Хмельницький національний університет
Факультет інформаційних технологій
Кафедра комп'ютерних наук

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

на тему *Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу*

Галузь знань *12 – Інформаційні технології*
Шифр і назва галузі знань

Спеціальність *122 – Комп'ютерні науки*
Шифр і назва спеціальності

Освітня програма *Комп'ютерні науки*
Назва освітньої програми

Виконав: *студент групи КН-22-1* (Група виконавця) _____ (Підпис) *Олександр ГЛАДУН* (Ім'я, ПРІЗВИЩЕ)

Керівник: *асистент каф. КН* (Науковий ступінь, посада) _____ (Підпис) *Ольга ЗАЛУЦЬКА* (Ім'я, ПРІЗВИЩЕ)

Нормоконтроль: *к.т.н., доц. каф. КН* (Науковий ступінь, посада) _____ (Підпис) *Руслан БАГРІЙ* (Ім'я, ПРІЗВИЩЕ)

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор _____ (Підпис) *Олександр БАРМАК* (Ім'я, ПРІЗВИЩЕ)
17 червня 2026 р.

Хмельницький 2026

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
Факультет *інформаційних технологій*
Кафедра *комп'ютерних наук*
Освітній ступінь *бакалавр*
Галузь знань *12 – Інформаційні технології*
Спеціальність *122 – Комп'ютерні науки*

ЗАТВЕРДЖУЮ
Завідувач кафедри *комп'ютерних наук*

(підпис)
д.т.н., професор Олександр БАРМАК
«22» січня 2026 року

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА

1. Тема кваліфікаційної роботи бакалавра: *«Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу»*

2. Завдання видано студенту *Олександру Гладуну*
(Ім'я, прізвище)

3. Керівник роботи *асистент кафедри КН Ольга Залуцька*
(посада, ім'я, прізвище)

4. Затверджено наказом університету від «20» січня 2026р. № 7

5. Дата видачі завдання студенту: «22» січня 2026р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета роботи – підвищення точності визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу. Для її реалізації слід: провести аналіз предметної області та сучасних підходів визначення психологічних та соціокультурних характеристик угрупувань за відеоданими;формалізувати задачу автоматизованого визначення психологічних та соціокультурних характеристик угрупувань за відеоданими;розробити метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу;реалізувати програмний засіб на базі запропонованого методу; провести експериментальне дослідження та оцінити ефективність розробленого методу.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	*Вибір напряму дослідження та узгодження теми кваліфікаційної роботи з керівником, складання календарного графіка виконання*	*січень 2026*	Виконано
2	*Ознайомлення з предметною областю, формулювання мети і задач дослідження, визначення об'єкта та предмета дослідження*	*лютий 2026*	Виконано
3	*Проєктування методу розв'язання задачі, опис архітектурних рішень, розроблення математичних моделей та алгоритмів.*	*березень 2026*	Виконано
4	*Обґрунтування інструментарію розробки, програмна реалізація розробленого методу, проведення експериментального тестування та оцінювання ефективності.*	*квітень 2026*	Виконано
5	*Написання тексту кваліфікаційної роботи, урахування зауважень керівника, оформлення згідно з вимогами*	*травень 2026*	Виконано
6	*Розроблення презентаційних матеріалів та попередній захист кваліфікаційної роботи*	*травень 2026*	Виконано
7	*Отримання відгуку керівника, рецензії, перевірка тексту кваліфікаційної роботи на плагіат, нормоконтроль*	*червень 2026*	Виконано
8	*Підготовка до захисту та захист кваліфікаційної роботи*	*червень 2026*	Виконано

Виконавець: *студент групи КН-22-1* (Група виконавця) _____ (Підпис) *Олександр ГЛАДУН* (Ім'я, ПРІЗВИЩЕ)

Керівник: *асистент каф. КН* (Науковий ступінь, посада) _____ (Підпис) *Ольга ЗАЛУЦЬКА* (Ім'я, ПРІЗВИЩЕ)

Анотація

Тема кваліфікаційної роботи бакалавра: *«Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу»*

Виконавець кваліфікаційної роботи бакалавра: *студент групи КН-22-1 Олександр Гладун*

Керівник кваліфікаційної роботи бакалавра: *асистент кафедри КН Ольга Залуцька*

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
69	20	7	50	2

Метою кваліфікаційної роботи бакалавра є підвищення точності визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

У межах роботи передбачається побудова підходу, орієнтованого на групову інтерпретацію подій у натовпі. Основу методу становитиме модель YOLO, призначена для детекції об'єктів і локалізації людей у кадрі, а також трансформерна модель ViT, що використовуватиметься для класифікації та формування інформативних ознак сцени.

Напрямками практичного використання розробленої інтелектуальної системи є автоматизований аналіз відеофіксації дій натовпу та визначення психологічних і соціокультурних характеристик угрупувань.

Ключові слова: нейромережа, комп'ютерний зір, аналіз натовпу, відеоаналітика, YOLO, Vision Transformer, психологічні характеристики, соціокультурні характеристики, інтелектуальна система.

