рис. 4.12. Матриця невідповідностей для тестової вибірки.

У таблиці 4.5 представлені результати оцінки ефективності роботи моделі пізнього злиття на тестовій вибірці з невідомими даними.

Модель досягла точності на рівні 85%, що свідчить про задовільну здатність правильно класифікувати більшість зразків.

Оцінка площі під ROC-AUC кривої склала 0,73, що вказує на помірну здатність моделі розрізняти класи — вона працює краще, ніж випадкове вгадування, однак має потенціал для подальшого вдосконалення.

Коефіцієнт кореляції Метьюса на рівні 0,68 свідчить про значну відповідність між прогнозованими та фактичними класами, а також підтверджує стійкість моделі навіть у випадках можливого дисбалансу між класами.

Таблиця 4.5. Звіт про метрики продуктивності моделі на тестової вибірки.

Метрики	**Клас**	
	0	**1**
Точність позитивних передбачень	0.83	0.87
Повнота.	0.88	0.82
F1-показникк	0.86	0.84
Загальна точність	0.85	
AUC	0.73	
MCC	0.68	

Отже, модель демонструє стабільну та високу ефективність для обох класів, незважаючи на дещо нижчі показники для класу 1, що може бути пов'язано з меншою кількістю даних цього класу. Загальні результати свідчать про добру здатність моделі працювати з мультимодальними даними та підтверджують її потенціал для подальшого використання у подібних задачах.

4.5. Висновки

У даному розділі розроблена архітектура системи, описаний та розроблений протокол для обміну даних, спроектована для майбутньої розробки архітектура мультимодальної системи з використання ансамблів моделей машинного навчання для опрацювання даних.

Розроблено архітектуру мультимодальних даних в системі опрацювання природніх мов, яка використана в інформаційній системі перетворення аудіоінформації в текст, та навпаки. Систему впроваджено в приватному підприємстві "Мережа Медичних Центрів Родина". Розроблену систему можна використовувати, як конструктор для оптимізації даних, які є незмінними, наприклад логи для статистики тощо. Використання методів ансамблювання дозволили підвищити стійкість отриманих запропонованих рішень.

В рамках цього розділу було спроєктовано інтерфейсні рішення, орієнтовані на зручність взаємодії користувача з системою та швидке інтерпретування отриманих результатів у формі мультимодального інтерфейсу.

Проведено апробацію запропонованого методу аналізу модальних даних на основі ансамблю моделей, що дозволило продемонструвати переваги мультимодального підходу та ефективність інтеграції різних типів даних у межах єдиної системи. Аналіз основних результатів роботи підтвердив доцільність використання запропонованої архітектури та методів ансамблювання для підвищення точності виявлення аномалій і прогнозування стану об'єктів спостереження.

Отримані результати свідчать про високу ефективність запропонованої інформаційної технології в умовах обробки складних модальних даних та її потенціал для практичного застосування в різних галузях, де важливо забезпечити своєчасний аналіз, діагностику й прогнозування на основі різнорідних інформаційних джерел.

Результати розділу опубліковано в наукових працях автора [119, 125, 129].

ОСНОВНІ РЕЗУЛЬТАТИ ТА ВИСНОВКИ

У дисертаційній роботі, на основі виконаних теоретичних досліджень та практичних експериментів, вирішено актуальну науково-прикладну задачу розроблення інформаційної технології опрацювання мультимодальних даних, з акцентом на пристосування та апрабацію в рамках україномовних корпусних різномодальних даних.

При цьому отримано такі основні результати.

1. На підставі проведеного аналізу відомих рішень встановлено, що наявний технології опрацювання великих мультимодальних даних, не можуть бути сповна використані у опрацювання мульитмоадльни даних через неможливість забезпечення прийнятного часу доступу до даних та швидкості опрацювання.
2. Розроблено модель мультимодальних україномовних наборів даних, яка складається з послідовності числових векторів, які містить у собі заголовки, метадані та дані. Це дає змогу зберігати дані різної структури у різних базах та сховищах даних, що дозволяє зменшити суперечність у даних, які надходять з різних джерел, систем та сенсорів.
3. Розроблено моделі системи autoencoder (кодування/декодування) наборів даних різних модальностей з застосуванням механізмів уваги, які охоплюють можливість до опрацювання текст, відео та аудіо даних.
4. Проведено аналіз методик ансамблювання унімодальних систем для досягнення достатнього рівня точної роботи системи у вигляді. Додатково розглянути різні моделі злиття векторів різнотипових даних, а саме early, late і hybrid fusion.
5. За результатами експериментів, було створено узагальнений конвеєр опрацювання різнотипових даних на основі моделі пізнього

змішування, а також запроектовано систему з уніфікованим інтерфейсом для опрацювання вхідних даних різних модальностей.

6. Вдосконалено метод обчислення вагових коефіцієнтів для різних наборів даних, в операції нейроподібних мереж за рахунок покращення методу сингулярного
7. Розроблено метод перевірки якості внесених даних, який полягає в пошуку аномалій та попереднього процесингу даних, яка будується на основі унікальних послідовностей опрацювання для даних конкретної модальності;
8. Розроблено алгоритм перевірки якості внесених даних, який базується на аналізі наявності дублікатів, суперечливих даних та даних з пропусками.
9. Розроблено архітектуру інформаційної системи для опрацювання україномовних мультимодальних даних, яка використана в інформаційній системі визначення емоційного забарвлення даних та психологічного стану особи. Наявні віповідні акти впровадження в приватному підприємстві "Мережа Медичних Центрів Родина". Додатково розроблену систему можна використовувати, як конструктор для синхронізації та узагальнанення різнотипових даних, логи для статистики тощо.

