

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук  
(повна назва)

Кафедра програмної інженерії  
(повна назва)

**КВАЛІФІКАЦІЙНА РОБОТА**  
**Пояснювальна записка**

рівень вищої освіти другий (магістерський)

Дослідження методів глибокого навчання  
для визначення емоцій під час реабілітації пацієнтів  
(тема)

Виконав:  
здобувач 2 року навчання  
групи ІПЗм-23-1  
Андрій ФЕДІН  
(Власне ім'я, ПРІЗВИЩЕ)

Спеціальність 121 – Інженерія програмного  
забезпечення  
(код і повна назва спеціальності)

Тип програми освітньо-наукова

Керівник проф. Наталія БІЛОУС  
(посада, прізвище)

Допускається до захисту  
Зав. кафедри

Кирило СМЕЛЯКОВ  
(підпис) (Власне ім'я, ПРІЗВИЩЕ)

2025 р.

Харківський національний університет радіоелектроніки  
 Факультет \_\_\_\_\_ комп'ютерних наук \_\_\_\_\_  
 Кафедра \_\_\_\_\_ програмної інженерії \_\_\_\_\_  
 Рівень вищої освіти \_\_\_\_\_ другий (магістерський) \_\_\_\_\_  
 Спеціальність \_\_\_\_\_ 121 – Інженерія програмного забезпечення \_\_\_\_\_  
 Тип програми \_\_\_\_\_ освітньо-наукова \_\_\_\_\_  
 Освітня програма \_\_\_\_\_ Інженерія програмного забезпечення \_\_\_\_\_  
 (шифр і назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
 (підпис)  
 «\_\_\_\_» \_\_\_\_\_ 2025 р.

### ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві \_\_\_\_\_ Федіну Андрію Володимировичу \_\_\_\_\_  
 (прізвище, ім'я, по батькові)

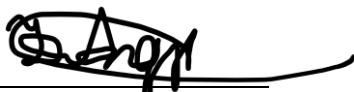
1. Тема роботи «Дослідження методів глибокого навчання для визначення емоцій під час реабілітації пацієнтів»  
 Затверджена наказом по університету від 15.04. 2024р. № 290 Ст
2. Термін подання студентом роботи до екзаменаційної комісії 20.06.2025
3. Вихідні дані до роботи опис досліджуваних методів глибокого навчання для визначення емоцій під час реабілітації пацієнтів для проведення досліджень за обраною предметною областю; мови програмування: Python, середовище розробки Visual Studio Code 2022.
4. Перелік питань, що потрібно опрацювати в роботі аналіз та порівняння існуючих методів глибинного навчання, вибір підходящих архітектур нейронних мереж для дослідження, проектування програмної системи для проведення експериментального дослідження, написання програмних рішень, проведення експериментів та аналіз отриманих результатів.

## КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	16.04.2025	<i>виконано</i>
2	Аналіз предметної галузі і постановка задачі	20.04.2025-04.05.2025	<i>виконано</i>
3	Аналіз існуючих методів	05.05.2025-10.05.2025	<i>виконано</i>
4	Теоретичне дослідження	10.05.2025-20.05.2025	<i>виконано</i>
5	Практичне дослідження	21.05.2025-05.06.2025	<i>виконано</i>
6	Підготовка пояснювальної записки	05.06.2026-14.06.2025	<i>виконано</i>
7	Підготовка презентації та доповіді	15.06.2025	<i>виконано</i>
8	Перевірка на плагіат	15.06.2025	<i>виконано</i>
9	Нормоконтроль	18.06.2025	<i>виконано</i>
10	Рецензування	18.06.2025	<i>виконано</i>
11	Попередній захист	18.06.2025	<i>виконано</i>
12	Занесення диплома в електронний архів	20.06.2025	<i>виконано</i>
13	Допуск до захисту у зав. кафедри	20.06.2025	<i>виконано</i>

Дата видачі завдання 16.04.2025р.

Студент

  
(підпис)

Андрій ФЕДІН

Керівник роботи

(підпис)

проф. Наталія БІЛОУС  
(посада, прізвище, ініціали)

## РЕФЕРАТ / ABSTRACT

Пояснювальна записка містить: 84 с., 20 рис., 5 табл., 20 джерел.

АНАЛІЗ ЕМОЦІЙ, ДОСЛІДЖЕННЯ, ОБРОБКА ЗОБРАЖЕНЬ, РЕАБІЛІТАЦІЯ, НЕЙРОННІ МЕРЕЖІ, AZURE, AZURE BLOB STORAGE, CNN, LSTM, RESNET, VGGNET.

Об'єктом дослідження є методи глибинного навчання розпізнавання емоцій пацієнтів під час фізичних навантажень у реабілітаційних програмах.

Метою роботи є аналіз ефективності застосування глибоких згорткових та рекурентних нейронних мереж для автоматичного розпізнавання емоцій пацієнтів на основі міміки обличчя.

Методами розробки та проектування є дослідження сучасних підходів до обробки зображень, вибір архітектури нейронних мереж, а також створення модульного програмного забезпечення для збору та аналізу даних.

У результаті роботи було розроблено систему, що використовує попередньо навчені моделі (ResNet, VGGNet, LSTM) для класифікації емоцій, з інтеграцією до реабілітаційних програм через RESTful API.

AZURE, AZURE BLOB STORAGE, CNN, EMOTION ANALYSIS, IMAGE PROCESSING, LSTM, NEURAL NETWORKS, REHABILITATION, RESEARCH, RESNET, RNN, VGGNET

The object of study is deep learning methods for recognizing patients' emotions during physical activity in rehabilitation programs.

The purpose of the study is to analyze the effectiveness of using deep convolutional neural networks for automatic recognition of patients' emotions based on facial expressions.

The development and design methods include the study of modern approaches to image processing, the choice of neural network architecture, and the creation of modular software for data collection and analysis.

As a result of the work, a system was developed that uses pre-trained models (ResNet, VGGNet, LSTM) to classify emotions, with integration into rehabilitation programs via RESTful API.

Завідувачу кафедри  
П  
(скорочена назва кафедри)  
проф. Кирилу СМЕЛЯКОВУ  
(вчене звання, сласне ім'я, прізвище)

### ЗАЯВА

щодо самостійності виконання кваліфікаційної роботи та можливості її публікації  
(та/або публікації анотації кваліфікаційної роботи) в електронному архіві  
відкритого доступу EIAr KhNURE

Я, Федін Андрій Володимирович, гр. ПЗм-23-1, здобувач вищої освіти на другому (магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів глибокого навчання для визначення емоцій під час реабілітації пацієнтів», що буде представлена для публічного захисту, виконана самостійно, не містить елементи плагіату і може бути опублікована в репозиторії EIArKhNURE. Всі запозичення з друкованих та електронних джерел мають відповідні посилання.

Я ознайомлений (а) з вимогами академічної доброчесності, згідно з якими виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

Дата 15.06.2025

Підпис



## ЗМІСТ

Вступ.....	9
1 Аналіз предметної галузі .....	11
1.1 Аналіз предметної галузі дослідження .....	11
1.2 Огляд існуючих підходів та їх ефективності .....	12
1.3 Перспективи та актуальність дослідження .....	14
1.4 Огляд основних джерел.....	18
1.5 Існуючі обмеження .....	24
1.6 Постановка задачі .....	25
2 Аналіз існуючих методів .....	27
2.1 Вибір датасету .....	27
2.2 Огляд існуючих методів .....	29
2.3 Порівняння моделей .....	33
3 Теоретичне дослідження .....	37
3.1 Архітектура та проектування ПЗ.....	37
3.2 Обмеження у процесі дослідження .....	39
3.3 Необхідні ресурси для виконання дослідження .....	40
3.4 Проектування системи.....	40
3.5 Проектування структури зберігання даних .....	41
3.6 UI/UX Дизайн Системи .....	44
3.7 Приклади алгоритмів та методів .....	45
3.8 Реалізація алгоритму розпізнавання емоцій .....	47
4 Практичне дослідження.....	52
Висновки .....	63
Перелік джерел посилання .....	65
Перелік джерел посилання за науковими напрямками керівника та науковців кафедри програмної інженерії .....	68
Додаток А Слайди презентації.....	69
Додаток Б Апробація результатів роботи.....	78
Додаток В Звіт з результатами перевірки на унікальність тексту в базі ХНУРЕ .	82

Додаток Г Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015 .....	84
--	----

## ВСТУП

Сучасний етап розвитку медицини та реабілітаційної інженерії характеризується активним впровадженням технологій штучного інтелекту та машинного навчання для підвищення ефективності лікування. Розпізнавання емоцій пацієнтів під час фізичних навантажень у реабілітаційних програмах має важливе значення для моніторингу їхнього психоемоційного стану та адаптації програм реабілітації відповідно до індивідуальних потреб. Це дозволяє оптимізувати процес відновлення, підвищити його ефективність і запобігти негативним емоційним реакціям, які можуть вплинути на прогрес пацієнта.

Особливу актуальність дослідження набуває у зв'язку з необхідністю розробки методів, які здатні враховувати специфіку емоційної реакції пацієнтів у реабілітаційному процесі. Використання методів машинного навчання для аналізу емоцій пацієнтів відкриває нові перспективи для покращення якості реабілітації та персоналізації лікування.

Основною метою роботи є розробка та аналіз методів розпізнавання емоцій пацієнтів під час фізичних навантажень у реабілітаційних програмах з використанням сучасних технологій машинного навчання та комп'ютерного зору. Для досягнення цієї мети необхідно виконати такі задачі:

- проаналізувати існуючі методи розпізнавання емоцій у здорових людей та їхній потенціал для адаптації у реабілітаційних програмах;
- розробити модель розпізнавання емоцій, яка враховує особливості фізичного навантаження;
- виконати тестування та верифікацію запропонованих методів на відповідних наборах даних;
- оцінити можливості впровадження розробленої моделі в практичну реабілітацію.

Об'єктом дослідження є емоційні реакції пацієнтів під час виконання фізичних вправ у реабілітаційних програмах та відслідковування стану здоров'я пацієнтів. Предметом дослідження є методи машинного навчання та

комп'ютерного зору, які використовуються для розпізнавання емоцій у реабілітаційних умовах. Для дослідження та аналізу використовуються наступні методи:

- методи машинного навчання, включаючи згорткові (CNN) та рекурентні (RNN) нейронні мережі для аналізу зображень обличчя та ;
- методи статистичного аналізу для оцінки ефективності розроблених моделей.

Очікується, що у результаті дослідження будуть отримані результати, які дозволять визначити найефективніші підходи до розпізнавання емоцій пацієнтів під час фізичних навантажень. Базуючись на виборі ефективного підходу, застосувати його у розробці практичних рекомендацій щодо інтеграції нових методів у реабілітаційні програми. Також передбачається можливість персоналізації процесу реабілітації та підвищенню його ефективності, орієнтуючись на індивідуальні потреби та результати.

Основні результати даної кваліфікаційної роботи пройшли апробацію на Міжнародній науковій конференції «SCIENTIA» 13 червня 2025 року.

Результати дослідження матимуть практичне значення для медичних установ та реабілітаційних центрів, спрямованих на використання сучасних технологій для покращення здоров'я пацієнтів, їх швидшу реабілітацію та спрощення загального процесу догляду за пацієнтами.

## 1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

### 1.1 Аналіз предметної галузі дослідження

Розуміння емоційного стану пацієнта під час реабілітаційних вправ надзвичайно важливе, оскільки це допомагає лікарям та медичним працівникам вчасно реагувати на стресові або негативні емоції, які можуть виникнути під час лікування. Такі технології дозволяють автоматизувати процес моніторингу емоційного стану пацієнтів, адаптувати реабілітаційні програми під їхні потреби і забезпечити більш ефективний та комфортний процес відновлення.

Існують різні категорії підходів, які направлені на аналіз емоціонального стану людини. Деякі з них мають істотно різну направленість. Перш за все, найкращим варіантом буде згадка про найбільш точний метод – психофізіологічний, який включає в себе різні методи вимірювання фізіологічних змін, серед них: електроенцефалографія, функціональна магнітно-резонансна томографія, гальванічна реакція шкіри, електроміографія та інші. Перелічені методи дозволяють надати досить високу точність показників вимірювань та вони можуть бути використані у реабілітації для моніторингу стану пацієнтів. Але даний підхід має найбільш складні вимоги. По-перше, це вартість обладнання, без якого ці дослідження не будуть мати ніякого сенсу. По-друге, більшість підходів потребують великої кількості сенсорів та датчиків, що значно зменшує зручність та мобільність рухів пацієнта. Окрім того, рухливість має бути обмежена під час деяких досліджень задля виключення неточностей показників.

Ще однією опцією є методи направлені на аналіз голосу та мовлення. Це набагато простіший варіант, ніж попередній, оскільки потребує значено менше обладнання та навичок. Він добре підходить для оцінки вербальної комунікації та використовує штучний інтелект для автоматизації процесу. Але найбільшою перешкодою у запровадженні методів обробки голосу та мовлення є висока чутливість до шуму та якості запису, залежність від окремої конкретної мови, вплив особливостей вимови, інтонацій та змішаної мови, що значно ускладнює

навчання моделей для аналізу аудіо. Це робить його неуніверсальним та досить складним у реалізації.

Особливу роль у сучасному світі віддають комп'ютерному аналізу міміки обличчя, який є одним із найбільш ефективних способів визначення емоцій. Особливо важливим є застосування цього підходу в умовах реабілітації, коли пацієнти виконують фізичні вправи і їхні емоційні реакції можуть варіюватися в залежності від інтенсивності навантаження[1]. Інтеграція цієї технології в процес реабілітації дозволяє надавати більш точні рекомендації щодо подальших кроків у відновленні. Окрім того, найбільшою перевагою є доступність, бо наявність камери для зйомки відео – мінімальна умова для впровадження підходу із аналізу міміки обличчя.

## 1.2 Огляд існуючих підходів та їх ефективності

Існують різні системи з розпізнавання емоцій. Одна з них, Brand24 – це програмний інструмент для моніторингу онлайн-репутації, що аналізує емоційний контекст згадок брендів у соціальних мережах. Хоча ця система не зосереджується безпосередньо на розпізнаванні міміки обличчя, вона використовує аналіз тексту та аудіовізуальних даних для визначення емоційного забарвлення. Аналогічний підхід можна використовувати для інтеграції текстових описів пацієнтів із відео аналізом для розширення функціональності системи розпізнавання емоцій.

Інша – MorphCast for Zoom – це інтерактивна платформа для відеоконференцій, яка використовує технологію розпізнавання емоцій, щоб покращити взаємодію між учасниками. Її алгоритми аналізують вирази обличчя в реальному часі, що дозволяє виявляти емоційний стан кожного учасника. Такий підхід особливо корисний у реабілітаційних програмах, де потрібно адаптувати фізичні вправи залежно від емоційного стану пацієнта.

FaceReader – це одна з найпоширеніших програм для автоматичного аналізу емоцій за мімікою обличчя. Вона використовує передові методи комп'ютерного зору та згорткові нейронні мережі для класифікації емоцій, таких як радість, сум,

злість, здивування, страх тощо. FaceReader також забезпечує аналіз мікроекспресій, що дозволяє виявляти приховані емоції, важливі в реабілітації для розуміння психоемоційного стану пацієнта.

Ще одна технологія розпізнавання емоцій була розроблена компанією Emotient, яку згодом придбала Apple. Хоча детальна інформація про її поточне застосування обмежена, відомо, що недоліками, пов'язаними з її використанням, є потенційне обмеження роботи лише в межах екосистеми Apple, а також закритість алгоритмів та можлива відсутність публічного API для інтеграції зі сторонніми системами.