Виконавець: *студент групи КН-22-1* (Група виконавця) _____ (Підпис) *Олександр ГЛАДУН* (Ім'я, ПРІЗВИЩЕ)

Зміст

Перелік скорочень

Скорочення, термін, позначення	Пояснення
КН	Комп’ютерні науки
ХНУ	Хмельницький національний університет.
AI	Artificial Intelligence
ViT	Vision Transformer
YOLO	You Look Only Once
mAP	mean Average Precision
AP	Average Precision
IoU	Intersection over Union
STAL	Small-Target-Aware Label Assignment
SiLU	Sigmoid Linear Unit
DFL	Distribution Focal Loss
GELU	Gaussian Error Linear Unit

Вступ

Актуальність. У сучасних умовах зростає потреба в автоматизованому аналізі поведінки людей у місцях масового скупчення, під час публічних заходів, соціальних акцій, спортивних подій та інших ситуацій, де дії натовпу можуть мати важливе соціальне, поведінкове та безпекове значення. Визначення психологічних і соціокультурних характеристик угрупувань за відеоданими дає змогу отримувати важливу інформацію про особливості групової взаємодії, рівень організованості, емоційний стан та характер поведінки учасників натовпу.

Традиційно аналіз таких характеристик здійснюється експертами на основі візуального спостереження або ручного перегляду відеоматеріалів. Такий підхід є трудомістким, потребує значних часових витрат, значною мірою залежить від суб’єктивної оцінки фахівця та є обмеженим в умовах великих обсягів відеоінформації. У задачах, де необхідно оперативно опрацьовувати значну кількість відеоданих, ручний аналіз стає малоефективним і не завжди забезпечує потрібний рівень точності та об’єктивності.

Розвиток методів комп’ютерного зору та глибокого навчання відкриває можливість створення інтелектуальних систем, здатних автоматично виявляти об’єкти у відеопотоці, аналізувати просторово-часові особливості сцени та визначати характерні ознаки поведінки груп людей. Використання сучасних нейромережевих моделей дає змогу підвищити точність аналізу, зменшити вплив людського чинника та забезпечити більш швидке й системне опрацювання відеоданих.

Актуальність теми зумовлена необхідністю створення підходів і засобів, які дозволяють автоматизувати процес визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу, що є важливим для задач моніторингу масових заходів, аналізу колективної поведінки, підтримки прийняття рішень у сфері безпеки, а також для соціологічних і поведінкових досліджень.

Об’єкт дослідження – процес нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Предмет дослідження – методи та засоби комп’ютерного зору на основі глибокого навчання для визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Мета кваліфікаційної роботи бакалавра – підвищення точності визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Завдання кваліфікаційної роботи бакалавра:

– провести аналіз предметної області та сучасних підходів визначення психологічних та соціокультурних характеристик угрупувань за відеоданимиіз застосуванням методів комп’ютерного зору та глибокого навчання;

– формалізувати задачу автоматизованого визначення психологічних та соціокультурних характеристик угрупувань за відеоданими;

– розробити метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу;

– реалізувати програмний засіб, що забезпечує обробку відеопотоку врежимі реального часу;

– провести експериментальне дослідження та оцінити ефективність розробленого методу.

Практичне значення одержаних результатів полягає у розробці цілісного нейромережевого методу та створенні на його основі програмного застосунку, призначеного для автоматизованого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу. Запропоноване рішення забезпечує послідовну обробку відеоданих, що включає детекцію учасників у кадрі, визначення їхнього емоційного стану, вікової групи та гендерних характеристик, а також подальше узагальнення отриманих результатів для формування інтегрального профілю угрупування.

Розроблений метод може застосовуватися фахівцями у сфері громадської безпеки, операторами систем відеоспостереження, аналітиками масових заходів та дослідниками колективної поведінки. Використання моделей YOLO та Vision Transformer дозволяє автоматично опрацьовувати значні обсяги відеоматеріалів, зменшувати залежність результатів від суб'єктивної оцінки людини та скорочувати час, необхідний для ручного перегляду записів. Система може бути використана для моніторингу масових заходів, аналізу ситуацій у громадських місцях, транспортній інфраструктурі, торговельних центрах, спортивних спорудах, а також у межах концепції Smart City та соціологічних досліджень.

Важливою практичною перевагою розробленого застосунку є можливість аналізу звичайних відеозаписів, статичних зображень і потоку зі стандартної вебкамери без використання спеціалізованих датчиків або складних багатокамерних комплексів. Це знижує вимоги до технічного оснащення та спрощує впровадження системи у наявну інфраструктуру відеоспостереження. Модульна структура програмного забезпечення також дає змогу замінювати нейромережеві моделі, розширювати перелік характеристик, що визначаються, та адаптувати розроблений метод до різних умов практичного застосування.

РОЗДІЛ 1 Характеристика предметної області: аналіз моделей, методів та реалізацій

1.1 Аналіз предметної області

У контексті комп’ютерних наук визначення характеристик угрупувань за відеоданими розглядається як задача інтелектуального аналізу відеопотоку, що передбачає формалізоване подання вхідної інформації, виділення інформативних візуальних ознак та їх автоматизовану інтерпретацію. Тому аналіз предметної області має охоплювати інформаційні моделі досліджуваних об’єктів, особливості обробки відеоданих, методи комп’ютерного зору та нейромережеві підходи до розпізнавання групових характеристик.

У сучасному суспільстві, що характеризується високим рівнем урбанізації, інтенсивністю масових заходів та постійним зростанням соціальної мобільності, особливої актуальності набуває питання забезпечення громадської безпеки у місцях скупчення людей [1]. Одним із найбільш поширених джерел інформації про поведінку людських угрупувань є відеофіксація, яка забезпечує можливість оперативного, безперервного та об'єктивного отримання даних про дії учасників масових заходів [2]. Однак ефективність використання таких даних значною мірою залежить від якісних характеристик їх обробки, зокрема – від релевантності виявлених ознак реальним психологічним та соціокультурним властивостям угрупування.