Результати дисертаційної роботи також були використані та апробовані в рамках двох національних науково-дослідних проєктах, а саме «Технологія опрацювання мультимодальних українськомовних наборів даних для визначення рівня стресу» (№ державного реєстру 0123U100231). А також впроваджені в навчальному процесі Національного університету «Львівська політехніка» при викладанні дисципліни «Машинне навчання» та «Обробка й аналіз цифрових сигналів» для студентів освітньо-кваліфікаційного рівня «бакалавр», що навчаються за напрямом 122 "Комп’ютерні науки", на кафедрі систем штучного інтелекту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Karpathy and L. Fei-Fei, “Deep visual-semantic alignmentsfor generating image descriptions,” in Proceedings of the IEEEComputer Society Conference on Computer Vision and PatternRecognition (CVPR), 2015, pp. 3128–3137

2. Daxin Tan, Liqun Deng, Yu Ting Yeung, Xin Jiang, Xiao Chen,and Tan Lee, “Editspeech: A text based speech editing systemusing partial inference and bidirectional fusion,” arXiv preprintarXiv:2107.01554, 2021.

3. M. Oncescu, A. S. Koepke, J. F. Henriques, Z. Akata, andS. Albanie, “Audio Retrieval with Natural Language Queries,”in Proceedings of Conference of the International Speech Com-munication Association, 2021, pp. 2411–2415.

4. Ian Goodfellow, Yoshua Bengio, Aaron Courville, and YoshuaBengio, Deep learning, vol. 1, MIT press Cambridge, 2016

5. Бехта І. А., Матвієнків О. С. Структурно-семантичні типи фразеологізмів в англійськомовному художньому прозовому тексті. Вчені записки ТНУ імені В. І. Вернадського. Серія: Філологія. Соціальні комунікації, УДК 821.111-112. 81’42, DOI https://doi.org/10.32838/2663-6069/2020.2-2/05

6. Mienye, Ibomoiye Domor, Theo G. Swart, and George Obaido. 2024. "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications" Information 15, no. 9: 517. https://doi.org/10.3390/info15090517

7. Balit, E., & Chadli, A. (2020). GMFNet: Gated multimodal fusion network for visible-thermal semantic segmentation. In Proceedings of the 16th European Conference on Computer Vision (pp. 1–4).

8. Cheng S-T, Lyu Y-J, Teng C. Image-Based Nutritional Advisory System: Employing Multimodal Deep Learning for Food Classification and Nutritional Analysis. Applied Sciences. 2025; 15(9):4911. https://doi.org/10.3390/app15094911

9. Havryliuk, M., Dumyn, I., Vovk, O. (2023). Extraction of Structural Elements of the Text Using Pragmatic Features for the Nomenclature of Cases Verification. In: Hu, Z., Wang, Y., He, M. (eds) Advances in Intelligent Systems, Computer Science and Digital Economics IV. CSDEIS 2022. Lecture Notes on Data

Engineering and Communications Technologies, vol 158. Springer, Cham. https://doi.org/10.1007/978-3-031-24475-9_57

10. Vitaly Yakovyna, Natalya Shakhovska, "Software failure time series prediction with RBF, GRNN, and LSTM neural networks", Procedia Computer Science 207(4):837-847, DOI:10.1016/j.procs.2022.09.139.

11. Nataliya Shakhovska, et. al.: "The Developing of the System for Autimatic Audio to Text Conversion", IT&AS'2021: Symposium on Information Technologies and Applied Sciences, March 5–6, 2021, Bratislava, Slovak Republic.

12. Zoryana Rybchak, et. al. "Analysis of methods and means of text mining". ECONTECHMOD, 6(2), 2017, pp. 73-78.

13. P. Zdebskyi, V. Lytvyn,Y. Burov, and et. Intelligent system for semantically similar sentences identification and generation based on machine learning methods, CEUR Workshop Proceedings, 2020, pp. 317–346.

14. Naihan Li, Shujie Liu, Yanqing Liu, Sheng Zhao, and MingLiu, "Neural speech synthesis with transformer network," inProceedings of the AAAI Conference on Artificial Intelligence,2019, vol. 33, pp. 6706–6713.

15. Cao, F., Shi, T., Han, K., Wang, P., & An, W. (2023). RDFM: Robust deep feature matching for multi-modal remote sensing images. IEEE Geoscience and Remote Sensing Letters.

16. N. Shakhovska, N. Boyko, P. Pukach. The Information Model of Cloud Data Warehouses International Conference on Computer Science and Information Technologies, CSIT 2018, September 11-14, Lviv, Ukraine, 2019, pp. 182-191.

17. Balit, E., & Chadli, A. (2020). GMFNet: Gated multimodal fusion network for visible-thermal semantic segmentation. In Proceedings of the 16th European Conference on Computer Vision (pp. 1–4).

18. S. Chowdhury and J. Sil, "FACERECOGNITION from NON-FRONTALIMAGES Using DEEP NEURALNETWORK," in 2017 Ninth InternationalConference on Advances in PatternRecognition (ICAPR), 2017, pp. 1-6.

19. I. Zheliznyak, Z. Rybchak, I. Zavuschak, Analysis of clustering algorithms, 2017. Advances in Intelligent Systems and Computing, 2017, pp. 305–314.

20. Abdulmajeed N. Q. A review on voice pathology: taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions / N. Q. Abdulmajeed, B. Al-Khateeb, M. A. Mohammed // Journal of Intelligent Systems. — 2022. — Vol. 31, No. 1. — P. 855–875.