Таблиця 1.1 – Недоліки програмних систем (таблиця виконана самостійно)

Назва системи	Недоліки
Brand24	<ul style="list-style-type: none"> <li>- зосереджується лише на текстових даних, без аналізу міміки чи відео;</li> <li>- відсутність точного аналізу емоцій окремих користувачів;</li> </ul>
MorphCast for Zoom	<ul style="list-style-type: none"> <li>- високі вимоги до ресурсів пристрою для аналізу в реальному часі;</li> <li>- низька точність в умовах поганого освітлення чи часткового закриття обличчя;</li> </ul>
FaceReader	<ul style="list-style-type: none"> <li>- обмеженість у використанні мультимодальних даних (аудіо, сенсори);</li> </ul>
Emotient (Apple)	<ul style="list-style-type: none"> <li>- обмеження використання лише в екосистемі Apple;</li> <li>- закритість алгоритмів і відсутність API для інтеграції.</li> </ul>

Хоча на даний момент і присутня різноманітність підходів та рівень розвитку технологій розпізнавання емоцій, але водночас їхні обмеження вказують на наявність невирішених проблем та відкривають простір для подальших

досліджень, зокрема у контексті реабілітації пацієнтів під час фізичних навантажень.

### 1.3 Перспективи та актуальність дослідження

Галузь розпізнавання емоцій, особливо при її застосуванні у медичному та реабілітаційному контекстах, демонструє стрімкий розвиток та постійне збагачення новими підходами і технологічними рішеннями. Цей прогрес значною мірою зумовлений зростаючою потребою в об'єктивних інструментах оцінки стану пацієнтів та міждисциплінарним характером досліджень, що поєднують медицину, інженерію та штучний інтелект. Серед актуальних тенденцій, які визначають майбутні перспективи цієї сфери, можна виділити декілька ключових напрямків, що обіцяють суттєво покращити якість реабілітаційного процесу.

Однією з найважливіших сучасних тенденцій є прагнення до інтеграції мультимодальних даних. Це означає, що для більш глибокого та всебічного аналізу емоційного стану пацієнта пропонується використовувати не лише візуальну інформацію. Це особливо є цінним в реабілітації, де емоційні реакції можуть бути складними та багатограними, а їх правильна інтерпретація є критичною для адаптації терапевтичних програм. Звісно, обробка та синхронізація мультимодальних даних представляє собою додаткові технічні виклики, проте потенційні переваги значно їх переважають.

Іншою значущою перспективою є розробка та вдосконалення адаптивних моделей розпізнавання емоцій, зокрема тих, що базуються на архітектурах згорткових нейронних мереж (CNN). Головна ідея полягає у створенні систем, здатних не просто застосовувати загальні шаблони, а навчатися та динамічно підлаштовуватися під індивідуальні характеристики миміки, фізіологічних реакцій та емоційних проявів кожного окремого пацієнта. Такі персоніфіковані моделі обіцяють значно покращити точність аналізу емоцій, оскільки вони можуть враховувати унікальні фізичні та емоційні зміни, що відбуваються з людиною протягом реабілітаційного курсу. Це відкриває шлях до дійсно індивідуалізованої медицини.

Паралельно активно розвивається напрямок автоматизації аналізу емоційного стану в режимі реального часу. Можливість отримувати актуальну інформацію про емоції пацієнта безпосередньо під час виконання реабілітаційних процедур дозволяє медичному персоналу негайно реагувати на будь-які значущі зміни. Наприклад, при виявленні ознак сильного болю, страху чи надмірного стресу, система може сигналізувати про необхідність коригування інтенсивності фізичних навантажень, зміни типу вправ або надання пацієнту відпочинку. Згорткові нейронні мережі відіграють тут ключову роль, оскільки їхня здатність до швидкої та ефективної обробки візуальних даних є фундаментальною передумовою для створення таких інтерактивних систем, що функціонують за принципом замкненого контуру, де стан пацієнта безпосередньо впливає на параметри терапії.

Розглядаючи питання інноваційності та практичності дослідження методів розпізнавання емоцій пацієнта при навантаженнях під час реабілітації, важливо наголосити, що саме використання згорткових нейронних мереж для аналізу міміки є високотехнологічним та новаторським підходом у цій специфічній сфері. Такий підхід знаменує собою перехід від переважно суб'єктивних оцінок стану пацієнта до об'єктивних, кількісних методів, що базуються на даних. Це відкриває широкі можливості для значного покращення адаптації реабілітаційних програм, надаючи медичним фахівцям гнучкі та зручні інструменти для моніторингу та управління емоційним станом пацієнтів. Інноваційність полягає також у спробі застосувати передові розробки в галузі глибокого навчання до вирішення конкретних проблем пацієнтів, що проходять складний шлях відновлення, часто в умовах, далеких від ідеальних лабораторних.

Практична цінність подібних досліджень є багатогранною і полягає, перш за все, у можливості створення та впровадження дійсно персоналізованих програм реабілітації. Завдяки точному та об'єктивному аналізу виразів обличчя, технології машинного навчання, і зокрема CNN, можуть стати основою для розробки програм, які динамічно враховують індивідуальні емоційні реакції пацієнтів під час виконання фізичних навантажень. Це, у свою чергу, неминуче призводить до

підвищення загальної ефективності лікувального процесу, оскільки дозволяє уникнути ситуацій, що викликають надмірний дискомфорт або негативні емоції, які можуть демотивувати пацієнта та сповільнити його прогрес. Персоналізація може проявлятися у регулюванні складності та тривалості вправ, чергуванні навантажень та відпочинку, або навіть у виборі специфічних методик, що найкраще відповідають поточному стану пацієнта.

Одним з ключових аспектів інноваційності та практичної значущості є адаптація методів розпізнавання емоцій до специфічних умов фізичних навантажень, характерних для реабілітаційного процесу. На відміну від багатьох існуючих моделей, які переважно розроблені для розпізнавання "класичних" емоцій у здорових людей в спокійних, контрольованих умовах, пропонується підхід має враховувати набагато складніший контекст. Необхідно брати до уваги такі додаткові фактори, як можливий фізичний біль, втома або стрес, що часто супроводжують виконання фізичних вправ пацієнтами з обмеженими можливостями. Емоційні прояви в таких умовах можуть відрізнятися від стандартних, і саме здатність системи розрізняти ці нюанси підвищує точність визначення емоційного стану та дозволяє створювати більш гнучкі та ефективні реабілітаційні програми.

Інтеграція системи розпізнавання емоцій для роботи в реальному часі є ще однією важливою практичною перевагою, яка напряму впливає на якість реабілітації. Здатність згорткових нейронних мереж до швидкого, практично миттєвого аналізу послідовності відеокадрів дає змогу системі негайно фіксувати та інтерпретувати емоційні зміни на обличчі пацієнта під час виконання вправ. Це, в свою чергу, дозволяє медичному персоналу або автоматизованій системі адаптувати інтенсивність та характер фізичного навантаження "на льоту", забезпечуючи не тільки більш ефективне та безпечне відновлення, але й значно підвищуючи психологічний комфорт пацієнта. Ефективність такого підходу значною мірою залежить від використання сучасних та добре оптимізованих архітектур CNN, наприклад, таких як ResNet та VGGNet, або їх більш легких варіацій, які здатні досягати високої точності розпізнавання емоцій навіть за умов

обмежених обчислювальних ресурсів. Дані мережі, завдяки своїй глибині та архітектурним особливостям, аналізують складні просторові взаємозв'язки між різними частинами обличчя, що дозволяє точніше інтерпретувати емоційні вирази навіть у складних, неідеальних умовах спостереження.

Також слід відзначити значний потенціал до самонавчання та самовдосконалення таких інтелектуальних систем. Модель, що використовує CNN, може бути розроблена таким чином, щоб вона поступово адаптувалася до індивідуальних особливостей міміки та емоційних проявів конкретного пацієнта через механізми зворотного зв'язку. Цей зворотний зв'язок може бути як явним (наприклад, періодичне підтвердження або коригування оцінок системи з боку медичного персоналу), так і неявним (система аналізує кореляції між певними типами вправ, фізіологічними показниками та емоційними реакціями). Це означає, що система здатна неперервно вдосконалюватися з часом, накопичуючи досвід та враховуючи специфічні емоційні реакції конкретного пацієнта на фізичні навантаження. Такий підхід веде до суттєвого підвищення точності розпізнавання та рівня персоналізації реабілітаційних програм у довгостроковій перспективі.

Таким чином, дослідження та розробка методів розпізнавання емоцій на базі згорткових нейронних мереж в умовах реабілітації є надзвичайно актуальним та інноваційним науково-практичним напрямком. Він має значний потенціал не лише для підвищення якості та ефективності самого реабілітаційного процесу для окремих пацієнтів, але й для збору цінних даних, які можуть сприяти глибшому розумінню психоемоційних аспектів відновлення та розробці нових, більш досконалих методик реабілітації в майбутньому.

Але ефективність практичного застосування буде також залежати від використання сучасних архітектур CNN, таких як ResNet та VGGNet, що дозволяють досягати високої точності розпізнавання емоцій. Ці мережі аналізують взаємодію між різними частинами обличчя, що дає змогу точніше інтерпретувати емоційні вирази, навіть у складних умовах. Сама модель буде мати здатність до самонавчання: вона адаптується до індивідуальних

особливостей пацієнта через зворотний зв'язок. Це дозволяє моделі вдосконалюватися з часом, враховуючи емоційні реакції пацієнта на фізичні навантаження, що підвищує точність та персоналізацію реабілітаційних програм.

#### 1.4 Огляд основних джерел

Для систематизованого аналізу було відібрано декілька джерел, які задовольняють таким критеріям:

- авторитетність – публікації рецензовані, мають посилання на авторитетні наукові ресурси;
- актуальність – усі джерела опубліковані в період 2019–2025 років, що забезпечує врахування останніх досягнень у галузі;
- об'єктивність – дослідження базуються на чітких експериментальних методах і представленні даних;
- достовірність інформації – джерела містять детальний опис застосованих методів, що підвищує надійність результатів.

Огляд обраних джерел висвітлює сучасні досягнення у сфері розпізнавання емоцій із використанням глибокого навчання та мультимодального аналізу. У роботі "Advancements in Emotion Classification via Facial and Body Gesture Analysis: A Survey" (2024) [2] запропоновано новітні підходи до аналізу емоцій через поєднання міміки та жестів. Особлива увага приділяється мультимодальному аналізу, який включає інтеграцію різних джерел інформації, а також полегшеним моделям CNN для роботи в реальному часі. Це є ключовим фактором для інтерактивних додатків, таких як реабілітація та інші сценарії, що вимагають швидкої обробки даних.

Інший важливий внесок представлено у статті "Deep Learning Based Emotion Recognition and Visualization of Figural Representation" (2022) [3], де розглядається поліпшена модель CNN-BiLSTM для розпізнавання емоцій, що досягає точності 98.75%. Основна увага зосереджена на розпізнаванні емоцій через обробку мовлення та зображень, а також їх візуалізацію. Робота демонструє

перспективність інтеграції глибокого навчання у процеси оцінки емоцій, що може бути адаптовано до реабілітаційних програм для пацієнтів.

У статті "Sensing Human Emotion using Emerging Machine Learning Techniques" (2024) [4] досліджується методика використання CNN для автоматичного розпізнавання емоцій. Основна увага приділена попередній обробці даних, класифікації емоцій та їх практичному застосуванню. Результати демонструють потенціал таких систем для покращення якості взаємодії людини з машиною, зокрема у медичних контекстах, де точне розпізнавання емоцій може суттєво вплинути на ефективність лікування.

У статті "Дослідження методів для розробки програмної системи розпізнавання емоцій та визначення стану здоров'я людини" (2020) [5] досліджуються нейронні мережі для класифікації емоцій, зокрема для виявлення болю через аналіз негативних емоцій. Автори тренували моделі на датасетах "fer2013" і "RAV-F", виявивши, що спеціалізований набір даних показує кращу ефективність у розпізнаванні болю. Використані методи включали очищення зображень, вирівнювання облич і налаштування параметрів тренування. Експерименти підтвердили, що якість розпізнавання значно залежить від специфічності набору даних, що робить це дослідження цінним для розробки систем медичного моніторингу.

Тематичне групування джерел дозволяє виокремити кілька ключових напрямків дослідження. Перш за все, глибокі нейронні мережі, такі як CNN-ViLSTM, демонструють значний потенціал для розпізнавання емоцій, особливо в реабілітаційних умовах, де необхідна висока точність і адаптивність.

Ще одним важливим аспектом є мультимодальний аналіз емоцій, який поєднує дані про міміку, жести та мовлення. Такий підхід забезпечує більш точне й комплексне розуміння емоційного стану пацієнтів, що особливо важливо в контексті їх реабілітації.

Практичне застосування досліджень у медичній сфері включає інтеграцію розроблених систем для моніторингу емоцій пацієнтів у реальних умовах. Це

відкриває перспективи для підвищення якості взаємодії між пацієнтом і медичним персоналом, а також сприяє покращенню реабілітаційних процесів.

Основні теорії, концепції та моделі в дослідженнях розпізнавання емоцій зосереджені на використанні глибоких нейронних мереж, мультимодального аналізу та методів обробки даних. У роботі "Deep Learning Based Emotion Recognition and Visualization of Figural Representation" (2022) представлено модель CNN-BiLSTM, яка поєднує конволюційні нейронні мережі для виділення особливостей зображень із довготривалою пам'яттю для аналізу послідовних даних. Ця модель виявилася надзвичайно ефективною у мультимодальному аналізі, демонструючи точність понад 98.75%.

Мультимодальний підхід до аналізу емоцій також висвітлюється у статті "Advancements in Emotion Classification via Facial and Body Gesture Analysis: A Survey" (2024). Дослідження акцентує увагу на інтеграції аналізу міміки та жестів, що дозволяє значно підвищити точність класифікації емоцій. Особливо перспективною виявилася концепція часткового трансферного навчання, яка сприяє адаптації моделей до нових наборів даних і розширює можливості їх практичного застосування.

Методи обробки даних та автоматизації класифікації емоцій розглядаються у роботі "Sensing Human Emotion using Emerging Machine Learning Techniques" (2024). У цьому дослідженні висвітлено підхід до попередньої обробки зображень та використання CNN для ідентифікації семи базових емоцій, таких як злість, радість і страх. Використання алгоритмів машинного навчання забезпечує високу точність, що підкреслює практичну цінність цих технологій у різних сферах, зокрема в медицині.

Розглядаючи іншу статтю – «Transfer Learning for Facial Expression Recognition» [6] слід виокремити використання трансферного навчання з архітектурами VGG-19 та ResNet-152. Замість того, щоб навчати складну нейронну мережу з нуля, що вимагало б величезних обсягів даних, буде взято попередньо навчені моделі й "донавчено" їх на своїх даних. Стаття показує, що такий підхід дає хороші результати, що значно покращує отримання необхідних

даних. Використовуючи цей підхід у даному дослідженні, значно зменшуються необхідні комп'ютерні потужності.

Дуже цінними є конкретні технічні рекомендації щодо підготовки даних та налаштування моделей. Наприклад, автори говорять про важливість специфічної для набору даних попередньої обробки, зокрема зміну розміру зображень (вони використовували 124x124 пікселі), аугментацію даних (штучне збільшення кількості прикладів шляхом їх модифікації, наприклад, горизонтального віддзеркалення) та вибіркоче "заморожування" певних шарів нейромережі під час донавчання. Деякі шари, які вже навчилися розпізнавати загальні риси на великих датасетах, залишаються незмінними, а донавчаються лише верхні шари, що спеціалізуються на конкретній задачі. Це допомагає уникнути перенавчання та покращити узагальнюючу здатність моделі.

Серед інших кроків у алгоритмі розпізнавання емоцій, автори статті використовують попереднє знаходження обличчя на зображенні перед тим, як розпізнавати емоції, а саме за допомогою алгоритму Віюлі-Джонса.

Щодо результатів та очікувань, стаття демонструє високу точність на контрольованих датасетах, як-от СК+ (точність 0.98), але водночас чесно говорить про виклики при роботі з більш "дикими", зашумленими даними з інтернету (датасет FER2013, де точність була значно нижчою - 0.44). Це досить важлива інформація для даного дослідження, оскільки умови реабілітації можуть бути далекими від ідеальних лабораторних – різне освітлення, часткові перекриття обличчя, рухи пацієнта. Стаття також підкреслює важливість збалансованості датасетів, оскільки дисбаланс класів емоцій може впливати на продуктивність моделі. Тому це буде взято до уваги при виборі датасету.