Під натовпом [3] у контексті соціальних досліджень зазвичай розуміють просторово локалізоване скупчення людей, що демонструє певні емерджентні властивості, не зведені до сумарних характеристик окремих осіб. Аналіз поведінки натовпу передбачає виявлення колективних патернів руху, щільності, спрямованості переміщення та емоційного забарвлення дій [4]. Цей напрям дозволяє здійснювати об'єктивне оцінювання стану громадського простору, уникаючи суб'єктивних чинників, притаманних традиційним формам спостереження. Водночас, дослідження поведінки угрупувань висуває вимоги до

структурної організації вихідних даних, відповідності виявлених ознак цільовим завданням аналізу та обґрунтованості інтерпретаційних висновків.

Аналіз поведінки угрупувань реалізується через виявлення різних типів характеристик, кожна з яких виконує певну функцію. Кількісні характеристики, такі як щільність та чисельність учасників, переважно застосовуються для оцінювання рівня заповненості простору з метою забезпечення безпеки. Динамічні характеристики, що відображають траєкторії руху та швидкісні параметри, дозволяють виявляти аномальні ситуації та використовуються на етапі попереджувального моніторингу [5, 6]. Психологічні характеристики, що відображають емоційний стан та рівень напруженості учасників, слугують інструментом для виявлення прихованих конфліктних ситуацій. Соціокультурні характеристики, що включають вікові, цисгендерні та культурно-символічні ознаки, дають змогу оцінити не лише поточний стан, а й типові поведінкові сценарії угрупування, що є особливо цінним при комплексному аналізі громадського простору [7].

Визначення психологічних та соціокультурних характеристик угрупувань передбачає встановлення ступеня відповідності виявлених у відеоматеріалах ознак певним семантичним категоріям. Релевантна характеристика повинна прямо або опосередковано відображати реальні властивості угрупування і бути спрямованою на формування саме тих понять і логічних зв'язків, що становлять зміст аналітичної задачі. Високий рівень релевантності дозволяє забезпечити обґрунтованість висновків як основи для прийняття управлінських рішень, тоді як її відсутність призводить до викривлення результатів аналізу, зниження довіри до системи та порушення принципу справедливості при оцінюванні соціальних явищ.

Автоматизація процесу визначення характеристик угрупувань є важливою задачею у контексті забезпечення громадської безпеки [8]. Вона дає змогу суттєво зменшити часові витрати на експертну обробку значних обсягів відеоматеріалів, підвищити рівень обґрунтованості висновків та їх відповідності реальним подіям, а також адаптувати аналітичні засоби до динамічних змін у структурі суспільних процесів. Інформаційні моделі, які реалізують механізми

зіставлення виявлених у відеопотоці ознак із психологічними та соціокультурними категоріями, є основою для побудови подібних автоматизованих рішень.

Отже, аналіз інформаційних моделей у сфері автоматизованого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу вказує на необхідність розробки методів, здатних виявляти змістовні відповідності між елементами візуального потоку та аналітичними об'єктами предметної області.

1.2 Огляд теоретичних підходів до розв’язку подібних задач

Сучасні методи штучного інтелекту знаходять широке застосування у задачі визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу, оскільки її вирішення передбачає автоматизовану обробку великих обсягів візуальної інформації з метою виявлення складних семантичних властивостей групової поведінки [9]. Складність цієї задачі обумовлює необхідність одночасного розв'язання двох взаємопов'язаних підзадач – виявлення учасників угрупування у відеопотоці та інтерпретації їхніх колективних ознак, таких як емоційний стан, цисгендер та вік учасників. Традиційні алгоритми обробки зображень не забезпечують достатнього рівня узагальнення для розпізнавання таких багатовимірних характеристик, що обумовлює вибір нейромережевих підходів як основного інструментарію.

Базовим теоретичним підґрунтям для першого етапу слугують згорткові нейронні мережі, клас глибоких нейронних мереж, спеціально розроблений для обробки даних із просторовою структурою, зокрема зображень та відео [10]. Принцип роботи згорткових нейронних мереж ґрунтується на послідовному застосуванні операцій згортки, що дозволяє автоматично виявляти ознаки зростаючої складності – від базових на перших шарах до семантично значущих на глибших рівнях мережі [11]. Завдяки механізму спільного використання вагових коефіцієнтів згорткові нейронні мережі суттєво скорочують кількість

параметрів порівняно із звичайними повнозв'язними мережами, що робить їх ефективними при роботі з відеоданими високої роздільної здатності.

На основі архітектури згорткових нейронних мереж побудована модель YOLO – нейромережева модель для детекції об'єктів у реальному часі [12]. Ключовою особливістю YOLO є принцип одного проходу – модель аналізує зображення цілісно за один прохід через мережу, розбиваючи його на сітку комірок, кожна з яких прогнозує координати обмежувальних рамок та ймовірності належності виявлених об'єктів до певного класу [13]. Такий підхід забезпечує значно вищу швидкість обробки порівняно з двоетапними детекторами при збереженні прийнятної точності, що є критично важливим для аналізу відеопотоку в режимі реального часу. У рамках даної роботи YOLO виконує роль першого етапу обробки – визначає місцезнаходження учасників натовпу у кожному кадрі та формує обмежувальні рамки навколо виявлених угрупувань, які передаються на вхід наступного етапу – семантичного розпізнавання.