21. Panek D. Acoustic analysis assessment in speech pathology detection / D. Panek, A. Skalski, J. Gajda, R. Tadeusiewicz // International Journal of Applied Mathematics and Computer Science. — 2015. — Vol. 25, No. 3. — P. 631–643.

22. Hossain M. S. Healthcare big data voice pathology assessment framework / M. S. Hossain, G. Muhammad // IEEE Access. — 2016. — Vol. 4. — P. 7806–7815.

23. Verde L. Voice disorder identification by using machine learning techniques / L. Verde, G. De Pietro, G. Sannino // IEEE Access. — 2018. — Vol. 6. — P. 16246–16255.

24. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O., 2019. nuscenes: A multimodal dataset for autonomous driving. arXiv preprint arXiv:1903.11027 .

25. Alhussein M. Voice pathology detection using deep learning on mobile healthcare framework / M. Alhussein, G. Muhammad // IEEE Access. — 2018. — Vol. 6. — P. 41034–41041.

26. Рузакова О.В. Використання апаратів штучного інтелекту для формалізації фінансових об'єктів при побудові СППР / О. В. Рузакова, Н. П. Юрчук // Вісник Хмельницького національного університету: Технічні науки. – 2021. № 1. – С. 45-511.

27. Vázquez-Romero, A.; Gallardo-Antolín, A. Automatic detection of depression in speech using ensemble convolutional neural networks. Entropy 2020, 22, 688. https://doi.org/10.3390/e22060688.

28. Yin, F.; Du, J.; Xu, X.; Zhao, L. Depression detection in speech using transformer and parallel convolutional neural networks. Electronics 2023, 12, 328. https://doi.org/10.3390/electronics12020328.

29. Miao, X.; Li, Y.; Wen, M.; Liu, Y.; Julian, I.N.; Guo, H. Fusing features of speech for depression classification based on higher-order spectral analysis. Speech Commun. 2022, 143, 46–56. https://doi.org/10.1016/j.specom.2022.07.006.

30. Zhao, Y.; Liang, Z.; Du, J.; Zhang, L.; Liu, C.; Zhao, L. Multi-head attention-based long short-term memory for depression detection from speech. Front. Neurorobotics 2021, 15, 684037. https://doi.org/10.3389/fnbot.2021.684037.

31. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

32. Басистюк О. А., Мельникова Н. І. Мультимодальне розпізнавання мовлення на основі звукових і текстових даних // Вісник Хмельницького національного університету. Серія: Технічні науки. – 2022. – № 5 (313). – С. 22–25.

33. Олег Басистюк, Наталія Мельникова, Ірина Думин, Андрій Думин. Ансамблевий підхід у мультимодальній обробці даних на основі Google API // Вісник Хмельницького національного університету. Серія: Технічні науки. – 2024. – № 4 (339). – С. 235–238

34. Basystiuk, O., Melnykova, N. (2023, October). Multimodal Learning Analytics: An Overview of the Data Collection Methodology. In 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT) (pp. 1-4). IEEE.

35. Jiang, Q., Chi, Z., Ma, X., Mao, Q., Yang, Y., & Tang, J. (2025). Rebalanced Multimodal Learning with Data-aware Unimodal Sampling. arXiv preprint arXiv:2503.03792.

36. Index.dev. (2024, April 16). Comparing unimodal vs. multimodal models in generative AI. Index.dev Blog. https://www.index.dev/blog/comparing-unimodal-vs-multimodal-models

37. Alhussein M. Automatic voice pathology monitoring using parallel deep models for smart healthcare / M. Alhussein, G. Muhammad // IEEE Access. — 2019. — Vol. 7. — P. 46474–46479.

38. Al-Dhief F. T. A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms / F. T. Al-Dhief, N. M. A. Latiff, N. N. N. Abd. Malik, [et al.] // IEEE Access. — 2020. — Vol. 8. — P. 64514–64533.

39. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

40. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1810.04805

41. Al-Dhief F. T. Voice pathology detection and classification by adopting online sequential extreme learning machine / F. T. Al-Dhief, M. M. Baki, N. M. A. Latiff, [et al.] // IEEE Access. — 2021. — Vol. 9. — P. 77293–77306.

42. Alzubaidi, Laith, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions." Journal of big Data 8 (2021): 1-74.

43. Sidhu, Manjit Singh, Nur Atiqah Abdul Latib, and Kirandeep Kaur Sidhu. "MFCC in audio signal processing for voice disorder: a review." Multimedia Tools and Applications (2024): 1-21.

44. Rao, Chengping, and Yang Liu. "Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization." Computational Materials Science 184 (2020): 109850.

45. Islam R. A novel convolutional neural network based dysphonic voice detection algorithm using chromagram / R. Islam, M. Tarique // International Journal of Electrical and Computer Engineering (IJECE). — 2022. — Vol. 12, No. 5. — P. 5511.

46. Tuncer T. Novel multi center and threshold ternary pattern based method for disease detection method using voice / T. Tuncer, S. Dogan, F. Ozyurt, [et al.] // IEEE Access. — 2020. — Vol. 8. — P. 84532–84540.

47. Park, J.; Moon, N. Design and implementation of attention depression detection model based on multimodal analysis. Sustainability 2022, 14, 3569. https://doi.org/10.3390/su14063569.

48. Mao, K.; Zhang, W.; Wang, D.B.; Li, A.; Jiao, R.; Zhu, Y.; Wu, B.; Zheng, T.; Qian, L.; Lyu, W.; et al. Prediction of depression severity based on the prosodic and semantic features with bidirectional LSTM and time distributed CNN. IEEE Trans. Affect. Comput. 2022, 14, 2251–2265. https://doi.org/10.1109/taffc.2022.3154332.