«Bilous, N., Malko, V., Frohme, M., & Nechiporenko, A. (2024) Comparison of CNN-based architectures for object detection to define what class of object is, pp1-15, AI (in printed)» [7] - дана стаття проводить порівняльний аналіз сучасних CNN-архітектур, таких як різні версії YOLO (від v4 до v8), Faster R-CNN, SSD та EfficientDet. Розпізнавання обличчя (як об'єкта) та його ключових рис є першим кроком перед аналізом міміки для визначення емоцій, тож це становить

практичну цінність при підготовці зображень. Розуміння того, які архітектури краще справляються з детекцією в складних умовах, може допомогти обрати надійний "фундамент" для майбутньої програмної системи. Окремо слід звернути увагу на висновок авторів про те, що хоча різні нейромережі показують кращі результати за окремими метриками, YOLO загалом перевершує їх за всіма показниками, демонструючи найкращі результати за точністю (mAP), F1-Score та швидкістю обробки (FPS).

Стаття наголошує на важливості таких метрик оцінки, як середня точність (mAP), Precision, Recall, F1-Score та здатність мережі працювати в реальному часі (FPS). Ці метрики є стандартними для задач детекції об'єктів, і ви можете використовувати їх не тільки для оцінки якості детекції облич, але й адаптувати для оцінки якості розпізнавання емоцій. Особливо важливим для реабілітаційного процесу є аспект роботи в реальному часі, оскільки емоційна реакція пацієнта може бути швидкою, і система повинна вчасно її фіксувати. Той факт, що YOLOv8 показує 48 FPS, є хорошим показником для таких завдань.

Автори використовували різноманітні набори даних для тестування моделей, включаючи COCO2017, SARD, SeaDronesSee та VisDrone2019, які представляють складні та реалістичні сценарії. Хоча ці датасети не специфічні для емоцій, досвід роботи з ними та підходи до підготовки даних (нормалізація, аугментація, розділення на вибірки у співвідношенні 80:10:10) можуть бути корисними при роботі з даними зображень пацієнтів. Стаття підкреслює, що умови середовища суттєво впливають на продуктивність моделей, і це треба враховувати, адже реабілітаційні зали також можуть мати різне освітлення, фони та інші фактори, що ускладнюють розпізнавання.

Розглядаючи внесок у область дослідження можна виділити наступні можливі покращення базуючись на обраній літературі:

- розробка ефективних моделей глибокого навчання для розпізнавання емоцій у реальному часі значно покращує можливості моніторингу емоцій пацієнтів у реабілітаційних умовах;

- інтеграція мультимодального підходу дозволяє отримати більш повну картину емоційного стану пацієнтів, враховуючи як візуальні, так і невербальні сигнали;
- використання трансферного навчання та легких архітектур нейронних мереж підвищує адаптивність моделей до нових умов.

Характеристика методів та підходів, описаних у розглянутих джерелах, відображає широкий спектр інструментів і технологій, що використовуються для розпізнавання емоцій. Методи аналізу зображень засновані на використанні конволюційних нейронних мереж (CNN), які дозволяють виділяти ключові характеристики обличчя, такі як контури, форма очей і рота. Зокрема, моделі на основі CNN-BiLSTM демонструють високу ефективність у розпізнаванні послідовних емоційних сигналів, що робить їх перспективними для задач, пов'язаних із аналізом у реальному часі.

Мультимодальні підходи, що поєднують аналіз міміки та жестів, дозволяють враховувати ширший спектр емоційних сигналів. Це не лише підвищує точність класифікації, але й поглиблює розуміння емоційного стану, створюючи більш комплексні системи аналізу.

Попередня обробка даних відіграє ключову роль у забезпеченні якісного функціонування цих методів. Усі розглянуті джерела підкреслюють важливість очищення зображень, вирівнювання облич та видалення шумів на етапі підготовки даних. Ці процедури не лише покращують якість вхідної інформації, але й сприяють підвищенню загальної точності аналізу емоцій, що є критичним для практичних застосувань у медицині та інших сферах.

Оцінка ефективності методів показала, що моделі на основі CNN-BiLSTM досягли найвищої точності серед усіх досліджених алгоритмів, значно знижуючи рівень помилкових класифікацій. Використання трансферного навчання дозволяє зменшити потребу у великих наборах даних, що є особливо важливим для медичних застосувань. Крім того, мультимодальні підходи демонструють більшу адаптивність до реальних умов завдяки інтеграції різних джерел інформації.

Розглянуті джерела, є надзвичайно актуальними для сучасного стану проблеми. Використання штучного інтелекту, зокрема глибокого навчання, для розпізнавання емоцій стає критично важливим у контексті реабілітаційної медицини, де точний моніторинг емоційного стану пацієнтів є ключовим фактором успішного відновлення. Зростаюча роль мультимодальних технологій, яка підкреслюється у дослідженнях, відповідає сучасним викликам інтерактивної медицини та дистанційної підтримки. Це доводять результати дослідження, що поєднують аналіз міміки, жестів і мовлення для отримання повнішої картини емоційного стану.

### 1.5 Існуючі обмеження

Існуючі рішення у системах не є ідеальними та мають власні обмеження. По-перше, слід згадати залежність від умов зйомки. Оскільки CNN обробляють зображення, результати розпізнавання емоцій можуть залежати від таких умов, як освітлення, якість зображення та наявність рухів пацієнта під час виконання вправ.

Суттєвим обмеженням є також кількість та якість даних, доступних для навчання моделей. Для ефективного навчання згорткових нейронних мереж, особливо для досягнення високої точності та здатності до узагальнення, потрібні великі, різноманітні та ретельно розмічені набори зображень або відео. Збір таких даних може бути особливо проблематичним для специфічних випадків, до яких належать реабілітаційні програми. Важко зібрати достатню кількість реалістичних даних з коректними емоційними мітками безпосередньо від пацієнтів під час фізичних навантажень, дотримуючись при цьому етичних норм та забезпечуючи приватність.

Персоналізація моделей – більшість моделей CNN розроблені для загальних умов, але вони можуть не враховувати індивідуальні особливості пацієнтів у реабілітації, такі як фізичні обмеження або специфічні емоційні реакції, пов'язані з їхнім станом здоров'я.

Також одним з обмежень є кількість даних для навчання: Для навчання CNN потрібні великі набори помічених зображень, що може бути проблемою для специфічних випадків, як-от реабілітаційні програми, де важко зібрати достатньо даних з мітками для точного навчання моделі.

## 1.6 Постановка задачі

Метою цієї кваліфікаційної роботи є розробка методу автоматичного розпізнавання емоцій пацієнтів під час фізичних навантажень у реабілітаційних програмах на основі аналізу міміки обличчя. Правильне оцінювання емоційного стану пацієнта є важливим аспектом при коригуванні реабілітаційних навантажень, що дозволяє підвищити ефективність лікування, поліпшити фізичний та психологічний комфорт пацієнта. У рамках цього проєкту буде створена система, яка здатна аналізувати вирази обличчя в реальному часі, щоб адаптувати інтенсивність вправ до емоційного стану пацієнта.

Попередні дослідження в цій галузі показали, що подібні системи вже були розроблені, зокрема для розпізнавання емоцій на основі міміки обличчя, але вони часто мають обмежену точність або недостатню адаптивність до умов реабілітації, де емоційні реакції пацієнтів можуть бути складними та змінюватися залежно від навантаження. Тому необхідно не лише створити автоматичну систему розпізнавання емоцій, але й покращити її точність та адаптивність порівняно з існуючими рішеннями.

Для реалізації цієї мети необхідно вирішити ряд задач. По-перше, аналіз існуючих методів розпізнавання емоцій. Здійснити огляд основних підходів до розпізнавання емоцій через міміку обличчя, визначити їх сильні та слабкі сторони та оцінити можливості адаптації цих методів до умов реабілітації. Також слід розробити модель розпізнавання емоцій. Це передбачає створення системи, яка буде використовувати згорткові нейронні мережі (CNN) для аналізу виразів обличчя, щоб класифікувати емоції, такі як радість, сум, гнів, здивування, страх тощо. Модель повинна бути здатною працювати в реальному часі з відео або зображеннями пацієнтів під час фізичних вправ.

Наступним кроком буде інтеграція та тестування моделі. Зробити інтеграцію моделі в систему реабілітації, щоб вона могла на основі виявлених емоцій коригувати фізичні навантаження. Модель має бути перевірена на реальних даних і надано оцінки її точності та ефективності.

Після виконання усіх попередніх пунктів необхідно буде виконати роботи щодо верифікації та покращення результатів. Після інтеграції проведемо тестування в реальних умовах реабілітації для оцінки результатів розпізнавання емоцій і можливості адаптації навантажень на основі цих результатів. Важливо буде оцінити точність та швидкодію моделі в умовах різних типів навантажень.

Наостанок необхідно провести оцінку практичної ефективності. Зробити висновки щодо доцільності використання розробленого підходу в реабілітаційних програмах, оцінити можливість інтеграції моделі в різні медичні установи та її потенціал для поліпшення результатів лікування.

## 2 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ

### 2.1 Вибір датасету

Якість та характеристики наборів даних (датасетів) відіграють фундаментальну роль у розробці та навчанні ефективних моделей машинного навчання, особливо в такій складній задачі, як розпізнавання емоцій. Для даного дослідження, що фокусується на аналізі виразів обличчя, вибір відповідних датасетів є критичним етапом, який значною мірою визначатиме надійність та узагальнюючу здатність розробленої моделі. Загалом, датасети для розпізнавання емоцій можуть містити статичні зображення облич або відеопослідовності, зібрані як у контрольованих лабораторних умовах, так і в більш природних, "диких" умовах (in-the-wild). Емоційні мітки в таких датасетах можуть представляти базові категорії емоцій, наприклад, за класифікацією Пола Екмана, або описуватися в рамках багатовимірних моделей, таких як простір валентності-збудження.

Для попереднього навчання та оцінки моделей у даній роботі розглядається використання загальнодоступних наборів даних, серед яких особливу увагу приділено FER-2013 та AffectNet. Ці датасети широко використовуються в дослідженнях з розпізнавання емоцій і слугують важливими бенчмарками для порівняння ефективності різних архітектур згорткових нейронних мереж.

Датасет FER-2013 [8] (Facial Expression Recognition 2013) виник у рамках конкурсу на платформі Kaggle і складається зі значної кількості зображень людських облич. Кожне зображення представляє одну з семи базових емоцій: злість, відраза, страх, радість, сум, здивування та нейтральний стан. Зображення в FER-2013 є відносно невеликими за розміром (зазвичай 48x48 пікселів) і представлені у відтінках сірого. Важливою характеристикою цього датасету є те, що він містить зображення, зібрані "в дикій природі", тобто в різноманітних неконтрольованих умовах освітлення, ракурсів та якості. Це робить його складним, але водночас цінним для тестування здатності моделей узагальнювати знання на реалістичні дані. Разом з тим, через спосіб збору та автоматичне мічення, FER-2013 може містити певну кількість зашумлених або некоректно

класифікованих прикладів, що є викликом при навчанні моделей. Дослідження, в яких автори тренували моделі на датасеті "fer2013", підтверджують його використання в академічній спільноті.

AffectNet [9] є одним із найбільших на сьогоднішній день датасетів для розпізнавання емоцій за виразами обличчя, що також містить зображення, зібрані з різних джерел. Він був створений шляхом збору зображень з Інтернету за допомогою ключових слів, пов'язаних з різними емоціями. AffectNet включає сотні тисяч зображень, що значно перевищує обсяги багатьох інших датасетів. Емоційні мітки в AffectNet охоплюють як категоріальні класифікації (наприклад, базові емоції), так і мітки в просторі валентності та збудження, що надає додаткові можливості для дослідження. Перевагами AffectNet є його великий розмір та значна різноманітність зображень, що сприяє навчанню більш робастних моделей. Однак, як і багато великих датасетів, зібраних автоматизовано, він може характеризуватися певним дисбалансом класів емоцій та потенційною неоднорідністю якості міток.

При виборі датасетів для даного дослідження враховувалися такі критерії, як релевантність до задачі розпізнавання емоцій за мімікою обличчя, наявність достатньої кількості прикладів для навчання згорткових нейронних мереж, доступність даних та їх відповідність умовам, в яких планується тестування та застосування розроблюваної системи. Оскільки метою є створення системи, здатної працювати в реабілітаційних умовах, які можуть бути неконтрольованими, перевага надається датасетам, що відображають різноманітність реальних життєвих ситуацій.

Після вибору основного датасету або датасетів, наступним важливим етапом є підготовка даних для подачі в нейронну мережу. Цей процес зазвичай включає декілька ключових кроків. Спочатку виконується детекція обличчя на зображенні, для чого можуть використовуватися такі алгоритми, як каскади Хаара або більш сучасні підходи на основі CNN, наприклад, MTCNN. Після локалізації обличчя часто проводиться його вирівнювання для нормалізації пози та ракурсу. Далі зображення обличчя нормалізуються за розміром до стандартного, який

вимагає обрана архітектура нейронної мережі, та може здійснюватися перетворення у відтінки сірого для зменшення обчислювальної складності. Також важливими є нормалізація яскравості та контрасту для зменшення впливу умов освітлення, а також можливе застосування фільтрів для зменшення шуму. Для збільшення обсягу навчальної вибірки та підвищення стійкості моделі до різних варіацій широко застосовуються техніки аугментації даних, такі як випадкові повороти, зсуви, масштабування, зміна яскравості або горизонтальне віддзеркалення зображень. Усі ці кроки спрямовані на те, щоб забезпечити модель якісними та консистентними даними, що є запорукою успішного навчання.

## 2.2 Огляд існуючих методів

Для вирішення задачі автоматичного розпізнавання емоцій пацієнтів на основі аналізу міміки обличчя основна увага приділяється використанню згорткових нейронних мереж (Convolutional Neural Networks, CNN). Цей вибір ґрунтується на їхніх доведених перевагах та високій ефективності в задачах комп'ютерного зору, особливо в аналізі зображень.

Згорткові нейронні мережі, або CNN, зарекомендували себе як провідна технологія в галузі комп'ютерного зору. Їхня особлива ефективність проявляється у задачах аналізу зображень, до яких належить і розпізнавання емоцій за виразами обличчя. Адаптивність та результативність CNN роблять їх ключовим інструментом для розробки систем автоматичного розпізнавання емоцій, особливо важливих у контексті реабілітаційних програм.

Згорткові нейронні мережі мають низку суттєвих переваг, що роблять їх оптимальним вибором для розпізнавання емоцій пацієнтів. Однією з ключових переваг є автоматичне вилучення ознак. На відміну від традиційних методів машинного навчання, CNN не потребують ручного визначення та вилучення специфічних ознак з зображення, таких як відстані між ключовими точками обличчя чи текстурні дескриптори. Згорткові шари мережі автоматично навчаються ідентифікувати релевантні патерни – від простих елементів, як краї та кути на початкових шарах, до складніших структур, наприклад, очі, ніс, рот, та

їхніх конфігурацій, що відповідають певним емоціям, на глибших шарах. Це значно спрощує процес розробки системи та часто призводить до вищої точності, оскільки мережа самостійно знаходить найбільш інформативні ознаки.

Також важливим є те, як архітектура CNN природним чином враховує двовимірну структуру зображень, забезпечуючи врахування просторової ієрархії та локальності. Згорткові операції застосовуються до невеликих локальних ділянок вхідного зображення, так званих рецептивних полів, що дозволяє моделі ефективно виявляти локальні патерни, наприклад, форму губ чи вигин брів. Послідовне застосування згорткових шарів та шарів підвибірки, відомих як *pooling*, створює ієрархічне представлення ознак, де глибші шари агрегують інформацію з ширших областей вхідного зображення. Така ієрархічна структура добре відповідає природі візуальної інформації, зокрема організації рис обличчя при вираженні емоцій.

Крім того, завдяки використанню шарів підвибірки, CNN здатні демонструвати певну стійкість, або інваріантність, до невеликих зсувів, поворотів, масштабування та деформацій об'єктів на зображенні. Це є надзвичайно важливим для розпізнавання емоцій в реальних умовах, оскільки обличчя пацієнта не завжди буде ідеально вирівняне відносно камери, може змінювати розмір або бути частково деформованим через міміку.