Для розв'язання задач другого етапу – розпізнавання психологічних та соціокультурних характеристик угрупувань – застосовується архітектура ViT [14]. ViT є адаптацією трансформерної архітектури до задач комп'ютерного зору і принципово відрізняється від згорткових нейронних мереж підходом до аналізу зображення. Вхідний фрагмент, виділений YOLO, розбивається на фіксовані патчі, які перетворюються на послідовність векторних представлень і подаються на вхід механізму самоуваги [15]. Механізм самоуваги дозволяє моделі виявляти глобальні залежності між будь-якими патчами зображення незалежно від їхнього просторового розташування – що є критично важливим для розпізнавання контекстуальних зв'язків між учасниками натовпу, їхніх поз, жестів, взаємного розташування та емоційного забарвлення дій [16]. На відміну від згорткових нейронних мереж, які обмежені локальним рецептивним полем, ViT здатний аналізувати глобальний контекст сцени, що забезпечує якісніше розуміння семантики групової поведінки та дозволяє формувати інтегральні характеристики угрупування – рівень агресивності, напруженості, соціокультурні ознаки учасників тощо.

Таким чином, у роботі реалізується двоетапний нейромережевий конвеєр – на першому етапі YOLO здійснює детекцію та локалізацію учасників угрупування у відеопотоці, виділяючи релевантні фрагменти сцени, а на другому етапі ViT виконує глибоке семантичне розпізнавання психологічних та соціокультурних характеристик виявлених угрупувань на основі глобального аналізу контексту.

1.3 Аналіз існуючих програмних засобів та наукових рішень

У контексті розробки методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу важливим етапом є аналіз існуючих наукових публікацій, що висвітлюють подібні дослідницькі задачі. Розгляд опублікованих наукових робіт дозволяє виявити сучасні підходи до розв'язання задач детекції учасників натовпу та розпізнавання їхніх характеристик, а також встановити обмеження існуючих рішень, які потребують подальшого дослідження.

У статті [17] автори запропонували підхід до детекції людей у відеопотоці на основі моделей YOLO. У дослідженні розглянуто застосування моделей YOLO для виявлення та підрахунку людей у замкнених приміщеннях, проведено порівняльний аналіз ефективності різних версій моделі. Автори наголошують на високій швидкості обробки відеопотоку та точності детекції в умовах реального часу. Перевагою роботи є детальний експериментальний аналіз продуктивності моделей YOLO для задачі детекції людей. Недоліком є те, що дослідження обмежується задачею кількісної оцінки натовпу і не охоплює задачі визначення психологічних чи соціокультурних характеристик виявлених угрупувань.

У роботі [18] досліджено можливості моделі YOLOv8 для точного виявлення та підрахунку людей у статичних зображеннях і відеопотоці. Автори зосереджуються на проблемах перекриття людей у щільному натовпі та значних варіаціях розміру і форми постатей через перспективні спотворення. Перевагою роботи є врахування специфіки роботи у щільних натовпах. Недоліком є відсутність аналізу інтегральних характеристик угрупування – робота

сфокусована виключно на задачі підрахунку індивідуальних осіб без формування семантичного профілю групи.

Автори [19] запропоновано підхід до виявлення поведінки натовпу на основі трансформерної архітектури Swin Transformer, що використовує карти підрахунку людей та оптичного потоку для класифікації поведінки за розміром натовпу та рівнем агресивності. Автори демонструють перевагу трансформерних архітектур у задачах аналізу складної поведінки угрупувань. Перевагою роботи є застосування сучасної трансформерної архітектури для семантичного аналізу натовпу. Недоліком є те, що дослідження зосереджене на класифікації агресивності і не охоплює визначення таких характеристик угрупування, як емоційний стан, цисгендерний склад чи вікові характеристики.

У роботі [20] представлено підхід до розпізнавання групових емоцій у відеозаписах натовпу на основі глибокого об'єднання просторово-часових ознак. Автори представили датасет GECV із 627 відеозаписів, у якому кожен запис розмічено на трьох рівнях: окремі обличчя, група людей та кадр у цілому. Перевагою роботи є визнання важливості агрегування характеристик на рівні угрупування, а не лише окремих осіб. Недоліком є те, що дослідження обмежується розпізнаванням лише емоційних характеристик без врахування соціокультурних ознак, таких як вік та цисгендер.

У статті [21] досліджено задачу визначення цисгенедру та вікових характеристик в умовах щільного натовпу в режимі реального часу. Автори демонструють, що стандартні моделі аналізу облич суттєво знижують точність у сценах з перекриттям та частковою видимістю учасників. Перевагою роботи є детальний аналіз специфіки задачі в умовах реального натовпу та запропоновані оптимізації для роботи у складних сценаріях. Недоліком є те, що дослідження зосереджене лише на цисгендерних та вікових атрибутах і не враховує психологічних характеристик угрупування як єдиного цілого.

Серед програмних засобів у даній предметній області варто виділити платформу Face++ [22] – веб-сервіс для аналізу атрибутів обличчя на основі штучного інтелекту (рисунок 1.1). Платформа забезпечує визначення емоційного стану, віку та цисгендеру осіб на зображеннях і у відеопотоці – зокрема

розпізнавання таких емоцій, як радість, смуток, гнів, страх, здивування та нейтральний стан – з поверненням числових оцінок впевненості для кожного виявленого атрибута. Система підтримує одночасний аналіз кількох облич у кадрі та здійснює виявлення більше 70 ключових точок обличчя, що забезпечує точність розпізнавання атрибутів.