49. Shbair, Wazen M., Thibault Cholez, Jerome Francois, and Isabelle Chrisment. "A multi-level framework to identify HTTPS services." In NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium, pp. 240-248. IEEE, 2016.

50. Iyortsuun, N.K.; Kim, S.-H.; Yang, H.-J.; Kim, S.-W.; Jhon, M. Additive cross-modal attention network (ACMA) for depression detection based on audio and textual features. IEEE Access 2024, 12, 20479–20489. https://doi.org/10.1109/access.2024.3362233.

51. Roberts, L. Understanding the Mel Spectrogram. 2020. Available online: https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53 (accessed on 22 December 2024).

52. Yang, J.; Luo, F.-L.; Nehorai, A. Spectral contrast enhancement: Algorithms and comparisons. Speech Commun. 2003, 39, 33–46. https://doi.org/10.1016/s0167-6393(02)00057-2.

53. Basystiuk, O.; Melnykova, N.; Rybchak, Z. Multimodal Learning Analytics: An Overview of the Data Collection Methodology 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT), https://doi.org/10.1109/CSIT61576.2023.10324177.

54. Jaafar, N.; Lachiri, Z. Multimodal fusion methods with deep neural networks and meta-information for aggression detection in surveillance. Expert Syst. Appl. 2022, 211, 118523. https://doi.org/10.1016/j.eswa.2022.118523.1

55. Abdulmajeed N. Q. A review on voice pathology: taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions

56. Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A., & Ganaie, I. A. (2023). Stock prediction by integrating sentiment scores of financial news and MLP-regressor: A machine learning approach. Procedia Computer Science, 218, 1067-1078.

57. Meng, Tong, Xuyang Jing, Zheng Yan, and Witold Pedrycz. "A survey on machine learning for data fusion." Information Fusion 57 (2020): 115-129.

58. Lahat, Dana, Tülay Adali, and Christian Jutten. "Multimodal data fusion: an overview of methods, challenges, and prospects." Proceedings of the IEEE 103, no. 9 (2015): 1449-1477.

59. N. Q. Abdulmajeed, B. Al-Khateeb, M. A. Mohammed // Journal of Intelligent Systems. — 2022. — Vol. 31, No. 1. — P. 855–875. https://www.kaggle.com/datasets/iamhungundji/dysarthria-detection/data

60. Lu W. A cnn-lstm-based model to forecast stock prices / W. Lu, J. Li, Y. Li, [et al.] // Complexity. — 2020. — Vol. 2020. — P. 1–10.

57. Tutorial on lstms: a computational perspective | by manu rastogi | towards data science / .

58. Intuitive understanding of mfccs. the mel frequency cepstral coefficients… | by emmanuel deruty | medium /

59. Panek D. Acoustic analysis assessment in speech pathology detection / D. Panek, A. Skalski, J. Gajda, R. Tadeusiewicz // International Journal of Applied Mathematics and Computer Science. — 2015. — Vol. 25, No. 3. — P. 631–643.

60. Al-Selwi, Safwan Mahmood, Mohd Fadzil Hassan, Said Jadid Abdulkadir, Amgad Muneer, Ebrahim Hamid Sumiea, Alawi Alqushaibi, and Mohammed Gamal Ragab. "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review." Journal of King Saud University-Computer and Information Sciences (2024): 102068.

61. Hossain M. S. Healthcare big data voice pathology assessment framework / M. S. Hossain, G. Muhammad // IEEE Access. — 2016. — Vol. 4. — P. 7806–7815.

62. Verde L. Voice disorder identification by using machine learning techniques / L. Verde, G. De Pietro, G. Sannino // IEEE Access. — 2018. — Vol. 6. — P. 16246–16255.

63. Yu, Yong, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. "A review of recurrent neural networks: LSTM cells and network architectures." Neural computation 31, no. 7 (2019): 1235-1270.

64. Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586.

65. Sherstinsky, Alex. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." Physica D: Nonlinear Phenomena 404 (2020): 132306. Alhussein M. Automatic voice pathology monitoring using parallel deep models for smart healthcare / M. Alhussein, G. Muhammad // IEEE Access. — 2019. — Vol. 7. — P. 46474–46479.

66. Al-Dhief F. T. A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms / F. T. Al-Dhief, N. M. A. Latiff, N. N. N. Abd. Malik, [et al.] // IEEE Access. — 2020. — Vol. 8. — P. 64514–64533.

67. Pavlyshenko, Bohdan. "Using stacking approaches for machine learning models." In 2018 IEEE second international conference on data stream mining & processing (DSMP), pp. 255-258. IEEE, 2018.

68. Tuncer T. Novel multi center and threshold ternary pattern based method for disease detection method using voice / T. Tuncer, S. Dogan, F. Ozyurt, [et al.] // IEEE Access. — 2020. — Vol. 8. — P. 84532–84540.

69. Al-Dhief F. T. Voice pathology detection and classification by adopting online sequential extreme learning machine / F. T. Al-Dhief, M. M. Baki, N. M. A. Latiff, [et al.] // IEEE Access. — 2021. — Vol. 9. — P. 77293–77306.

70. Islam R. A novel convolutional neural network based dysphonic voice detection algorithm using chromagram / R. Islam, M. Tarique // International Journal of Electrical and Computer Engineering (IJECE). — 2022. — Vol. 12, No. 5. — P. 5511.

71. Sill, Joseph, Gábor Takács, Lester Mackey, and David Lin. "Feature-weighted linear stacking." arXiv preprint arXiv:0911.0460 (2009).

72. Ko, Jiyoung, and Yung-Cheol Byun. "Analyzing factors affecting micro-mobility and predicting micro-mobility demand using ensemble voting regressor." Electronics 12, no. 21 (2023): 4410.