Різноманітні архітектури CNN, такі як VGG, ResNet, Inception, EfficientNet, продемонстрували високу точність у багатьох стандартних тестах, або бенчмарках, з розпізнавання зображень. Це включає бази даних емоцій, такі як FER2013, AffectNet, CK+[10], що підтверджує їхню здатність ефективно вирішувати складні задачі класифікації та забезпечувати сучасний рівень ефективності.

Значною перевагою є можливість перенесення навчання. Існує велика кількість моделей CNN, попередньо навчених на величезних наборах даних загального призначення, наприклад, ImageNet. Ці моделі вже навчилися розпізнавати загальні візуальні ознаки низького та середнього рівня. Їх можна успішно адаптувати, донавчити або здійснити "fine-tuning", для специфічної

задачі розпізнавання емоцій, навіть маючи обмежений обсяг цільових даних, наприклад, зображень пацієнтів у реабілітаційних умовах. Такий підхід значно прискорює процес навчання, зменшує потребу у великих спеціалізованих датасетах та часто призводить до покращення кінцевої точності моделі.

Розуміння технічних аспектів функціонування CNN також є важливим. Основними будівельними блоками, що визначають їхню функціональність, є згорткові шари. Вони застосовують набір фільтрів, або ядер згортки, до вхідного зображення, виділяючи важливі локальні ознаки, такі як контури, текстури, форми частин обличчя та інші характеристики, що є релевантними для розпізнавання емоційного виразу. Після згорткових шарів зазвичай розташовуються пулінгові шари, мета яких - зменшити просторову розмірність карт ознак, зберігаючи при цьому найважливішу інформацію. Це допомагає зменшити обчислювальні витрати, кількість параметрів у мережі та забезпечує певну інваріантність до зсувів. Після того, як згорткові та пулінгові шари виділили та узагальнили ієрархічні ознаки, повнозв'язні шари виконують безпосередньо класифікацію, визначаючи приналежність зображення до одного з класів емоцій на основі навчених ваг.

Поряд зі згортковими архітектурами, що аналізують статичні зображення, інший важливий напрямок у розпізнаванні емоцій пов'язаний з аналізом динаміки, для чого використовуються рекурентні нейронні мережі.

Окрім згорткових нейронних мереж, у задачах розпізнавання емоцій також розглядаються рекурентні нейронні мережі (RNN), зокрема їхній більш просунутий варіант - мережі з довгою короткостроковою пам'яттю (LSTM). На відміну від CNN, головна сила RNN полягає в аналізі послідовних даних, де важливий порядок елементів, наприклад, кадрів у відео. Оскільки RNN не призначені для вилучення просторових ознак безпосередньо з зображень, на практиці використовується гібридний підхід CNN-LSTM. У такій архітектурі CNN виступає в ролі екстрактора ознак: вона обробляє кожен окремий кадр і перетворює його на компактний векторний опис. Потім послідовність цих

векторів подається на вхід LSTM, яка аналізує динаміку зміни виразу обличчя в часі.

Теоретична перевага такого підходу полягає в здатності моделі розуміти контекст і розрізняти емоції, що розвиваються поступово, що може підвищити стійкість та точність прогнозів. Однак, цей метод є більш складним у реалізації та вимагає відео-датасетів для ефективного навчання.

Вибір CNN як основного методу для розпізнавання емоцій пацієнтів у цій магістерській роботі обґрунтований їхньою доведеною ефективністю в аналізі візуальної інформації, зокрема міміки. Мета роботи полягає у розробці системи, що автоматично аналізує вирази обличчя пацієнтів під час фізичних навантажень у реабілітаційних програмах. CNN дозволяють реалізувати таку систему, забезпечуючи високу точність класифікації емоцій. Планується використання попередньо навчених моделей CNN, таких як ResNet або VGGNet, з подальшим їх донавчанням на специфічних даних. Це узгоджується з перевагами підходу перенесення навчання, дозволяючи досягти хороших результатів навіть за умов обмеженої кількості спеціалізованих навчальних даних.

Незважаючи на значні переваги, використання CNN також пов'язане з певними обмеженнями та викликами. Одним з таких викликів є залежність від умов зйомки, таких як освітлення, ракурс камери, та рухи пацієнта. Іншим важливим аспектом є необхідність персоналізації моделей, оскільки стандартні моделі можуть не враховувати індивідуальні особливості міміки пацієнтів у реабілітації. Також слід зважати на вимоги до обсягу та якості навчальних даних, оскільки для ефективного навчання CNN потрібні великі набори ретельно розмічених зображень.

Порівнюючи CNN з іншими підходами, такими як психофізіологічні методи, можна відзначити, що останні, хоч і є високоточними, вимагають дорогого обладнання та обмежують рухливість пацієнта. На цьому тлі комп'ютерний аналіз міміки обличчя за допомогою CNN виглядає більш збалансованим рішенням для поставленої задачі. Цей метод є безконтактним, може працювати в реальному часі та, завдяки сучасним архітектурам і технікам

перенесення навчання, здатен досягати високої точності, що є ключовим для моніторингу стану пацієнта під час фізичних навантажень.

### 2.3 Порівняння моделей

Після огляду згорткових нейронних мереж як основного інструменту для вирішення поставленої задачі, важливо провести порівняльний аналіз ключових архітектур, що розглядаються в рамках даної роботи. Це дозволить глибше зрозуміти їхні сильні та слабкі сторони, а також обґрунтувати їх вибір для дослідження ефективності розпізнавання емоцій. Основну увагу слід приділити архітектурам VGGNet, ResNet та Inception, які стали класичними у сфері комп'ютерного зору.

Для наочного узагальнення їхніх ключових відмінностей, наведемо порівняльну таблицю.

Таблиця 2.3.1 – Порівняльний аналіз архітектур CNN (таблиця виконана самостійно)

Параметр	VGGNet	ResNet	Inception
Основна ідея	Простота та глибина рахунок послідовних малих фільтрів	Залишкові з'єднання (skip connections) для боротьби зі згасанням градієнтів у надглибоких мережах	"Ширина" мережі, паралельні згортки на різних масштабах в межах одного "Inception-модуля"
Глибина	Глибока (16-19 шарів), але подальше значне збільшення	Дуже глибока (50, 101, 152+ шарів)	Глибока (напр., 22 шари у GoogLeNet), але з меншою

Кінець таблиці 2.3.1

Параметр	VGGNet	ResNet	Inception
	проблематичне		кількістю параметрів, ніж у VGG
Обчислювальна складність	Висока, велика кількість параметрів	Помірна (ефективніша за VGG при тій самій глибині)	Помірна (розроблена як більш ефективна альтернатива VGG)
Ключова перевага	Надійність, архітектурна простота, чудова базова модель	Досягнення найвищої точності, здатність ефективно навчати надглибокі мережі	Висока ефективність при менших обчислювальних витратах, аналіз ознак на різних масштабах
Придатність для Transfer Learning	Дуже висока, наявність якісних попередньо навчених моделей	Дуже висока, часто є стандартом для досягнення високої точності "з коробки"	Висока

Ці архітектури базуються на різних підходах побудови глибоких нейронних мереж. Архітектура VGGNet дотримується принципу простоти та однорідності, демонструючи, що значне збільшення глибини мережі за допомогою послідовних згорткових фільтрів малого розміру (3x3) може призвести до значного покращення точності. На противагу цьому, ResNet вводить революційну на той час концепцію залишкових з'єднань, або "skip connections", яка дозволяє

градієнтам безперешкодно проходити через дуже глибокі шари, ефективно борючись із проблемою їх згасання, що була перешкодою для навчання надглибоких мереж. Архітектура Inception, в свою чергу, реалізує ідею "широких" мереж. В її основі лежить "Inception-модуль", який виконує паралельні згорткові операції з різними розмірами фільтрів в межах одного шару, що дозволяє моделі одночасно аналізувати візуальні ознаки на різних масштабах.

За критеріями глибини та складності, ResNet є безумовним лідером, оскільки ця архітектура дозволила успішно тренувати мережі глибиною понад сто шарів, що є критично важливим для вивчення тонких та складних мімічних нюансів. VGGNet також є глибокою мережею, проте її архітектура є більш вимогливою до обчислювальних ресурсів та пам'яті через велику кількість параметрів у повнозв'язних шарах. Архітектура Inception, особливо завдяки активному використанню згорток розміром 1x1 для зменшення розмірності перед більш ресурсоемними операціями, була розроблена як більш обчислювально ефективна альтернатива VGGNet, досягаючи порівнянної точності при меншій кількості параметрів.

Усі три архітектури продемонстрували високу точність у задачах розпізнавання образів[11] і стали надзвичайно популярними для застосування в режимі перенесення навчання (transfer learning). Доступність якісних попередньо навчених на великих датасетах моделей VGGNet та ResNet значно спрощує їх застосування для специфічних завдань, таких як розпізнавання емоцій. Це дозволяє скоротити час розробки та зменшити вимоги до обсягу навчальних даних, що є особливо актуальним для даного дослідження, де збір великої кількості специфічних медичних даних є ускладненим. Зазвичай ResNet пропонує одну з найвищих точностей "з коробки" завдяки своїй здатності до навчання дуже складних представлень ознак, тоді як VGGNet залишається надійною та стабільною базовою моделлю для порівняльних експериментів.

Таким чином, вибір між VGGNet, ResNet та Inception залежить від конкретних пріоритетів дослідження: балансу між точністю, доступними обчислювальними ресурсами та складністю реалізації. VGGNet є чудовим

вибором завдяки своїй архітектурній простоті, стабільності та доступності якісних попередньо навчених моделей. Inception пропонує кращий компроміс між точністю та обчислювальною ефективністю. ResNet, у свою чергу, є найкращим кандидатом для досягнення максимальної точності при роботі зі складними даними, де здатність до навчання глибоких ієрархій ознак є вирішальною.

Окрім класичних архітектур CNN, що аналізують кожне зображення як незалежну одиницю, для повноти порівняння варто розглянути і гібридний підхід. Він поєднує згорткові та рекурентні нейронні мережі (RNN), зокрема, моделі з довгою короткостроковою пам'яттю (LSTM). Такий підхід, відомий як CNN-LSTM, спрямований на вирішення одного з ключових недоліків стандартних CNN – їхньої нездатності аналізувати часову динаміку.

Суть цього методу полягає у двох-етапній обробці даних. На першому етапі згорткова нейронна мережа, наприклад VGG або ResNet, виступає в ролі екстрактора просторових ознак. Вона обробляє кожен окремий кадр відеопотоку і перетворює його на компактний векторний опис, що містить інформацію про ключові риси обличчя на цьому кадрі. На другому етапі послідовність цих векторів, отриманих з кількох кадрів, подається на вхід шару LSTM. LSTM, маючи внутрішню пам'ять, аналізує, як саме змінювалися риси обличчя в часі, що дозволяє їй вловлювати динаміку розвитку емоції, а не просто фіксувати статичний вираз.

Теоретично, такий метод має значні переваги. Аналізуючи динаміку, а не окремий кадр, модель може краще розуміти контекст, наприклад, відрізнити посмішку, що тільки з'являється, від широкої посмішки, або здивування, що переходить у страх. Окрім того, розгляд послідовності робить модель більш стійкою до випадкових шумів або невдалих кадрів, оскільки вона приймає рішення на основі кількох попередніх станів.

Для даної роботи, як зазначено в рефераті, розглядається використання VGGNet, ResNet та LSTM, що дозволить на практиці порівняти ефективність стабільної, класичної архітектури з більш сучасною та глибокою моделлю для вирішення задачі розпізнавання емоцій в умовах реабілітації.

### 3 ТЕОРЕТИЧНЕ ДОСЛІДЖЕННЯ

#### 3.1 Архітектура та проектування ПЗ

Розробка програмного забезпечення для автоматичного розпізнавання емоцій пацієнтів під час фізичних навантажень в реабілітаційних програмах передбачає створення гнучкої та ефективної архітектури, яка забезпечить високу точність і адаптивність системи до реальних умов. Архітектура системи повинна бути розроблена з урахуванням специфіки задачі – аналізу виразів обличчя пацієнтів та коригування інтенсивності фізичних навантажень [12, 13] у реальному часі.

Архітектура системи складається з кількох основних компонентів. Модуль збору та обробки даних відповідає за збір відео-потоків з камер, встановлених під час реабілітаційних вправ, та його попередню обробку[14]. Для цього в реалізації на мові Python використовуються бібліотеки комп'ютерного зору OpenCV (cv2) та dlib для визначення обличчя. Попередня обробка, що включає вирівнювання освітлення, зменшення шумів та нормалізацію зображень, також виконується засобами OpenCV та NumPy.

Модуль розпізнавання емоцій використовує нейронні мережі для аналізу міміки обличчя. Реалізація цього модуля виконана з використанням фреймворку TensorFlow та його високорізневого API Keras.

В рамках дослідження було реалізовано та порівняно моделі на основі згорткових архітектур ResNet (див. рис. 3.1.1) та VGGNet (див.рис.3.1.2), а також гібридної архітектури CNN-LSTM (див. рис. 3.1.3) для аналізу послідовностей. Для підвищення точності та ефективності системи застосовується метод трансферного навчання, що полягає у використанні моделей, попередньо навчених на великому наборі даних ImageNet, та їх подальшому донавчанні на спеціалізованому датасеті FER-2013.

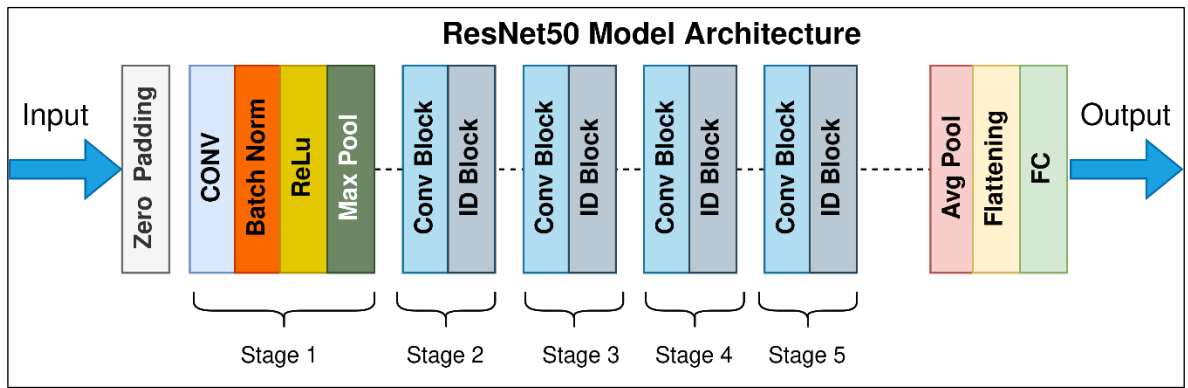


Рисунок 3.1.1 – Архітектура ResNet (за даними [15])

Для підвищення точності та ефективності системи можна застосовувати transfer learning з використанням попередньо навчених моделей на великих наборах даних, таких як FER-2013 або AffectNet.