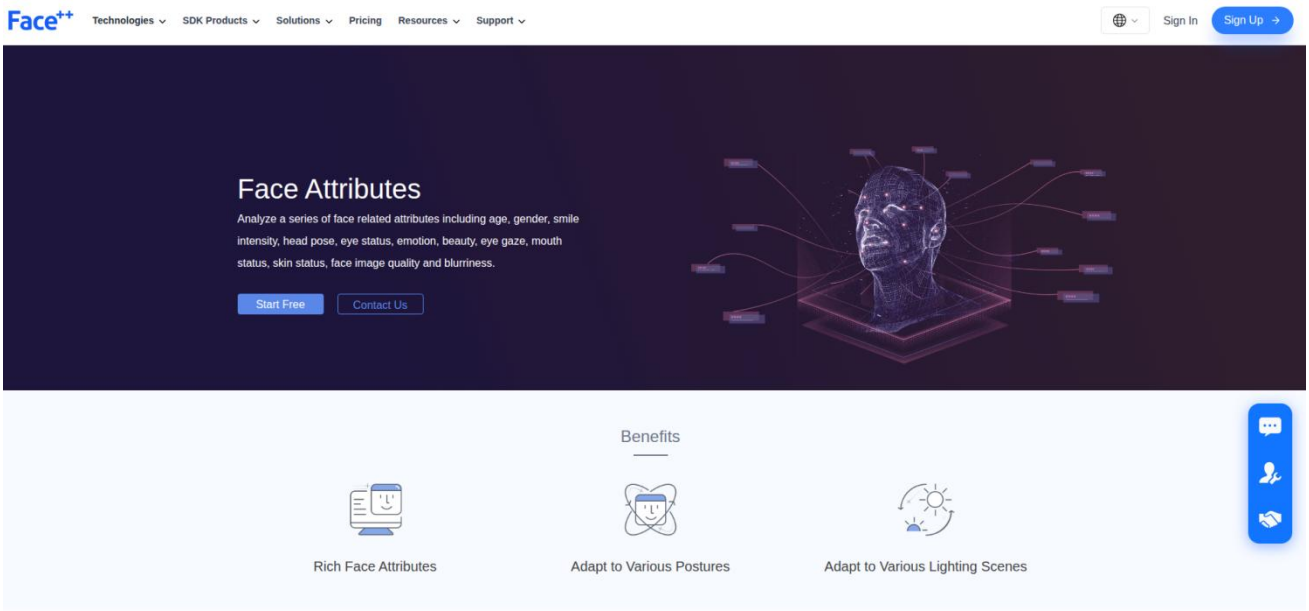


Рисунок 1.1 – Платформа Face++ [22]

До переваг платформи належать широкий набір атрибутів, що аналізуються, висока точність розпізнавання та можливість одночасної обробки кількох осіб у кадрі. Недоліком є те, що Face++ формує результат аналізу для кожного обличчя окремо і не агрегує виявлені характеристики на рівні угрупування. Окрім того, платформа не містить механізму детекції та виділення самих угрупувань у відеопотоці.

Іншим відомим рішенням у сфері відеоаналітики є платформа BriefCam [23] – програмний продукт для інтелектуального аналізу відеоспостереження, що реалізований як корпоративний застосунок з трьома модулями: REVIEW для прискореного пошуку подій у відеозаписах, RESPOND для сповіщень у режимі реального часу та RESEARCH для формування аналітичних звітів (рисунок 1.2). У версії 2024 R2 платформа отримала новий модуль Crowd Counting – алгоритм оцінювання розміру натовпу у режимі реального часу, що автоматично активується при перевищенні порогової кількості людей у кадрі та доповнює алгоритми підрахунку людей і виявлення груп. Платформа забезпечує

формування теплових карт щільності, відстеження траєкторій руху та виявлення аномальної поведінки.

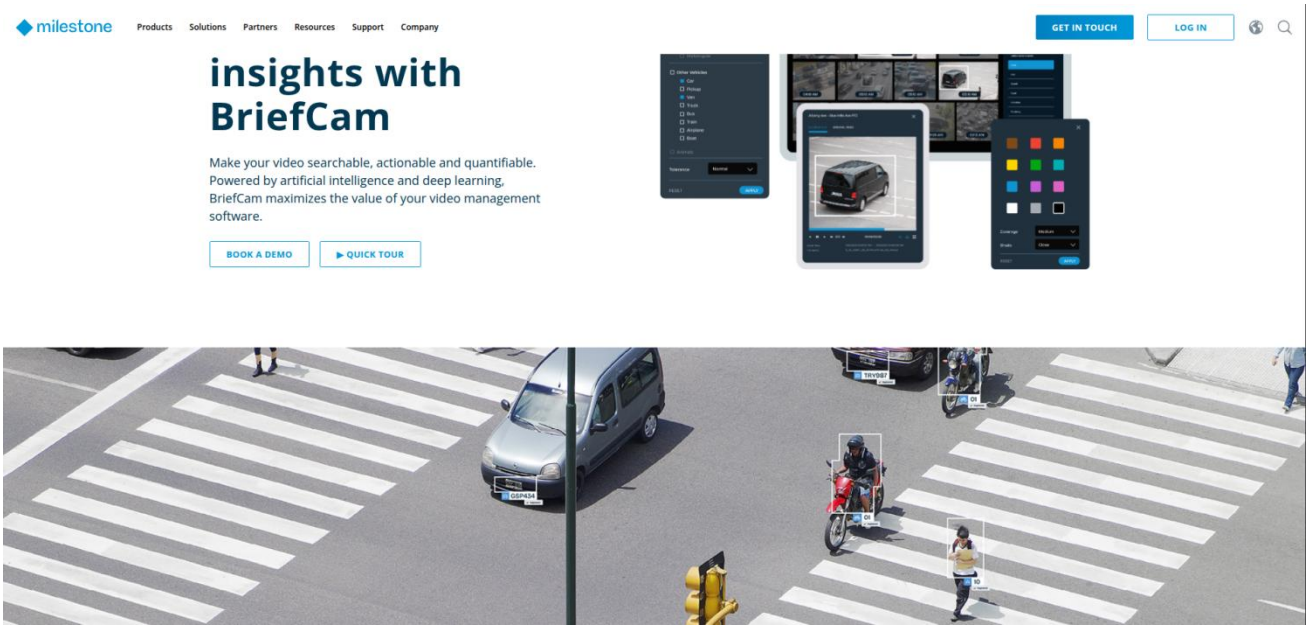


Рисунок 1.2 – Платформа BriefCam[23]