73. Dey, Rahul, and Fathi M. Salem. "Gate-variants of gated recurrent unit (GRU) neural networks." In 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS), pp. 1597-1600. IEEE, 2017.

74. Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. International Journal of Cognitive Computing in Engineering, 2, 40-46.

75. Daubechies, Ingrid, Ronald DeVore, Simon Foucart, Boris Hanin, and Guergana Petrova. "Nonlinear approximation and (deep) ReLU networks." Constructive Approximation 55, no. 1 (2022): 127-172.

76. Vijayakumar, A. K., Cogswell, M., Selvaraju, R. R., Sun, Q., Lee, S., Crandall, D., & Batra, D. (2016). Diverse beam search: Decoding diverse solutions from neural sequence models. arXiv preprint arXiv:1610.02424.

77. Liu, Yijin, Fandong Meng, Yufeng Chen, Jinan Xu, and Jie Zhou. "Scheduled sampling based on decoding steps for neural machine translation." arXiv preprint arXiv:2108.12963 (2021).

78. Zhou, Yuxuan, Margret Keuper, and Mario Fritz. "Balancing diversity and risk in llm sampling: How to select your method and parameter for open-ended text generation." arXiv preprint arXiv:2408.13586 (2024).

79. Amin, Mostafa M., and Björn W. Schuller. "On prompt sensitivity of ChatGPT in affective computing." arXiv preprint arXiv:2403.14006 (2024).

80. Kashyap, Kishore, Shikhar Kumar Sarma, and Mazida Akhtara Ahmed. "Improving translation between English, Assamese bilingual pair with monolingual data, length penalty and model averaging." International Journal of Information Technology 16, no. 3 (2024): 1539-1549.

81. Enkvist, Nils Erik. "Text and discourse linguistics, rhetoric and stylistics." In Discourse and literature, pp. 11-38. John Benjamins Publishing Company, 2011.

82. Keaton, Shaughan A., and Christopher C. Gearhart. "Identity formation, identity strength, and self-categorization as predictors of affective and psychological outcomes: A model reflecting sport team fans' responses to highlights and lowlights of a college football season." Communication & Sport 2, no. 4 (2014): 363-385.

83. Rasenberg, Marlou, Wim Pouw, Asli Özyürek, and Mark Dingemanse. "The multimodal nature of communicative efficiency in social interaction." Scientific Reports 12, no. 1 (2022): 19111.

84. Basystiuk O., Melnykova N. Development of the multimodal handling interface based on Google API // Комп'ютерні системи проектування. Теорія і практика. – 2024. – Вип. 6, № 1. – С. 216–223

85. Basystiuk O., Rybchak Z. Recommendation systems in e-commerce applications // Вісник Національного університету “Львівська політехніка”. Серія: Інформаційні системи та мережі. – 2024. – Вип. 15. – С. 252–259.

86. Basystiuk O., Melnykova N. Multimodal medical data learning approaches for digital healthcare // CEUR Workshop Proceedings. – 2024. – Vol. 3609: Proceedings of the 6th International conference on informatics & data-driven medicine, Bratislava, Slovakia, November 17-19, 2023. (IDDM 2023) (pp. 332–337).

87. Basystiuk, O., Melnykova, N. (2023, December). Multimodal Approaches for Natural Language Processing in Medical Data. In 5th International Conference on Informatics & Data-Driven Medicine (IDDM 2022) (pp. 246-252). IEEE.

84. Basystiuk O., Melnykova N. Development of the multimodal handling interface based on Google API // Комп'ютерні системи проектування. Теорія і практика. – 2024. – Вип. 6, № 1. – С. 216–223

85. Basystiuk O., Rybchak Z. Recommendation systems in e-commerce applications // Вісник Національного університету “Львівська політехніка”. Серія: Інформаційні системи та мережі. – 2024. – Вип. 15. – С. 252–259.

86. Basystiuk O., Melnykova N. Multimodal medical data learning approaches for digital healthcare // CEUR Workshop Proceedings. – 2024. – Vol. 3609: Proceedings of the 6th International conference on informatics & data-driven medicine, Bratislava, Slovakia, November 17-19, 2023. (IDDM 2023) (pp. 332–337).

87. Basystiuk, O., Melnykova, N. (2023, December). Multimodal Approaches for Natural Language Processing in Medical Data. In 5th International Conference on Informatics & Data-Driven Medicine (IDDM 2022) (pp. 246-252). IEEE.

88. Siddique, Nahian, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. "U-net and its variants for medical image segmentation: A review of theory and applications." IEEE access 9 (2021): 82031-82057.

89. Bergmann, Paul, Michael Fauser, David Sattlegger, and Carsten Steger. "MVTec AD--A comprehensive real-world dataset for unsupervised anomaly

detection." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9592-9600. 2019.

90. Rethage, Dario, Jordi Pons, and Xavier Serra. "A wavenet for speech denoising." In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5069-5073. IEEE, 2018.

91. Kang, Yue, Zhao Cai, Chee-Wee Tan, Qian Huang, and Hefu Liu. "Natural language processing (NLP) in management research: A literature review." Journal of Management Analytics 7, no. 2 (2020): 139-172.

92. Mei, Taiyuan, Yun Zi, Xiaohan Cheng, Zijun Gao, Qi Wang, and Haowei Yang. "Efficiency optimization of large-scale language models based on deep learning in natural language processing tasks." In 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE), pp. 1231-1237. IEEE, 2024.

93. Steinberg, Alan N., and Christopher L. Bowman. "Revisions to the JDL data fusion model." In Handbook of multisensor data fusion, pp. 65-88. CRC press, 2017.