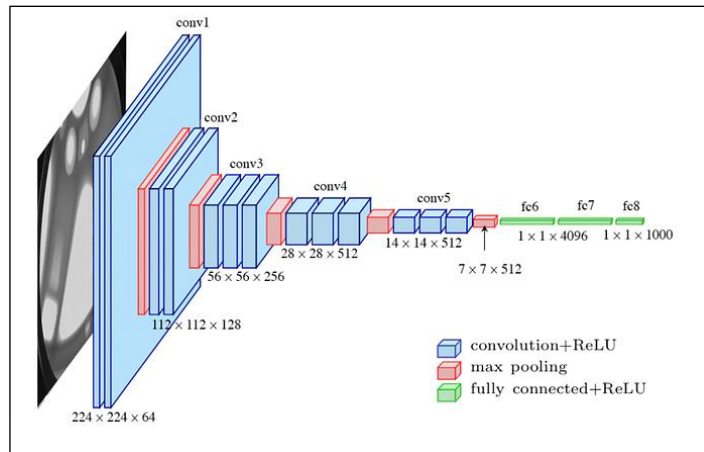


Рисунок 3.1.2 – Архітектура VGGNet (за даними [16])

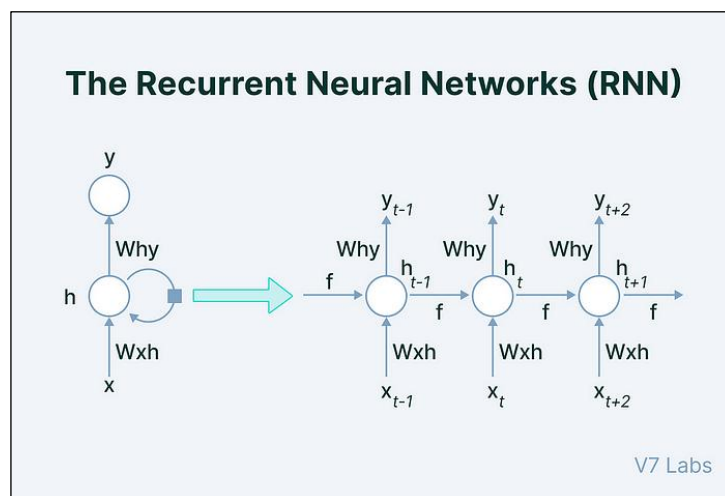


Рисунок 3.1.3 – Архітектура RNN (за даними [17])

Модуль адаптації навантажень обробляє отримані дані про емоції. Після того як емоція пацієнта виявлена, система передає результат до логічного блоку коригування фізичних навантажень. Цей модуль приймає рішення про зміну інтенсивності вправ в залежності від емоційного стану пацієнта (наприклад, якщо пацієнт відчуває біль чи стрес, навантаження повинно бути знижене). Для прототипу, реалізованого в рамках даної роботи, візуалізація стану системи та результатів розпізнавання здійснюється безпосередньо у вікні відеопотоку за допомогою функцій бібліотеки OpenCV.

Модуль зберігання даних організовує структуру зберігання інформації. Для повноцінної системи передбачається зберігання результатів аналізу та даних пацієнтів у базі даних, де фіксується історія розпізнаних емоцій та коригувань навантажень для кожного пацієнта. Для прототипу, розробленого в рамках дослідження, результати експериментів та сесій можуть зберігатися для подальшого аналізу.

### 3.2 Обмеження у процесі дослідження

Розробка програмної системи буде включати різні обмеження, які слід взяти до уваги.

Перш за все, наявне обмеження якості даних. Для тренування моделі необхідні великі та якісні набори даних із зображеннями та відео пацієнтів під час фізичних вправ. Це може ускладнювати процес, але використання готових датасетів спростить обробку і підготовку даних.

Також необхідно взяти до уваги залежність від зовнішніх умов. Точність розпізнавання емоцій залежить від умов, таких як освітлення чи кут зйомки. Для зменшення впливу цих факторів можна впровадити алгоритми попередньої обробки зображень за допомогою бібліотек.

Різноманітність емоційних реакцій. Фізіологічні реакції пацієнтів на навантаження можуть варіюватися, що потребує персоналізованого налаштування моделі. Це можна вирішити за допомогою спеціальних конфігурацій в додатку.

### 3.3 Необхідні ресурси для виконання дослідження

Серед необхідних ресурсів можна виділити технічні ресурси, дані, ресурси для тестування. Далі кожен з них розглядається більш детально.

До технічних ресурсів можна віднести:

- комп'ютер із потужним процесором і GPU для тренування моделей;
- використання TensorFlow та API Keras для створення моделей;
- інтеграція бібліотек для роботи з комп'ютерним зором та обробки зображень.

Розглядаючи дані, очікується наступні варіанти:

- набори даних із відео та зображеннями пацієнтів під час фізичних вправ, позначеними відповідно до емоцій;
- доступ до загальнодоступних наборів даних, таких як FER-2013 або AffectNet, для попереднього навчання моделей.

Час і ресурси для тестування:

- збір і обробка тестових даних за допомогою scikit-learn;
- аналіз результатів, коригування моделі за допомогою TensorFlow та Keras, з подальшою інтеграцією оновленого файлу моделі в програму для розпізнавання.

Розробка такого підходу дозволить створити ефективну систему, що інтегрується в процес реабілітації та враховує емоційний стан пацієнтів, що забезпечить підвищення ефективності лікування та поліпшить якість реабілітаційних процесів.

### 3.4 Проектування системи

Система розробляється із застосуванням модульного підходу, що дозволяє кожному компоненту працювати автономно та ефективно взаємодіяти з іншими модулями. Важливими аспектами проектування є гнучкість та масштабованість системи, а також можливість її інтеграції з іншими медичними системами для збору даних про пацієнтів.

Для проектування модулів систему та їх функціоналу наводиться use-case діаграму. Серед основних акторів наявні пацієнт - основний користувач системи, чий емоційний стан аналізується під час фізичних вправ; система розпізнавання емоцій - програмний компонент, що обробляє дані з камер, аналізує емоції та передає результати для адаптації фізичних навантажень; медичний працівник – використовує результати роботи системи для моніторингу та коригування реабілітаційної програми.

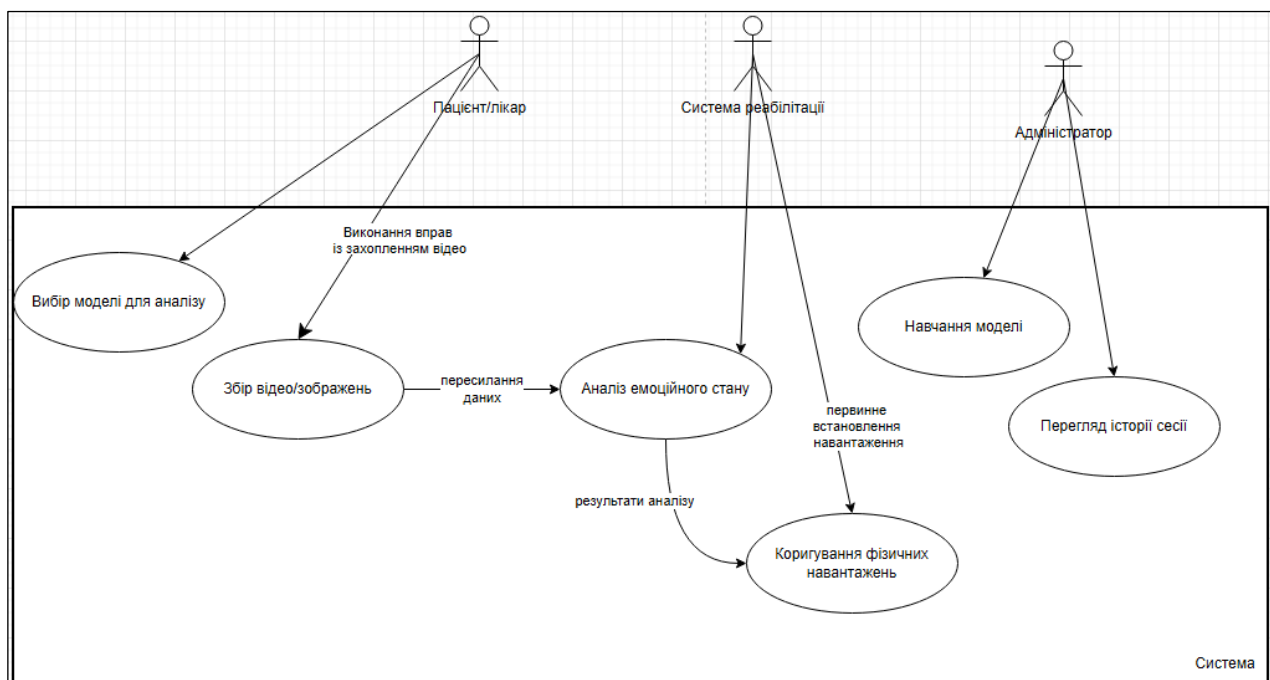


Рисунок 3.4.1 – Use-case діаграма системи (рисунок виконано самостійно)

Система повинна бути інтегрована з іншими медичними системами для збору даних про пацієнтів та історію їхнього лікування. Для цього можна застосувати RESTful API для обміну даними з іншими програмними комплексами, що використовуються в медичних установах. Інтеграція з такими системами дозволить використовувати дані пацієнтів для персоналізації реабілітаційних програм.

### 3.5 Проектування структури зберігання даних

Структура зберігання даних є важливою складовою частиною системи автоматичного розпізнавання емоцій пацієнтів під час фізичних навантажень у

реабілітаційних програмах. Правильне проектування бази даних забезпечить ефективне зберігання, обробку та доступ до інформації, що необхідна для аналізу емоційних виразів пацієнтів та коригування навантажень. Враховуючи специфіку задачі, структура бази даних повинна бути оптимізована для роботи з великими обсягами зображень та відео, а також підтримувати інтеграцію з іншими медичними системами.

Для зберігання даних про пацієнтів, їх емоційні вирази та історію фізичних навантажень, буде використано реляційну базу даних. Реляційна база даних є оптимальним вибором, оскільки дозволяє ефективно працювати з таблицями, які мають чітко визначену структуру, та надає можливість створення складних запитів для аналізу даних.

Проектування бази даних передбачає визначення основних сутностей та їх взаємозв'язків [18]. Основні сутності, які будуть використані у базі даних, включають: пацієнтів, відео та зображення, емоційні вирази, корегування навантаження (див. рис. 2).

Пацієнт (Patient) – сутність, яка містить інформацію про пацієнта. Це дозволяє зберігати базову інформацію про пацієнта для подальшого аналізу емоційних виразів та коригування фізичних навантажень. PatientId (PK) – Унікальний ідентифікатор пацієнта. FirstName – Ім'я пацієнта. LastName – Прізвище пацієнта. Age – Вік пацієнта. MedicalHistory – Історія хвороби пацієнта. Gender – Стать пацієнта. DateOfRegistration – Дата реєстрації пацієнта.

Відео/Зображення (Video/Images) – сутність, яка зберігає інформацію про відео або зображення, що використовуються для аналізу емоційних виразів пацієнта. MediaId (PK) – Унікальний ідентифікатор медіа. PatientId (FK) – Ідентифікатор пацієнта (зовнішній ключ). MediaType – Тип медіа (відео або зображення). MediaData – Дані медіа (може бути шлях до файлу або збережені в бінарному вигляді). CapturedAt – Дата та час запису відео або зображення.

Емоційні вирази (Emotions) – сутність, що містить інформацію про виявлені емоції пацієнта в конкретний момент часу. EmotionId (PK) – Унікальний ідентифікатор емоції. MediaId (FK) – Ідентифікатор медіа, яке було

проаналізовано. EmotionType – Тип емоції (радість, сум, страх, гнів, здивування тощо). ConfidenceLevel – Рівень впевненості в розпізнаній емоції (відсоток точності). Timestamp – Час, коли емоція була зафіксована.

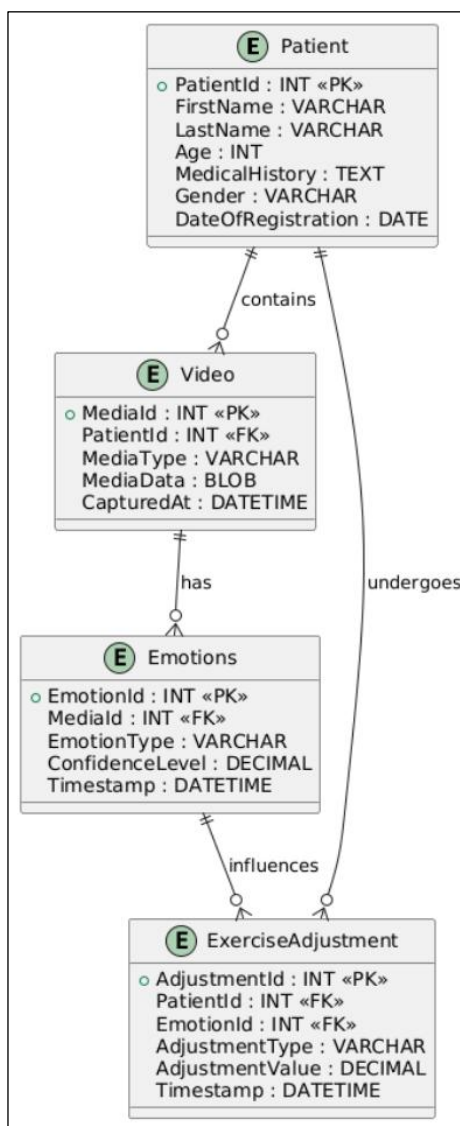


Рисунок 3.5.1 – ER-діаграма бази даних (рисунок виконано самостійно)

Коригування навантаження (ExerciseAdjustment) – сутність, що описує коригування фізичних навантажень на основі емоційного стану пацієнта. AdjustmentId (PK) – Унікальний ідентифікатор коригування. PatientId (FK) – Ідентифікатор пацієнта (зовнішній ключ). EmotionId (FK) – Ідентифікатор емоції, на основі якої здійснено коригування. AdjustmentType – Тип коригування (збільшення, зменшення навантаження). AdjustmentValue – Значення коригування

(наприклад, зміна інтенсивності вправ). Timestamp – Час, коли було застосоване коригування.

Ці сутності будуть мати наступні зв'язки між собою: один пацієнт може мати багато відео та зображень, що належать до нього; одне відео або зображення може мати кілька розпізнаних емоцій, оскільки під час аналізу можуть бути виявлені кілька емоційних виразів; одне коригування навантаження зв'язується з конкретним пацієнтом і конкретною емоцією, на основі якої було здійснено коригування фізичних вправ.

Для забезпечення ефективного масштабування системи та зберігання великих обсягів даних (зображення та відео) можуть використовуватись розподілені бази даних або хмарні рішення, такі як Azure Blob Storage для зберігання медіа-даних, в той час як реляційна база даних використовується для зберігання метаданих.

Для захисту даних пацієнтів буде застосовано шифрування при зберіганні конфіденційної інформації, а також використання безпечних з'єднань для передачі даних. Доступ до бази даних буде обмежений за допомогою ролей та дозволів для користувачів, що працюють з системою.

### 3.6 UI/UX Дизайн Системи

Дизайн користувацького інтерфейсу (UI) для системи спрямований на забезпечення зручності та ефективності використання як для лікарів, так і для пацієнтів. Особлива увага приділяється простоті візуалізації даних і легкості доступу до основних функцій.

Система буде мати два інтерфейси користувача в залежності від ролі – інтерфейс лікаря та інтерфейс пацієнта. Інтерфейс лікаря включає модулі для перегляду результатів розпізнавання емоцій пацієнтів і можливості коригування фізичних навантажень у реальному часі. Він передбачає:

- інтуїтивну панель управління для відображення ключових показників, таких як поточний емоційний стан, рівень фізичних навантажень та історія змін;

- інтерактивні графіки, які дозволяють аналізувати динаміку емоцій пацієнта протягом тренувань;
- кнопки та повзунки для швидкого налаштування інтенсивності вправ;
- другий інтерфейс буде призначений для пацієнта, який буде спрощений і зрозумілий інтерфейс, який інформує про їхній поточний стан та прогрес у реабілітації.

Основні особливості:

- мінімалістичний дизайн із фокусом на відображенні позитивних досягнень;
- інтерактивні підказки для забезпечення зворотного зв'язку під час виконання вправ.

Загалом користувацький досвід зосереджується на зручності та швидкості роботи із системою. UX-особливості включають інтерактивні елементи для адаптації навантажень у реальному часі, такі як перетягування повзунків або вибір готових профілів тренувань; підтримку різних мов для полегшення використання системи в різних регіонах; адаптивність інтерфейсу для роботи на різних пристроях: настільних комп'ютерах, планшетах та смартфонах.

### 3.7 Приклади алгоритмів та методів

У процесі розробки системи для розпізнавання емоцій пацієнтів під час реабілітаційних вправ важливим етапом є вибір і застосування ефективних алгоритмів та методів машинного навчання й обробки зображень. У цьому розділі розглянемо приклади алгоритмів, які можуть бути використані для розпізнавання емоцій, а також їх адаптацію для задачі.