До переваг платформи належать широкий функціонал аналізу відеоданих, інтеграція з існуючими системами відеоспостереження та можливість роботи у режимі реального часу. Недоліком є те, що BriefCam зосереджена на кількісних та просторових характеристиках натовпу – підрахунку, щільності, траєкторіях – і не забезпечує визначення психологічних характеристик угрупування, таких як емоційний стан, та соціокультурних ознак, зокрема вікового і цисгендерного складу групи як єдиної аналітичної одиниці.

Отже, аналіз розглянутих програмних засобів та наукових публікацій демонструє, що існуючі рішення зосереджуються переважно на окремих аспектах аналізу натовпу. Програмні засоби – зокрема Face++ та BriefCam – забезпечують або аналіз атрибутів окремих облич, або кількісну оцінку щільності натовпу, однак не формують інтегрального психологічного та соціокультурного профілю угрупування як цілісного об'єкта. Наукові публікації, своєю чергою, розглядають детекцію учасників, класифікацію поведінки або визначення окремих демографічних характеристик без комплексного підходу до одночасного аналізу емоційного стану, цисгендерного складу та вікових ознак групи. У зв'язку з цим виникає потреба у розробці спеціалізованого методу, що

дозволяє комплексно вирішити задачу визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

1.4 Мета, задачі та вимоги до реалізаціїінтелектуальної системи

Метою кваліфікаційної роботи бакалавра є підвищення точності визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Для реалізації поставленої мети необхідно виконати такі завдання:

– провести аналіз предметної області та сучасних підходів визначення психологічних та соціокультурних характеристик угрупувань за відеоданими із застосуванням методів комп’ютерного зору та глибокого навчання;

– формалізувати задачу автоматизованого визначення психологічних та соціокультурних характеристик угрупувань за відеоданими;

– розробити метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу;

– реалізувати програмний засіб, що забезпечує обробку відеопотоку в режимі реального часу;

– провести експериментальне дослідження та оцінити ефективність розробленого методу.

РОЗДІЛ 2 Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу

2.1 Формалізація задачі визначення психологічних та соціокультурних характеристик угрупувань за відеопотоком

2.1.1 Математична формалізація задачі

Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу орієнтований на автоматизоване формування інтегрального профілю угрупування – емоційного стану, цисгендерного складу та вікових ознак учасників – безпосередньо з відеопотоку без участі людини-експерта. Профіль формується для угрупування як цілісного об'єкта аналізу, а не для окремих осіб, що забезпечує отримання узагальненого опису групи у формі, зручній для подальшої інтерпретації та прийняття управлінських рішень.

Основна ідея методу полягає у двоетапній нейромережевій обробці відеопотоку, що поєднує дві сучасні архітектури глибокого навчання. На першому етапі здійснюється детекція та локалізація учасників угрупування у кадрі засобами моделі YOLO – нейромережевої архітектури, спеціалізованої на швидкісній детекції об'єктів у режимі реального часу за принципом одного проходу через мережу. Модель YOLO забезпечує одночасне визначення координат обмежувальних рамок та оцінку впевненості детекції для всіх об'єктів у кадрі, що дозволяє ефективно обробляти відеопотік навіть в умовах щільного натовпу. Результатом першого етапу є набір обмежувальних рамок навколо виявлених осіб разом з оцінками впевненості детекції, що передаються на вхід наступного етапу обробки.

На другому етапі сформовані фрагменти сцени, що відповідають виявленим учасникам угрупування, передаються на вхід моделі Vision Transformer, яка виконує глибокий семантичний аналіз та формує інтегральні характеристики угрупування – емоційний стан, цисгендерний склад та вікові ознаки учасників. Vision Transformer принципово відрізняється від традиційних

згорткових нейронних мереж підходом до обробки зображення – вхідні дані розбиваються на патчі фіксованого розміру, кожен з яких перетворюється на векторне представлення, після чого послідовність векторів подається на вхід механізму самоуваги. Завдяки цьому механізму модель здатна аналізувати глобальний контекст сцени та виявляти контекстуальні зв'язки між учасниками натовпу – їхні пози, жести, вирази облич та емоційне забарвлення дій. Це забезпечує якісніше розуміння семантики групової поведінки порівняно зі згортковими нейронними мережами, обмеженими локальним рецептивним полем.

Ключовою особливістю запропонованого методу є те, що характеристики формуються не для окремих осіб, а для угрупування як цілісного об'єкта – результатом є узагальнений профіль групи, а не набір індивідуальних атрибутів. Інтегральний профіль угрупування забезпечує стислу та інтерпретовану інформацію про стан групи, що значно скорочує час сприйняття результатів аналізу та підвищує зручність використання методу у складі прикладних систем.