94. Gheini, Mozhdeh, Xiang Ren, and Jonathan May. "Cross-attention is all you need: Adapting pretrained transformers for machine translation." arXiv preprint arXiv:2104.08771 (2021).

95. Kim, Hannah Halin, Shuzhi Yu, Shuai Yuan, and Carlo Tomasi. "Cross-attention transformer for video interpolation." In Proceedings of the Asian conference on computer vision, pp. 320-337. 2022.

96. Petit, Olivier, Nicolas Thome, Clement Rambour, Loic Themyr, Toby Collins, and Luc Soler. "U-net transformer: Self and cross attention for medical image segmentation." In Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12, pp. 267-276. Springer International Publishing, 2021.

97. Xu, H., Ma, J., Jiang, J., Guo, X., & Ling, H. (2020). U2Fusion: A unified unsupervised image fusion network. IEEE transactions on pattern analysis and machine intelligence, 44(1), 502-518.

98. Lahat, Dana, Tülay Adali, and Christian Jutten. "Multimodal data fusion: an overview of methods, challenges, and prospects." Proceedings of the IEEE 103, no. 9 (2015): 1449-1477.

99. Song, Jingqi, Yuanjie Zheng, Muhammad Zakir Ullah, Junxia Wang, Yanyun Jiang, Chenxi Xu, Zhenxing Zou, and Guocheng Ding. "Multiview multimodal network for breast cancer diagnosis in contrast-enhanced spectral mammography images." International Journal of Computer Assisted Radiology and Surgery 16, no. 6 (2021): 979-988.

100. Li, Jiayuan, Qingwu Hu, and Mingyao Ai. "RIFT: Multi-modal image matching based on radiation-variation insensitive feature transform." IEEE Transactions on Image Processing 29 (2019): 3296-3310.

101. Sen, Sangeeta, Devashish Katoriya, Animesh Dutta, and Biswanath Dutta. "RDFM: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data." Expert Systems with Applications 179 (2021): 115043.

102. Jiang, H., Karpur, A., Cao, B., Huang, Q., & Araujo, A. (2024). Omniglue: Generalizable feature matching with foundation model guidance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 19865-19875).

103. Sarlin, Paul-Edouard, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. "Superglue: Learning feature matching with graph neural networks." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4938-4947. 2020.

104. Helsinki-NLP/Ted_Iwlst2013 · Datasets at Hugging Face. 1 June 2022, huggingface.co/datasets/Helsinki-NLP/ted_iwlst2013.

105. TED2020/TED2020-en-hy-train.tsv.gz · Sentence-transformers/Parallel-sentences at E68a0af02fc4fc4aeb13d2016572e04071017f9c.

huggingface.co/datasets/sentence-transformers/parallel-sentences/blame/e68a0af02fc4fc4aeb13d2016572e04071017f9c/TED2020/TED2020-en-hy-train.tsv.gz.

106. Papers With Code - BABEL Dataset. paperswithcode.com/dataset/babel-1.

107. Mao, Ze, Yang Xu, and Erick Suarez. "Dataset Management Platform for Machine Learning." arXiv preprint arXiv:2303.08301 (2023).

108. Gadre, Samir Yitzhak, Gabriel Ilharco, Alex Fang, Jonathan Hayase, Georgios Smyrnis, Thao Nguyen, Ryan Marten et al. "Datacomp: In search of the next generation of multimodal datasets." Advances in Neural Information Processing Systems 36 (2023): 27092-27112.

109. Fellbaum, Christiane. "WordNet." In Theory and applications of ontology: computer applications, pp. 231-243. Dordrecht: Springer Netherlands, 2010.

110. Pulapakura, Raj. “Multimodal Models and Fusion - a Complete Guide | Medium.” Medium, 6 Mar. 2024, medium.com/@raj.pulapakura/multimodal-models-and-fusion-a-complete-guide-225ca91f6861.

111. Qaiser, Shahzad, and Ramsha Ali. "Text mining: use of TF-IDF to examine the relevance of words to documents." International journal of computer applications 181, no. 1 (2018): 25-29.

112. Qader, Wisam A., Musa M. Ameen, and Bilal I. Ahmed. "An overview of bag of words; importance, implementation, applications, and challenges." In 2019 international engineering conference (IEC), pp. 200-204. IEEE, 2019.

113. Church, Kenneth Ward. "Word2Vec." Natural Language Engineering 23, no. 1 (2017): 155-162.

114. Pratiwi, Heny, Agus Perdana Windarto, S. Susliansyah, Ririn Restu Aria, Susi Susilowati, Luci Kanti Rahayu, Yuni Fitriani, Agustiena Merdekawati, and Indra Riyana Rahadjeng. "Sigmoid activation function in selecting the best model of artificial neural networks." In Journal of Physics: Conference Series, vol. 1471, no. 1, p. 012010. IOP Publishing, 2020.

115. Suebsombut, Paweena, Aicha Sekhari, Pradorn Sureephong, Abdelhak Belhi, and Abdelaziz Bouras. "Field data forecasting using LSTM and Bi-LSTM approaches." Applied Sciences 11, no. 24 (2021): 11820.

116. Sleeman IV, William C., Rishabh Kapoor, and Preetam Ghosh. "Multimodal classification: Current landscape, taxonomy and future directions." ACM Computing Surveys 55, no. 7 (2022): 1-31.

117. Kobylyukh, L., Rybchak Z., Basystiuk, O. (2023, April). Analyzing the Accuracy of Speech-to-Text APIs in Transcribing the Ukrainian Language. In The 7th International Conference Computational Linguistics and Intelligent Systems» (CoLInS 2023) (pp. 217-227).