Перед застосуванням методів розпізнавання емоцій необхідно виконати попередню обробку даних, яка включає декілька етапів.

Обробка зображень та відео. Це включає градування сірого кольору – перетворення зображення у відтінки сірого для зменшення обчислювальної складності; нормалізацію яскравості – вирівнювання освітлення для зменшення

впливу зовнішніх факторів та зменшення шуму – використання фільтрів для видалення шуму.

Наступним алгоритмом для обробки буде виділення обличчя, щоб підготувати його для безпосередньо розпізнавальних алгоритмів. Для цього може використовуватися Haar Cascade Classifiers або MTCNN (Multi-Task Cascaded Convolutional Networks) – алгоритми для швидкого виділення області обличчя на зображенні.

Після виділення обличчя необхідно нормалізувати розміри обличчя за допомогою методів вирівнювання обличчя (face alignment), таких як аналіз ключових точок (landmarks).

Головною задачею реалізації роботи буде безпосередньо сам алгоритм розпізнавання емоцій у підготовленому зображенні.

Для підходу глибинного навчання буде використовуватися:

- автоматичне вилучення ознак – згорткові шари (CNN) автоматично навчаються на сирих пікселях зображення і самостійно вирішують, які ознаки є найважливішими для розпізнавання емоцій.
- інтегрована класифікація відбувається в тих самих повнозв'язних шарах, які є частиною цієї ж великої нейронної мережі.

Методи на основі глибокого навчання:

- CNN (Convolutional Neural Networks) – VGGFace або його модифікації – нейронна мережа, навчена для аналізу облич, може бути адаптована для розпізнавання емоцій;
- ResNet – використовується для розв'язання проблеми деградації глибоких нейронних мереж;
- Recurrent Neural Networks (RNN) – використовуються для аналізу послідовностей кадрів з відео. Зокрема, LSTM (Long Short-Term Memory) дозволяє виявляти часові залежності у зміні емоцій;
- Transfer Learning – використання попередньо навченої моделі, такої як EfficientNet або MobileNet, з подальшим перенавчанням на наборі даних емоцій пацієнтів.

Далі розглядаються методи аналізу емоцій – розпізнавання емоцій виконується за допомогою класифікаторів, які використовують як вхідні дані вектор ознак. Приклади ознак: геометричні ознаки (відстань між ключовими точками обличчя, кути між лініями, що проходять через ключові точки), текстурні ознаки (використання методів, таких як LBP (Local Binary Patterns) або HOG (Histogram of Oriented Gradients), для аналізу текстури обличчя), комбінація ознак (поєднання геометричних і текстурних характеристик для підвищення точності).

Для визначення емоційного стану використовуються багатокласові класифікатори, які розділяють емоції на такі категорії: радість, сум, страх, гнів, здивування тощо. У результаті модель видає тип емоції (клас), ймовірність (рівень впевненості) для кожного класу.

Також для підвищення продуктивності та точності системи використовуються такі методи оптимізації:

- аугментація даних – додавання до набору даних нових зображень шляхом поворотів, масштабування, зміни яскравості тощо;
- регуляризація – використання Dropout або L2-регуляризації для запобігання перенавчанню;
- гіперпараметрична оптимізація – використання методів, таких як Grid Search або Bayesian Optimization, для налаштування параметрів моделі.

### 3.8 Реалізація алгоритму розпізнавання емоцій

Загалом, надані програмні компоненти спрямовані на створення та підготовку інтелектуальної системи, здатної розпізнавати людські емоції за виразом обличчя. Ця задача є центральною для дослідження, яке має на меті розробити методи для аналізу емоційного стану пацієнтів під час реабілітаційних процедур.

Перший компонент вирішує завдання навчання "мозку" системи – тобто моделі, яка буде безпосередньо аналізувати зображення та визначати емоції. Цей процес можна розділити на кілька ключових етапів. Перш за все необхідно зібрати та попередня обробити навчального матеріалу. Система починає з

підготовки великого набору фотографій облич, де кожне зображення вже має позначку відповідної емоції (наприклад, "радість", "сум", "здивування").

Щоб модель краще навчалася та не залежала від дрібних деталей, ці зображення проходять етап "очищення" та стандартизації: їх розмір уніфікується, а значення яскравості пікселів приводяться до єдиного діапазону. Щоб навчити модель розпізнавати емоції в різноманітних умовах, вихідний набір зображень штучно розширюється. Наприклад, існуючі фотографії трохи повертають, злегка зсувають, масштабують або дзеркально відображають. Це допомагає моделі стати більш стійкою до незначних змін у реальних зображеннях.

Далі створюється сама модель – складна структура, здатна виявляти складні закономірності в даних. Для прискорення процесу та підвищення якості часто використовують вже перевірені архітектури таких моделей, які довели свою ефективність у схожих задачах аналізу зображень. Модель навчають підготовленими зображеннями, і вона поступово вчиться співвідносити візуальні ознаки на обличчі з відповідними емоціями. Процес навчання контролюється, щоб уникнути перенавчання, коли модель ідеально розпізнає лише ті зображення, які бачила раніше, але погано справляється з новими. Зберігається та версія моделі, яка показала найкращі результати на незалежному наборі тестових зображень під час кожної епохи.

Наступним кроком йде оцінка та перевірка знань. Після навчання якість роботи моделі ретельно перевіряється на зображеннях, які вона раніше не бачила. Оцінюється, наскільки точно вона може класифікувати емоції, і виявляються її сильні та слабкі сторони. Навчена модель готова до використання: їй можна подати нове зображення обличчя, і вона видає прогноз щодо емоції, яку виражає людина на цьому фото, разом із рівнем впевненості у своєму прогнозі.

Другий програмний компонент призначений для практичного застосування розпізнавання емоцій, наприклад, для аналізу виразів обличчя людини через веб-камеру в реальному часі. Імплементация буде включати такі кроки: захоплення відео потоку; попереднє виявлення обличчя на кадрі; аналіз ключових рис обличчя; інтерпретація емоцій. У цьому компоненті може використовуватися або

спрощений набір правил для визначення емоцій, або, що більш ефективно, інтегрована модель, навчена за допомогою першого компонента. Наступні кроки включають стабілізацію результату, бо вираз обличчя може швидко змінюватися або система може робити миттєві помилкові розпізнавання, результат "згладжується" – тобто враховується послідовність розпізнаних емоцій за короткий проміжок часу, щоб отримати більш стабільну оцінку та відображення інформації. Нижче наведений приклад реалізації для визначення характеристик обличчя, які будуть використовуватися у визначення різних емоцій:

```

eye_dist_x = landmarks.part(45).x - landmarks.part(36).x
    if eye_dist_x <= 1:
        eye_dist_x = 1

    mouth_opening_y = (landmarks.part(66).y -
landmarks.part(62).y)
        norm_mouth_opening = mouth_opening_y / eye_dist_x

        mouth_corners_y_avg = (landmarks.part(48).y +
landmarks.part(54).y) / 2
            inner_lips_y_avg = (landmarks.part(62).y +
landmarks.part(66).y) / 2
                smile_indicator = (inner_lips_y_avg - mouth_corners_y_avg) /
eye_dist_x

                eyebrow_left_y_avg = (landmarks.part(19).y +
landmarks.part(20).y) / 2
                    eyebrow_right_y_avg = (landmarks.part(23).y +
landmarks.part(24).y) / 2
                        eye_top_left_y_avg = (landmarks.part(37).y +
landmarks.part(38).y) / 2
                            eye_top_right_y_avg = (landmarks.part(43).y +
landmarks.part(44).y) / 2

                                brow_raise_left = (eye_top_left_y_avg - eyebrow_left_y_avg) /
eye_dist_x
                                    brow_raise_right = (eye_top_right_y_avg -
eyebrow_right_y_avg) / eye_dist_x
                                        avg_brow_raise = (brow_raise_left + brow_raise_right) / 2

                                            inner_brow_dist_x = landmarks.part(22).x -
landmarks.part(21).x
                                                norm_inner_brow_dist = inner_brow_dist_x / eye_dist_x

```

Основний користувацький інтерфейс буде зроблений прямо на відео елементі. Користувач бачить на екрані відео з камери, де обличчя може бути

виділено, а поряд відображається розпізнана емоція. Також може показуватися технічна інформація, наприклад, швидкість обробки кадрів.

Ці два компоненти разом створюють основу для системи, яка може бути використана в рамках дослідження для автоматичного моніторингу емоційного стану пацієнтів під час реабілітації, допомагаючи адаптувати програму відновлення та покращувати її ефективність.

Також буде окремо створений компонент для визначення пристосування до навантаження пацієнта. В залежності від того, які емоції будуть розпізнані, програма буде автоматично адаптувати складність тренувань. Конкретно для дослідження буде реалізований абстрактний параметр складності, який відображає цю умовний рівень вправ та фізичного навантаження.

Алгоритм для визначення адаптації буде базуватися на вибірці результатів розпізнавання емоцій за останній проміжок часу. Умовно позитивні та нейтральні емоції будуть сприяти збільшенню навантаженню, негативні – зменшенню. В залежності від відношення цих результатів, то висновок може бути незмінним, щоб не перевантажувати просто так пацієнти. Приклад реалізації цієї логіки наведений нижче:

```

if cooldown_timer == 0 and len(emotion_history) ==
DECISION_BUFFER_SIZE:
    negative_count = sum(1 for emo in emotion_history if emo in
NEGATIVE_STRAIN_EMOTIONS)
    positive_count = sum(1 for emo in emotion_history if emo in
POSITIVE_NEUTRAL_EMOTIONS)

    negative_ratio = negative_count / DECISION_BUFFER_SIZE
    positive_ratio = positive_count / DECISION_BUFFER_SIZE

    print(f"DEBUG: neg_ratio={negative_ratio:.2f} (Threshold:
{NEGATIVE_THRESHOLD}) | "
          f"pos_ratio={positive_ratio:.2f} (Threshold:
{POSITIVE_THRESHOLD})")

    if negative_ratio >= NEGATIVE_THRESHOLD:
        if current_level > 1:
            current_level -= 1
            last_decision = "DECREASE"
            cooldown_timer = COOLDOWN_FRAMES
            print(f"Decision: DECREASE Load to Level
{current_level}")

```

```
elif positive_ratio >= POSITIVE_THRESHOLD and negative_ratio
== 0:
    if current_level < 5:
        current_level += 1
        last_decision = "INCREASE"
        cooldown_timer = COOLDOWN_FRAMES
        print(f"Decision: INCREASE Load to Level
{current_level}")
    else:
        last_decision = "Maintain"
```

На додачу до проміжку часу, який береться до уваги під час розрахунків, буде введений також проміжок часу, під час якого не буде проводитися розрахунки. Це зроблено з метою, щоб дати час адаптуватися пацієнту до нового рівня фізичного навантаження та не враховувати первинний стрес. Цей час також буде виводитися користувачу.

## 4 ПРАКТИЧНЕ ДОСЛІДЖЕННЯ

Для розпізнавання емоцій виконувався покадровий аналіз зображень для розпізнавання емоцій. Визначення ключових точок та аналіз їх взаємного розташування використовується для первинного аналізу, щоб визначити емоції. Для реалізації цього використовувалася бібліотека OpenCV та DLib. Проміжний етап цього алгоритму включає розпізнавання ключових точок та визначення меж зображення (див. рис. 4.1 та 4.2).

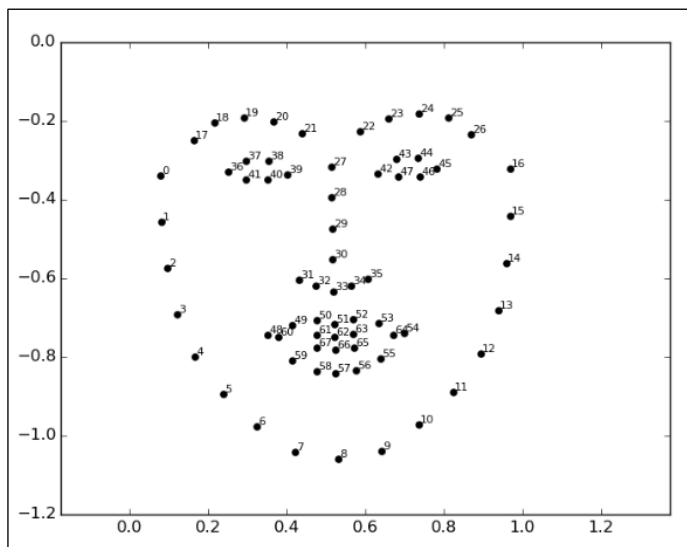


Рисунок 4.1 – Нормалізовані ключові точки з OpenCV та DLib (за даними [19])

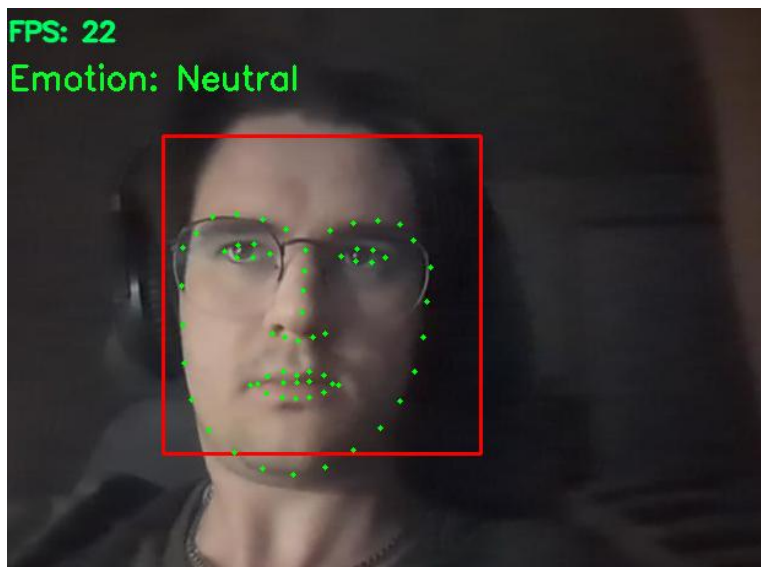


Рисунок 4.2 – Реалізація за допомогою OpenCV та DLib (рисунок виконано самостійно)

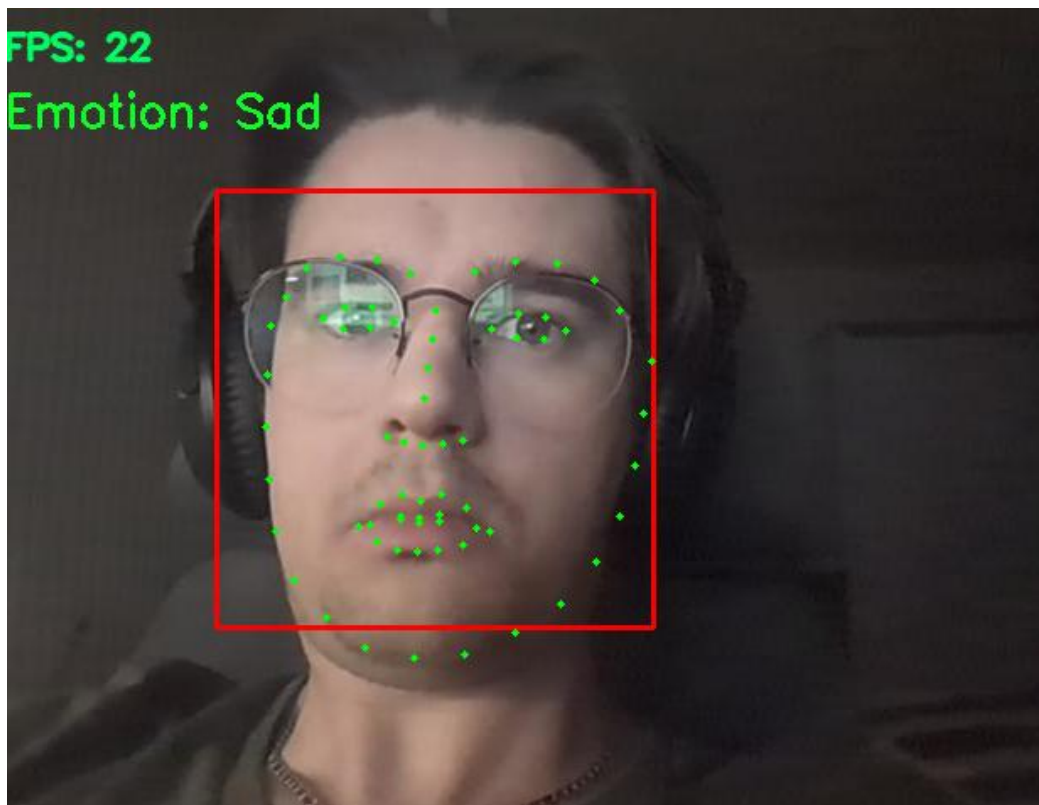


Рисунок 4.3 – Реалізація за допомогою OpenCV та DLib (рисунок виконано самостійно)

У результаті виконання програмної реалізації була розроблена програма система, що дозволяє підключатися до камери та розпізнавати емоції (див. рис. 4.4 та 4.5). Під час роботи програми, також здійснюється корекція фізичних навантажень (див. рис. 4.6).