Вхідними даними методу є відеофіксація дій натовпу у вигляді послідовності кадрів. Послідовність може бути отримана як з потокового джерела – камер відеоспостереження, що працюють у режимі реального часу, так і з попередньо записаних відеофайлів, що зберігаються в архівах систем спостереження. Метод не накладає жорстких обмежень на параметри вхідного відеопотоку – роздільну здатність, частоту кадрів чи кодек відео – проте якість результатів аналізу безпосередньо залежить від чіткості зображення та видимості учасників натовпу. Вихідними даними методу є структурований профіль угрупування, що містить три складові – переважний емоційний стан угрупування, його цисгендерний склад та переважні вікові характеристики учасників. Таким чином, метод реалізує повний аналітичний ланцюжок від необробленого відеопотоку до інтерпретованого психологічного та соціокультурного опису угрупування.

Узагальнена схема методу, що ілюструє двоетапну обробку відеопотоку – детекцію учасників угрупування засобами YOLO та подальше семантичне

розпізнавання характеристик засобами Vision Transformer – наведена на рисунку 2.1.

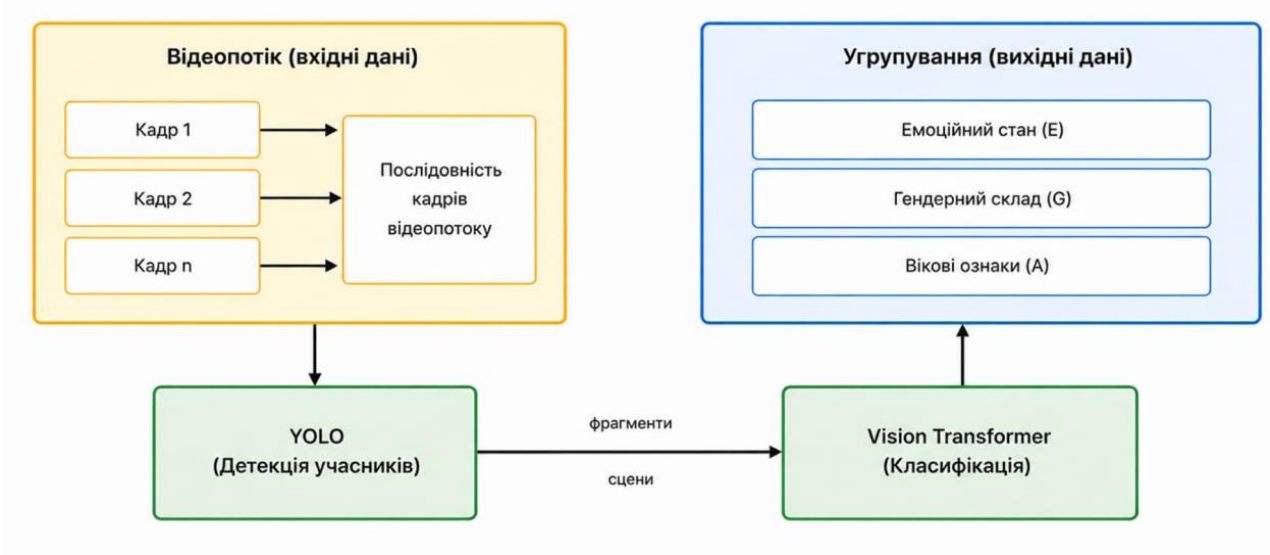


Рисунок 2.1 – Узагальнена схема методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань

Метод поєднує переваги двох сучасних нейромережевих архітектур – швидкодію YOLO у задачах детекції об'єктів та здатність Vision Transformer аналізувати глобальний контекст сцени. Завдяки спеціалізації моделі YOLO на задачі детекції забезпечується висока швидкість обробки навіть на потокових відеоданих у режимі реального часу, тоді як використання Vision Transformer на етапі семантичного аналізу гарантує високу якість розпізнавання характеристик угрупування з урахуванням глобальних залежностей у сцені. Таке поєднання дозволяє забезпечити комплексне визначення психологічних і соціокультурних характеристик угрупувань у відеопотоці з прийнятною точністю та швидкістю обробки.

Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань реалізує послідовну обробку відеоданих у чотири основні етапи – від необробленого відеопотоку до формування інтегрального профілю угрупування. Кожен етап методу виконує специфічну функцію в загальному аналітичному ланцюжку та формує проміжний результат, що передається на вхід наступного етапу. Послідовність етапів, їх взаємозв'язки та структура обробки даних наведені на рисунку 2.2.