118. Басистюк, Наталія Мельникова, Ірина Думин. Розроблення мультимодального інтерфейсу на основі Google API // Вісник Хмельницького національного університету. Серія: Технічні науки. – 2024. – № 3 (335). – С. 15–18.

119. Basystiuk, O., Farmaha I., Rybchak, Z. (2023, February). Exploring Multimodal Data Approach in Natural Language Processing Based on Speech Recognition Algorithms. In 2023, the 17th International Conference on the Experience of Designing and Application of CAD Systems (CADSM) (pp. 1-3). IEEE.

120. Zanevych Y., Yovbak V., Basystiuk O., Fedushko S., Shahovska N., Argyroudis S. Evaluation of pothole detection performance using deep learning models under low-light conditions // Sustainability. – 2024. – Vol. 16, iss. 24. – P. 1–15.

121. Boniface, Coline, Imane Fouad, Nataliia Bielova, Cédric Lauradoux, and Cristiana Santos. "Security analysis of subject access request procedures: How to authenticate data subjects safely when they request for their data." In Privacy Technologies and Policy: 7th Annual Privacy Forum, APF 2019, Rome, Italy, June 13–14, 2019, Proceedings 7, pp. 182-209. Springer International Publishing, 2019.

122. Phillips, Dusty. Python 3 Object-Oriented Programming.: Build robust and maintainable software with object-oriented design patterns in Python 3.8. Packt Publishing Ltd, 2018.

123. Kohut N., Basystiuk O., Melnykova N., Shahovska N. Detecting Plant Diseases Using Machine Learning Models // Sustainability. – 2024. – Vol. 16, iss. 24. – P. 1–15.

124. Këpuska, Veton, and Gamal Bohouta. "Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx)." Int. J Eng. Res. Appl 7, no. 03 (2017): 20-24.

125. Bisong, Ekaba. "An overview of google cloud platform services." Building Machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners (2019): 7-10.

126. “Text-to-Speech AI: Lifelike Speech Synthesis | Google Cloud.” Google Cloud, cloud.google.com/text-to-speech.

127. “AutoML Solutions - Train Models Without ML Expertise.” Google Cloud, cloud.google.com/automl.

128. Mittal, S. (2020). A survey on modeling and improving reliability of DNN algorithms and accelerators. Journal of Systems Architecture, 104, 101689.

129. Nykoniuk M., Basystiuk O., Shahovska N, Melnykova N. Multimodal Data Fusion for Depression Detection Approach // Computation. – 2024. – Vol. 16, iss. 24. – P. 1–15.

ДОДАТКИ

"ЗАТВЕРДЖУЮ"
Проректор з наукової роботи
Національного університету
"Львівська політехніка"
Іван ДЕМИДОВ
"18" 03 2025 р.

АКТ
про використання наукових результатів
дисертаційної роботи Басистюка Олега Андрійовича
представленої на здобуття наукового ступеня доктора філософії

Комісія в складі: голови комісії – начальник науково-дослідної частини д.т.н., с.н.с. Небесного Р.В. та членів комісії – завідувача кафедри СШІ Мельникової Н.І., старшого викладача кафедри СШІ Думин І.Б., доцента кафедри СШІ Хавалка В.М., доцента кафедри СШІ Кривенчука Ю.П. цим актом підтверджують, що результати дисертаційної роботи Басистюка Олега Андрійовича, зокрема:

- метод синхронізації мультимодальних даних;
- механізм уваги для виявлення найважливіших сегментів інформації;
- метод пошуку і виявлення аномалій в мультимодальних даних

використано у науково-дослідних роботах фінансованих Міністерством освіти і науки України, що виконувалась на кафедрі систем штучного інтелекту і включено до звіту «Технологія опрацювання мультимодальних українськомовних наборів даних для визначення рівня стресу» (№ державної реєстрації 0123U100231).

Отримані автором результати використано:

- при налаштуванні параметрів моделі для злиття мультимодальних даних;
- при проектуванні мультимодального інтерфейсу розроблюваної системи;
- при розробці методів класифікації емоцій для мультимодальних даних.

Голова комісії:
Начальник науково-дослідної частини
д.т.н., с.н.с. Роман НЕБЕСНИЙ

Члени комісії:

Завідувач кафедри СШІ Наталія МЕЛЬНИКОВА

Старший викладача кафедри СШІ Ірина ДУМИН

Доцент кафедри СШІ Віктор ХАВАЛКО

Доцент кафедри СШІ Юрій КРИВЕНЧУК

"ЗАТВЕРДЖУЮ"
Проректор з наукової роботи
Національного університету
"Львівська політехніка"
_____ Іван ДЕМИДОВ
"___" _____ 2025 р.

АКТ
про використання наукових результатів
дисертаційної роботи Басистюка Олега Андрійовича
представленої на здобуття наукового ступеня доктора філософії

Комісія в складі: голови комісії – начальник науково-дослідної частини д.т.н., с.н.с. Небесного Р.В. та членів комісії – завідувача кафедри СШІ Мельникової Н.І., директор ІКНІ Шаховська Н.Б., цим актом підтверджують, що результати дисертаційної роботи Басистюка Олега Андрійовича, зокрема:

- метод пошуку найважливіших сегментів інформації, на основі механізмів уваги;
- моделі змішування даних різних модальностей;
- принципи ансамблювання моделей машинного навчання.

використано у науково-дослідних роботі, що виконувалась на кафедрі систем штучного інтелекту і включено до звіту за договором «Комплексний догляд для наступного покоління» (C2020/1-8, iCare4Next) виконаної за договором № М/100-2024 від 19.07.2024 р.