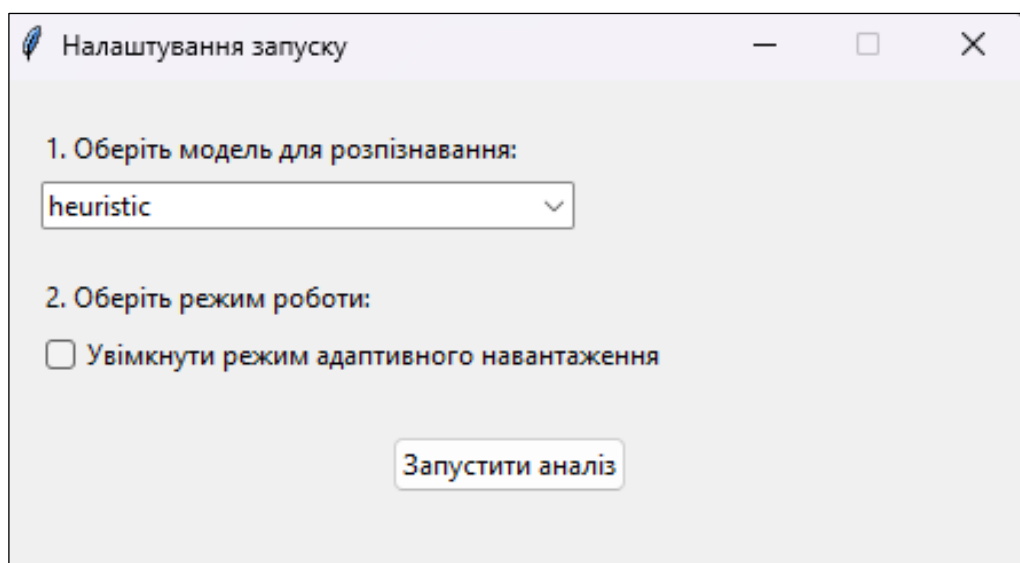


Рисунок 4.4 – Початковий інтерфейс користувача (рисунок виконано самостійно)

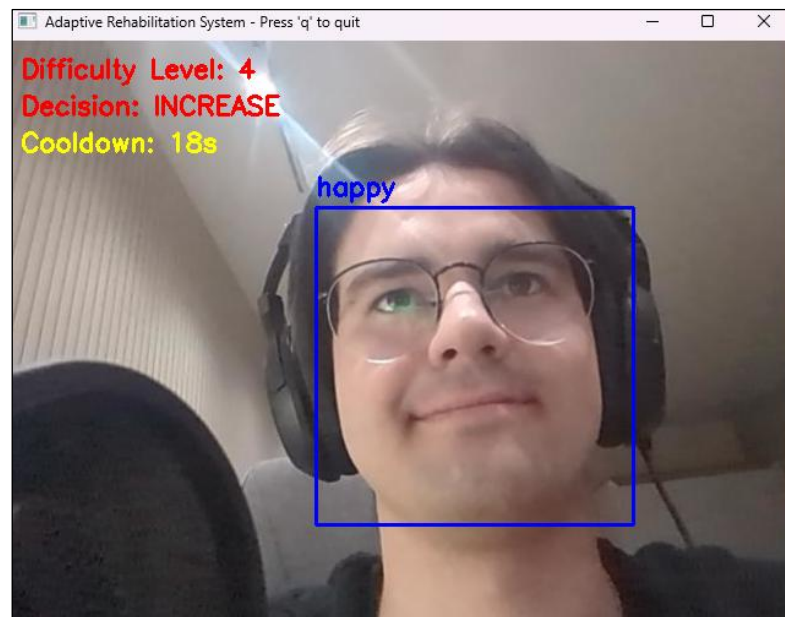


Рисунок 4.5 – Демонстрація роботи розпізнавання емоцій (рисунок виконано самостійно)

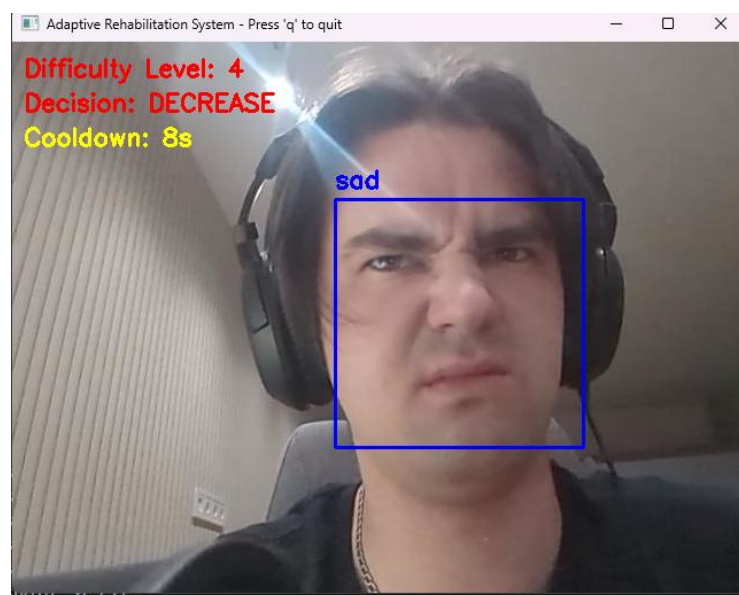


Рисунок 4.6 – Демонстрація зміни навантаження (рисунок виконано самостійно)

Для подальшого практичного дослідження першим ділом проведемо навчання моделей на тестових даних з датасету FER-2013. Кількість епох, яка використовувалася для навчання – 50, з попереднім закінченням навчання, за умови відсутності покращення результатів протягом декількох епох. Результати та параметри навчання можна побачити у консолі після кожної епохи. (див. рис. 4.7).

dropout (Dropout)	(None, 1024)	0
dense_1 (Dense)	(None, 7)	7,175

```

Total params: 15,247,175 (58.16 MB)
Trainable params: 532,487 (2.03 MB)
Non-trainable params: 14,714,688 (56.13 MB)
None
Epoch 1/50
C:\Users\andre\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\LocalCache\local-pac
kages\Python312\site-packages\keras\src\trainers\data_adapters\py_dataset_adapter.py:121: UserWarning: Your `
PyDataset` class should call `super().__init__(**kwargs)` in its constructor. `**kwargs` can include `workers
`, `use_multiprocessing`, `max_queue_size`. Do not pass these arguments to `fit()`, as they will be ignored.
  self._warn_if_super_not_called()
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0s 437ms/step - accuracy: 0.2964 - loss: 1.7567
Epoch 1: val_accuracy improved from -inf to 0.36356, saving model to ../Models/emotion_resnet50_model.keras
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 287s 633ms/step - accuracy: 0.2965 - loss: 1.7566 - val_accuracy: 0.3636 - val_l
oss: 1.6073
Epoch 2/50
 1/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 3:46 508ms/step - accuracy: 0.2656 - loss: 1.75562025-06-19 01:18:09.890319: W t
ensorflow/core/framework/local_rendezvous.cc:404] Local rendezvous is aborting with status: OUT_OF_RANGE: End
of sequence
[[{{node IteratorGetNext}}]]
C:\Program Files\WindowsApps\PythonSoftwareFoundation.Python.3.12_3.12.2800.0_x64-qbz5n2kfra8p0\lib\contextl

```

Рисунок 4.7 – Навчання моделі (рисунок виконано самостійно)

Наступним кроком після навчання необхідно провести багатокритеріальний аналіз за різними параметрами та характеристиками моделей. Для цього використаємо метод вагових коефіцієнтів, щоб представити наявні критерії у вигляді кількісних та вимірюваних оцінок.

```

Epoch 4: val_accuracy did not improve from 0.40000
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0s 113us/step - accuracy: 0.3438 - loss: 1.5585 - val_accuracy: 0.4000 - val_loss: 1.4984
Epoch 5/50
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0s 101ms/step - accuracy: 0.3509 - loss: 1.6308
Epoch 5: val_accuracy did not improve from 0.40000
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56s 125ms/step - accuracy: 0.3509 - loss: 1.6308 - val_accuracy: 0.3915 - val_loss: 1.5598
Epoch 6/50
 1/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43s 98ms/step - accuracy: 0.3906 - loss: 1.48892025-06-19 01:21:44.985910: W tensorflow/core/fr
amework/local_rendezvous.cc:404] Local rendezvous is aborting with status: OUT_OF_RANGE: End of sequence
[[{{node IteratorGetNext}}]]
2025-06-19 01:21:45.014678: W tensorflow/core/framework/local_rendezvous.cc:404] Local rendezvous is aborting with status: O
UT_OF_RANGE: End of sequence
[[{{node IteratorGetNext}}]]

Epoch 6: val_accuracy improved from 0.40000 to 0.70000, saving model to ../Models/emotion_resnet50_model.keras
Epoch 6: val_accuracy improved from 0.40000 to 0.70000, saving model to ../Models/emotion_resnet50_model.keras
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0s 520us/step - accuracy: 0.3906 - loss: 1.4889 - val_accuracy: 0.7000 - val_loss: 1.1587
Epoch 7/50
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0s 104ms/step - accuracy: 0.3622 - loss: 1.6149
Epoch 7: val_accuracy did not improve from 0.70000
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 58s 129ms/step - accuracy: 0.3622 - loss: 1.6149 - val_accuracy: 0.3982 - val_loss: 1.5524
Epoch 8/50
 1/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 45s 102ms/step - accuracy: 0.4688 - loss: 1.50812025-06-19 01:22:43.560144: W tensorflow/core/fr
amework/local_rendezvous.cc:404] Local rendezvous is aborting with status: OUT_OF_RANGE: End of sequence
[[{{node IteratorGetNext}}]]
2025-06-19 01:22:43.593256: W tensorflow/core/framework/local_rendezvous.cc:404] Local rendezvous is aborting with status: O
UT_OF_RANGE: End of sequence
[[{{node IteratorGetNext}}]]

Epoch 8: val_accuracy did not improve from 0.70000
448/448 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 0s 99us/step - accuracy: 0.4688 - loss: 1.5081 - val_accuracy: 0.7000 - val_loss: 1.2493
Epoch 9/50

```

Рисунок 4.8 – Приклад логів з показниками ефективності після епохи (рисунок виконано самостійно)

Першим кроком будується таблиця вагових коефіцієнтів для визначення найкращої моделі, включаючи різні ознаки. Спочатку розглядається найголовніший параметр – точність знайдених емоцій. Програма виводить два подібних параметри – точність на тренувальному наборі даних та на валідаційному. Для визначення ефективності розпізнавання емоцій для кожної моделі було обране друге значення, оскільки саме воно відображає те, як буде працювати модель у реальних задачах на різноманітних наборах даних. Оскільки модель навчається упродовж 50 епох, необхідно визначити найкраще значення під час найбільш показової епохи, коли модель показує найкращі показники розпізнавання емоцій після навчання. Окрім логів про показники ефективності (див. рис. 4.8), що виводить програма після кожної епохи, це можна побачити також на графіках з результатами навчання (див. рис. 4.9).

```
Restoring model weights from the end of the best epoch: 18.
113/113 ██████████ 12s 108ms/step - accuracy: 0.3346 - loss: 1.6386
Loss на валідаційному наборі: 1.5431
Ассурасу на валідаційному наборі: 40.23%
█
```

Рисунок 4.9 –Результати навчання моделі (рисунок виконано самостійно)

Як можна побачити з графіків точності та втрат, можна помітити, що ResNet та VggNet дещо краще визначають емоцію як під час тренувань, так і під час валідації на окремому датасеті. Хоча, незважаючи на нижчі показники точності, RNN показує кращі показники навчання у втратах, визначаючи неправильні емоції. На дистанції це дає кращий показник.

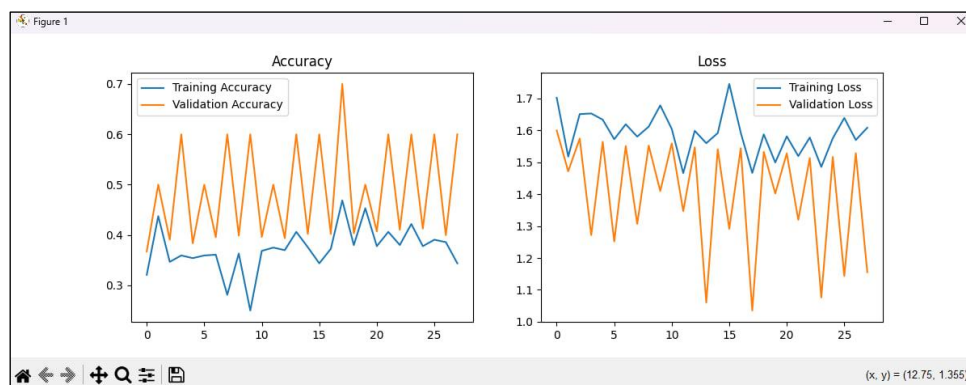


Рисунок 4.10 –Графік показників навчання ResNet (рисунок виконано самостійно)

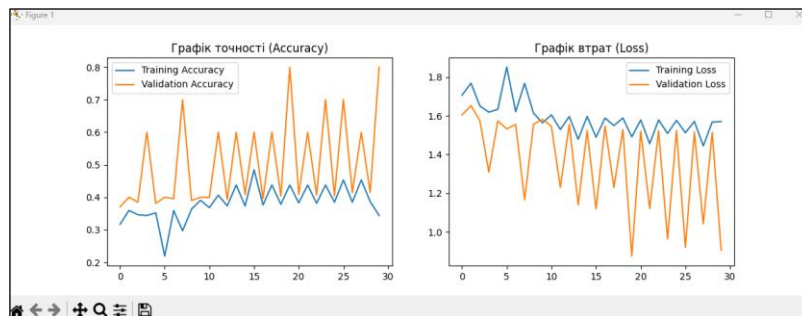


Рисунок 4.11 –Графік показників навчання VGGNet (рисунок виконано самостійно)

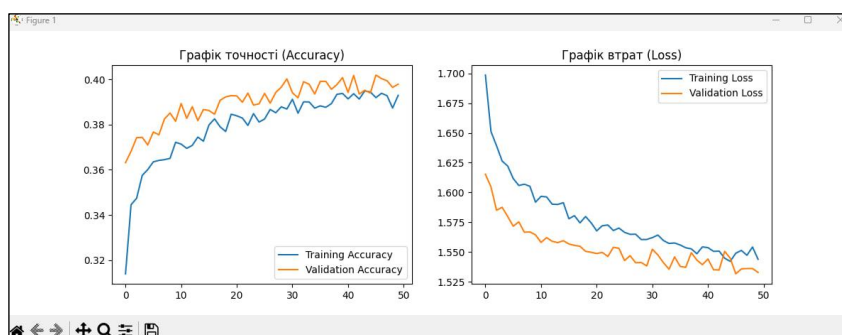


Рисунок 4.12 –Графік показників навчання RNN – LSTM (рисунок виконано самостійно)

Окрім показників точності та втрат слід проаналізувати також й інші параметри, які базуються на результатах навчання, як от стабільність роботи програми та коректність визначення різних класів емоцій. Для цього можна порівняти показник повноти – наскільки часто модель правильно визначила кожний окремий клас емоцій. Для цього виведемо дані зі звіту про класифікацію з найкращої збереженої моделі для кожної з архітектур.

```

_warn_prt(average, modifier, f {metric.capitalize()} is , len
precision    recall  f1-score   support

   angry      0.35     0.19     0.24       958
  disgust     0.00     0.00     0.00       111
    fear      0.29     0.14     0.19      1024
   happy      0.42     0.70     0.53      1774
  neutral     0.34     0.45     0.39      1233
    sad       0.40     0.23     0.29      1247
  surprise    0.51     0.57     0.54       831

 accuracy                    0.40       7178
 macro avg      0.33     0.32     0.31       7178
 weighted avg   0.38     0.40     0.37       7178

```

Рисунок 4.13 –Звіт про класифікацію ResNet (рисунок виконано самостійно)

Було зроблено подібні звіти і для VGGNet та LSTM, щоб отримати необхідні дані для порівняння обраних моделей за одними і тими ж параметрами після навчання на датасеті.

```

precision    recall  f1-score   support

   angry      0.40     0.16     0.23       958
  disgust     0.00     0.00     0.00       111
    fear      0.30     0.15     0.20      1024
   happy      0.42     0.75     0.53      1774
  neutral     0.36     0.43     0.39      1233
    sad       0.40     0.18     0.25      1247
  surprise    0.49     0.63     0.55       831

 accuracy                    0.40       7178
 macro avg      0.34     0.33     0.31       7178
 weighted avg   0.39     0.40     0.36       7178

```

Рисунок 4.14 –Звіт про класифікацію VGGNet (рисунок виконано самостійно)

```

Confusion Matrix:
[[ 49  0  56 326 143 118 266]
 [  3  0  9  20  12  1  66]
 [167  0  57 371 139 100 190]
 [158  1 169 813 269 189 175]
 [ 57  0  17 877 157  78  47]
 [ 60  0  37 490 498 100  62]
 [ 66  0  39 337 150 213  26]]

Classification Report:
              precision    recall  f1-score   support

   angry         0.09         0.05         0.06         958
   disgust        0.00         0.00         0.00         111
    fear         0.15         0.06         0.08        1024
   happy         0.25         0.46         0.32        1774
   neutral        0.11         0.13         0.12        1233
    sad          0.13         0.08         0.10        1247
   surprise       0.03         0.03         0.03         831

 accuracy                   0.17         7178
  macro avg         0.11         0.11         0.10         7178
 weighted avg       0.14         0.17         0.14         7178

```

Рисунок 4.15 –Звіт про класифікацію LSTM (рисунок виконано самостійно)

Вирахуємо середні значення показника повноти (Recall) для класу загалом, який відповідає за точність визначення окремого класу емоцій щоб отримати загальне значення по всіх класах для використаних моделей. Було взято середні значення оскільки кожна модель має визначати увесь спектр емоцій, а не якусь певну окремо, оскільки не можливо дізнатися за тип емоцій до безпосереднього визначення.

Приклад розрахунку середнього значення повноти (формула 4.1):

$$Recall_{avg} = \frac{\sum_{i=0}^n recall_i}{n} \quad (4.1)$$

де  $n$  – кількість класів емоцій;

$recall_i$  – повнота для окремого класу емоцій.

У даному випадку використовується датасет із 7 типів емоцій ( $n = 7$ ): злість, огида, страх, щастя, нейтральне, сум, здивування.

Також включимо параметр швидкодії для визначення оптимального методу з точки зору швидкості обробки кадрів.

Таблиця 4.1 – Порівняльна таблиця різних архітектур (таблиця виконана самостійно)

Критерій/Модель	CNN(VggNet)	CNN(ResNet)	LSTM
Точність	0.7	0.7	0.40190
Швидкодія (кадри/с)	14	15	12
Втрати	0.990	1.0602	1.5316
Розмір моделі	63.799MB	63.8MB	17.827MB
Повнота	0.3286	0.3257	0.1157

Після заповнення таблиці необхідно нормалізувати отримані дані та перевести у бальну шкалу отримані значення. Для цього розподілимо значення від 1 до 5, де 1 – найменше, а 5 – найбільше, для кожного критерію.

Таблиця 4.2 – Розподілення вагових коефіцієнтів за критеріями (таблиця виконана самостійно)

Критерій	Обґрунтування пріоритету	Ваговий коефіцієнт
Точність (Accuracy)	Найважливіший показник, що визначає якість класифікації.	0.35
Швидкодія (FPS)	Критично важливий параметр для роботи системи в реальному часі.	0.25
Повнота (Recall)	Показує, наскільки добре модель знаходить емоції, що важливо для виявлення негативних станів.	0.2
Втрати (Loss)	Демонструє, наскільки сильно модель помиляється.	0.1
Розмір моделі	Важливий для розгортання на пристроях з обмеженими ресурсами.	0.1
РАЗОМ		1

Логіка оцінювання наступна: для точності, швидкодії та повноти вище значення отримує вищий бал, а втрати та розмір моделі – навпаки, вище значення

отримує менший бал. Оновлена таблиця містить готові значення для розрахунків та аналізу кращої опції (див. табл. 4.3).

Таблиця 4.3 – Таблиця із ваговими коефіцієнтами (таблиця виконана самостійно)

Критерій/Модель	CNN (VggNet)	CNN (ResNet)	LSTM
Точність	5	5	1
Швидкодія (кадри/с)	4	5	1
Повнота (Recall)	5	4	1
Втрати (Loss)	5	4	1
Розмір моделі	2	1	5

Тепер необхідно розрахувати фінальний зважений бал для кожної моделі, помноживши її бальну оцінку на ваговий коефіцієнт критерію та підсумувавши результати.

Розрахунок для CNN (VggNet):  $Score_{VGG} = (5 \times 0.35) + (4 \times 0.25) + (5 \times 0.20) + (5 \times 0.10) + (2 \times 0.10) = 1.75 + 1.0 + 1.0 + 0.5 + 0.2 = 4.45$

Розрахунок для CNN (ResNet):  $Score_{ResNet} = (5 \times 0.35) + (5 \times 0.25) + (4 \times 0.20) + (4 \times 0.10) + (1 \times 0.10) = 1.75 + 1.25 + 0.8 + 0.4 + 0.1 = 4.30$

Розрахунок для RNN:  $Score_{RNN} = (1 \times 0.35) + (1 \times 0.25) + (1 \times 0.20) + (1 \times 0.10) + (5 \times 0.10) = 0.35 + 0.25 + 0.20 + 0.10 + 0.50 = 1.40$

Отже, можна зробити висновок, що архітектури CNN схожі за результатами і показали високий результат по різних критеріям. Обидва мають високу точність розпізнавання емоцій, але мають різні значення інших показників. RNN архітектура показала гірші результати майже по всім критеріям. Хоча вона й має малий розмір моделі, що зручно для розгортання на мобільних пристроях, але її

точність та ефективність є сумнівними, що робить її менш оптимальним вибором для даної задачі.

## ВИСНОВКИ

У межах даної кваліфікаційної роботи було визначено ключову мету – створення системи автоматичного розпізнавання емоцій пацієнтів під час виконання фізичних навантажень у реабілітаційних програмах та дослідження методів глибинного навчання для реалізації системи. Досягнення цієї мети є важливим для підвищення ефективності лікування, покращення фізичного і психологічного комфорту пацієнтів та оптимізації процесу реабілітації.

Аналіз сучасних підходів до розпізнавання емоцій вказав на значний потенціал згорткових нейронних мереж (CNN) у вирішенні поставлених задач. CNN здатні ефективно обробляти зображення обличчя, витягуючи важливі ознаки, що дозволяє точно класифікувати емоції, такі як радість, сум, гнів, здивування, страх тощо.

Виявлено декілька етапів реалізації проекту.

- перш за все потрібно зробити аналіз існуючих підходів: виявлення сильних і слабких сторін сучасних систем, таких як MorphCast і FaceReader, з метою врахування їхнього досвіду в розробці;
- розробка моделі, створення нейронної мережі на основі CNN та RNN, здатної обробляти дані в реальному часі, забезпечуючи високу точність і продуктивність у реабілітаційних умовах;
- інтеграція системи: поєднання моделі з реабілітаційними програмами для автоматичного регулювання фізичних навантажень відповідно до емоційного стану пацієнта;
- тестування та оптимізація: перевірка роботи моделі на реальних даних, коригування алгоритмів і покращення точності.

Разом із тим, було виявлено низку обмежень і викликів. Серед яких якість даних, бо для тренування моделі потрібні великі та якісні набори зображень пацієнтів у реабілітаційних умовах. Це ускладнює процес збору та підготовки даних. Також вплив зовнішніх факторів, наприклад, освітлення, положення обличчя та інші зовнішні умови можуть негативно впливати на точність

розпізнавання. Це вирішується шляхом попередньої обробки зображень за допомогою комп'ютерного зору. Різноманітність емоційних реакцій у пацієнтів має оброблюватися системою та надавати можливість до персоналізації.

Практичні дослідження показали, що моделі на основі CNN значно перевершили гібридну модель RNN за ключовими показниками. Такий результат для RNN пояснюється тим, що статичний характер навчальних даних не дозволив повною мірою використати головну перевагу рекурентних мереж – аналіз часових послідовностей.

Для визначення найбільш оптимальної архітектури було застосовано метод вагових коефіцієнтів, що враховував точність, швидкодію, повноту, втрати та розмір моделі. За результатами інтегральної оцінки, найбільш збалансованим та ефективним рішенням для поставленої задачі виявилася архітектура CNN (VggNet), яка незначно випередила ResNet завдяки кращим показникам повноти та втрат при прийнятній швидкодії.

Підсумовуючи, дана робота демонструє високий потенціал застосування згорткових нейронних мереж для створення інтелектуальних систем у сфері реабілітації. Проведений порівняльний аналіз та розроблений прототип підтверджують, що такі технології можуть значно покращити якість медичних послуг і сприяти впровадженню індивідуалізованих підходів до відновлення пацієнтів.

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Bang, G.-S., & Park, S.-B. (2024). Workout Classification Using a Convolutional Neural Network in Ensemble Learning. *Sensors*, 24(10), 3133. URL: <https://doi.org/10.3390/s24103133> (дата звернення 10.06.2025).
2. Diwan, Anjali & Sunil, Reshma & Mer, Parita & Mahadeva, Rajesh & Patole, Shashikant. (2024). Advancements in Emotion Classification via Facial and Body Gesture Analysis: A Survey. *Expert Systems*. 42. 10.1111/exsy.13759.
3. Deep Learning Based Emotion Recognition and Visualization of Figural Representation. URL: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.818833/full> (дата звернення 10.06.2025).
4. Perwej, Dr. Yusuf & Gupta, Dileep & Agarwal, Devendra & Vishwakarma, Opinder & Mishra, Priya & Nitya. (2024). Sensing Human Emotion using Emerging Machine Learning Techniques. *International Journal of Scientific Research in Science, Engineering and Technology*. Volume 11. Pages 80-91. 10.32628/IJSRSET24114104.
5. Білоус Н. В., Рассоха О. В., Агекян І. А., Грамм О. В. (2020). Дослідження методів для розробки програмної системи розпізнавання емоцій та визначення стану здоров'я людини. URL: [https://doi.org/10.30837/bi.2020.1\(94\).10](https://doi.org/10.30837/bi.2020.1(94).10) (дата звернення 10.06.2025).
6. Transfer Learning for Facial Expression Recognition. URL: <https://www.mdpi.com/2078-2489/16/4/320> (дата звернення 10.06.2025).
7. Nataliya, Bilous & Malko, Vladyslav & Frohme, Marcus & Nechyporenko, Alina. (2024). Comparison of CNN-Based Architectures for Detection of Different Object Classes. *AI*. 5. 2300-2320. 10.3390/ai5040113.
8. Facial Expression Recognition (FER) on FER2013. URL: <https://paperswithcode.com/sota/facial-expression-recognition-on-fer2013> (дата звернення 10.06.2025)
9. Facial Expression Recognition (FER) on AffectNet. URL: <https://paperswithcode.com/sota/facial-expression-recognition-on-affectnet> (дата звернення 10.06.2025).

10. Facial Expression Recognition (FER) on CK+. URL: <https://paperswithcode.com/sota/facial-expression-recognition-on-ck> (дата звернення 10.06.2025).

11. Face recognition using CNN and siamese network. URL: <https://www.sciencedirect.com/science/article/pii/S2665917423001368> (дата звернення 10.06.2025).

12. Nataliya, Bilous & Ahekan, Iryna & Kaluhin, V.. (2023). DETERMINATION AND COMPARISON METHODS OF BODY POSITIONS ON STREAM VIDEO. Radio Electronics, Computer Science, Control. 52. 10.15588/1607-3274-2023-2-6.

13. Nataliya, Bilous & Svidin, Oleh & Ahekan, Iryna & Malko, Vladyslav. (2024). A skeleton-based method for exercise recognition based on 3D coordinates of human joints. IAES International Journal of Artificial Intelligence (IJ-AI). 13. 1805. 10.11591/ijai.v13.i2.pp1805-1816.

14. Nataliya Bilous, Vladyslav Malko and Nazarii Moshenskyi. (2025). Search and Detection of People in the Water Using YOLO Architectures: A Comparative Analysis from YOLOv3 to YOLOv8. URL: <https://doi.org/10.15588/1607-3274-2020-3-1> (дата звернення 10.06.2025).

15. VGG-Net Architecture Explained. URL: <https://medium.com/@siddheshb008/vgg-net-architecture-explained-71179310050f> (дата звернення 10.06.2025).

16. The Annotated ResNet-50. URL: <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758/> (дата звернення 10.06.2025).

17. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks: A Deep Dive. URL: <https://medium.com/@sarthakjamdar9561/recurrent-neural-networks-rnns-and-long-short-term-memory-lstm-networks-a-deep-dive-85817329d0a6> (дата звернення 10.06.2025).

18. Beaulieu A. Learning SQL: Generate, Manipulate, and Retrieve Data 3rd Edition. Sebastopol: O'Reilly Media, 2020. 377 с.

19. Open Source Computer Vision Library – GitHub. URL: <https://github.com/opencv/opencv/tree/master> (дата звернення 10.06.2025).
20. GitHub – Emotion Recognition. GitHub. URL: [https://github.com/afkInTheBox/emotion\\_recognition](https://github.com/afkInTheBox/emotion_recognition) (дата звернення 10.06.2025).

## ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ

5. Білоус Н.В., Рассоха О.В., Агемян І.А., Грамм О.В. (2020). Дослідження методів для розробки програмної системи розпізнавання емоцій та визначення стану здоров'я людини. URL: [https://doi.org/10.30837/bi.2020.1\(94\).10](https://doi.org/10.30837/bi.2020.1(94).10) (дата звернення 10.06.2025).

7. Nataliya, Bilous & Malko, Vladyslav & Frohme, Marcus & Nechyporenko, Alina. (2024). Comparison of CNN-Based Architectures for Detection of Different Object Classes. AI. 5. 2300-2320. 10.3390/ai5040113.

12. Nataliya, Bilous & Ahegian, Iryna & Kaluhin, V.. (2023). DETERMINATION AND COMPARISON METHODS OF BODY POSITIONS ON STREAM VIDEO. Radio Electronics, Computer Science, Control. 52. 10.15588/1607-3274-2023-2-6.

13. Nataliya, Bilous & Svidin, Oleh & Ahegian, Iryna & Malko, Vladyslav. (2024). A skeleton-based method for exercise recognition based on 3D coordinates of human joints. IAES International Journal of Artificial Intelligence (IJ-AI). 13. 1805. 10.11591/ijai.v13.i2.pp1805-1816.

14. Nataliya Bilous, Vladyslav Malko and Nazarii Moshenskyi. (2025). Search and Detection of People in the Water Using YOLO Architectures: A Comparative Analysis from YOLOv3 to YOLOv8. URL: <https://doi.org/10.15588/1607-3274-2020-3-1> (дата звернення 10.06.2025).