Отримані автором результати використано:

- при налаштуванні гіперпараметрів моделей контрольоване навчання;
- при проектуванні моделей неконтрольованого навчання для класифікації даних;
- при розробці гібридні ансамблі машинного навчання.

Голова комісії:
Начальник науково-дослідної частини
д.т.н., с.н.с. Роман НЕБЕСНИЙ

Члени комісії:

Директор ІКНІ Наталія ШАХОВСЬКА

Завідувач кафедри СШІ Наталія МЕЛЬНИКОВА

"ЗАТВЕРДЖУЮ"
Генеральний директор
Приватного підприємства
"Мережа Медичних Центрів Родина"
_____ **Андрій ХАНИК**
"11" квітня 2025 р.

ПП "МЕРЕЖА МЕДИЧНИХ ЦЕНТРІВ "РОДИНА" вул. Б. Хмельницького, 31 м. Львів 79019

АКТ

про впровадження результатів дисертаційної роботи
аспіранта кафедри «Систем штучного інтелекту»
Національного університету «Львівська політехніка»
Басистюка Олега Андрійовича

Цей акт підтверджує, що результати дисертаційної роботи **Басистюка О.А.** були використані для розроблення системи створення емоційного портрету та визначення психологічного стану пацієнта, в м. Львів за 2024-2025р.

Впровадження дисертаційних досліджень Басистюка Олега Андрійовича полягає в наступному:

- Розроблено метод аналізу даних різних модальностей, на основі ансамблю моделей машинного навчання, з метою підвищення ефективності опрацювання опитувальників пацієнта.
- Розроблено архітектуру інформаційної системи для визначення емоційного стану пацієнта, запропоновано програмні модулі для автоматизації опрацювання даних різних модальностей.

Даний акт не є підставою для взаємних фінансових розрахунків.

Генеральний директор
ПП «Мережа Медичних Центрів Родина» Андрій ХАНИК

ПРИВАТНЕ ПІДПРИЄМСТВО МЕРЕЖА МЕДИЧНИХ ЦЕНТРІВ «РОДИНА» 35472249 УКРАЇНА, м. ЛЬВІВ

"ЗАТВЕРДЖУЮ"
Проректор з науково-педагогічної
роботи та міжнародних зв'язків
Національного університету
"Львівська політехніка"
_____ **Наталія ЧУХРАЙ**
"___" _____ 2025 р.

АКТ
про використання наукових результатів
дисертаційної роботи Басистюка Олега Андрійовича
представленої на здобуття наукового ступеня доктора філософії

Комісія в складі: голови комісії – керівник Проєктного офісу Національного університету «Львівська політехніка» к.т.н., доцент Андрущак Н.А. та членів комісії – завідувача кафедри СШІ Мельникової Н.І., професор кафедри СШІ Шаховська Н.Б., що результати дисертаційної роботи Басистюка Олега Андрійовича, зокрема:

- механізм уваги для виявлення найважливіших сегментів інформації;
- метод трансформерів для кодування та декодування модальностей даних;
- метод пошуку і виявлення аномалій в мультимодальних даних.

використано у освітньому проєкті профінансованого в рамках Erasmus+ KA220-HED програм, що виконувалась на кафедрі систем штучного інтелекту і включено до звіту «Effectiveness of Medicine E-learning Distance Courses» (№ проєкту 2022-1-IT02-KA220-HED-000087665).
Отримані автором результати використано:

- при транскрибуванні сценаріїв та змісту дистанційних курсів;
- при складенні методичних рекомендацій по роботі з модальними даними;
- при розробці модуля штучного інтелекту.

Голова комісії:
Керівник Проєктного офісу
к.т.н., доцент — Назарій АНДРУЩАК

Члени комісії:

Завідувач кафедри СШІ
д.т.н., доцент — Наталія МЕЛЬНИКОВА

Професор кафедри СШІ
д.т.н., професор — Наталія ШАХОВСЬКА

“ЗАТВЕРДЖУЮ”
Проректор з науково-педагогічної роботи
Національного університету
"Львівська політехніка"
Олег ДАВИДЧАК
"04" 2025 р.

АКТ

про впровадження в навчальний процес результатів
дисертаційної роботи
Басистюка Олега Андрійовича

Цей акт складено про те, що результати дисертаційної роботи Басистюка Олега Андрійовича впроваджено у навчальний процес кафедри "Систем штучного інтелекту" Національного університету "Львівська політехніка".

Впровадження результатів дисертаційної роботи полягає в їхньому використанні при викладанні навчальних дисциплін як окремих розділів лекційних курсів, так і в циклах лабораторних робіт.

Зокрема для викладання дисципліни "Машинне навчання" для студентів освітньо-кваліфікаційного рівні "бакалавр", що навчаються за напрямком 122 "Комп'ютерні науки" використано такі результати:

- методи попередня обробка даних різних модальностей;
- відсіювання аномальних даних на основі підходів Z-standardization;
- принципи використання рекурентних мереж для опрацювання часових рядів,

А також у курсі «Обробка й аналіз цифрових сигналів» для студентів освітньо-кваліфікаційного рівня «бакалавр», що навчаються за спеціальністю 122 «Комп’ютерні науки», використано такі результати дисертаційного дослідження.

- метод дискретизація неперервних сигналів;
- спосіб ранжування вхідних даних із використанням механізмів уваги;
- методи класифікації емоцій у транскриптах.

Голова комісії:
Директор ІКНІ
д.т.н., професор Наталія ШАХОВСЬКА

Члени комісії:

Завідувач кафедри СШІ
д.т.н., доцент Наталія МЕЛЬНИКОВА

Професор кафедри СШІ
д.т.н., професор Віталій ЯКОВИНА