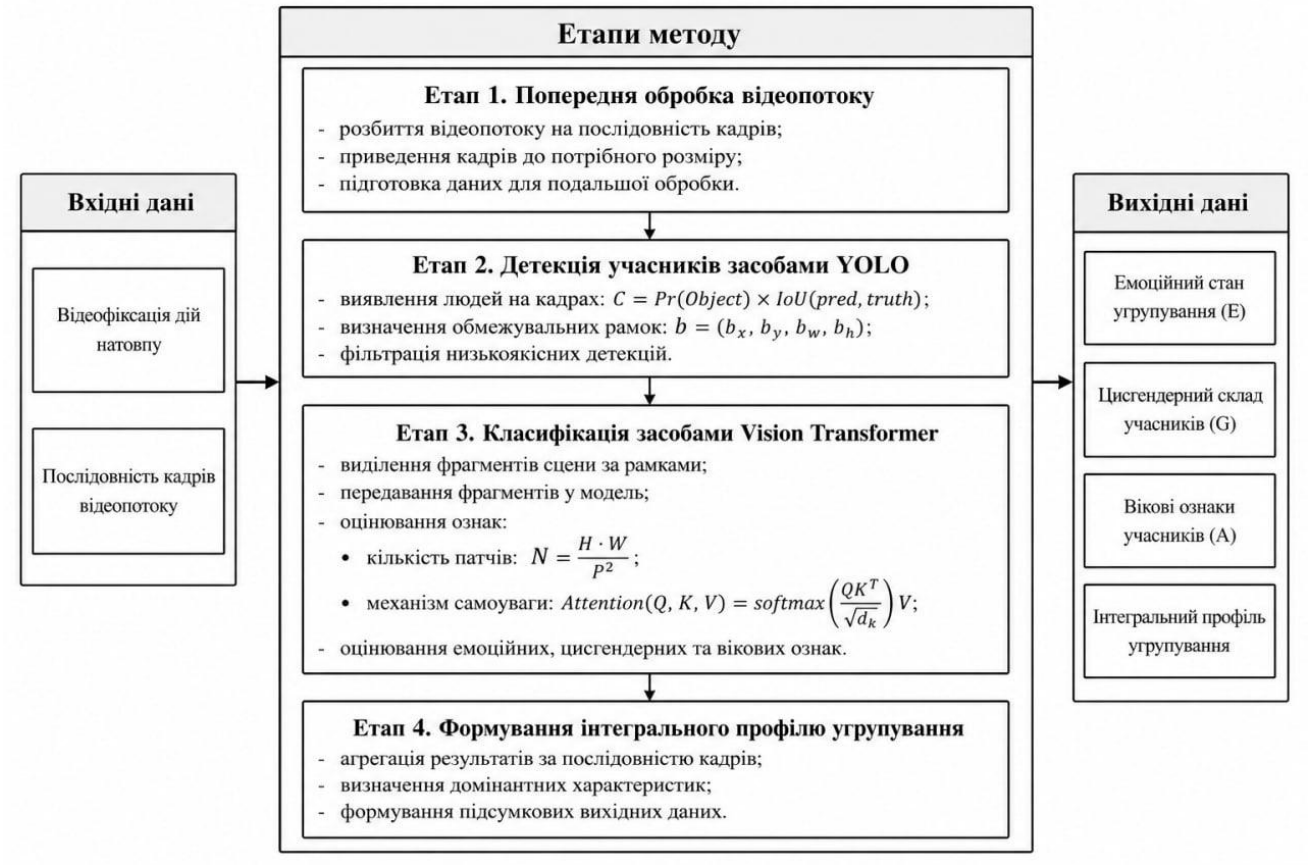


Рисунок 2.2 – Схема етапів методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань

На кроці 1 виконується попередня обробка відеопотоку, що є необхідним підготовчим етапом для подальшої роботи нейромережевих моделей. Вхідний відеопотік розбивається на послідовність окремих кадрів із заданою частотою, що залежить від типу аналізованої сцени та вимог до швидкодії системи. Зменшення частоти вибірки кадрів дозволяє суттєво скоротити обчислювальні витрати, проте може призвести до пропуску короткочасних змін у поведінці угрупування, тоді як обробка кожного кадру забезпечує максимальну деталізацію аналізу за рахунок збільшення часу обробки. Подальша обробка кожного кадру передбачає приведення зображення до єдиного розміру, прийнятного для подальшої роботи нейромережевих моделей, та нормалізацію значень пікселів у діапазон, у якому навчалися моделі YOLO та Vision Transformer. Нормалізація є критично важливою для коректної роботи нейромережевих моделей, оскільки відхилення розподілу вхідних даних від розподілу навчальної вибірки може призвести до суттєвого зниження точності прогнозу.

На кроці 2 здійснюється детекція учасників угрупування засобами моделі YOLO. Детекція базується на принципі регресії обмежувальних рамок – зображення розбивається на сітку комірок розміром S × S, і для кожної комірки модель прогнозує координати обмежувальних рамок та оцінку впевненості детекції. Кожна комірка сітки відповідає за виявлення об'єктів, центр яких потрапляє в межі цієї комірки, що забезпечує просторову локалізацію об'єктів без необхідності генерації регіонів-кандидатів, як це роблять двоетапні детектори. Оцінка впевненості визначається за формулою [24]:

$$C = Pr(Object) \times IOU(pred, truth), \quad (2.1)$$

де *Pr*(*Object*) – ймовірність наявності об'єкта у комірці сітки; *IOU(pred, truth)* – значення перетину між передбаченою та реальною обмежувальними рамками.

Така формула забезпечує одночасне врахування двох аспектів якості детекції – ймовірності правильної ідентифікації об'єкта та точності локалізації його меж. Якщо у комірці відсутній об'єкт, ймовірність *Pr(Object)* прямує до нуля, що автоматично занулює оцінку впевненості та виключає таку детекцію з подальшої обробки. Локалізація кожного виявленого учасника задається вектором, що містить координати центру, ширину та висоту обмежувальної рамки відносно розміру зображення [24]:

$$b = (b_x, b_y, b_w, b_h), \quad (2.2)$$

де b_x, b_y – координати центру рамки відносно комірки сітки; b_w, b_h – ширина та висота рамки відносно розміру зображення.

Координати b_x, b_y задаються відносно меж конкретної комірки сітки, що забезпечує локальну прив'язку детекції та спрощує процес навчання моделі. Ширина та висота рамки нормуються відносно розміру всього зображення, що дозволяє моделі коректно працювати з об'єктами різних масштабів – від великих постатей на передньому плані до малих фігур у глибині сцени. Після отримання детекцій виконується фільтрація низькоякісних результатів – видаляються рамки з оцінкою впевненості нижче заданого порогу, що дозволяє уникнути аналізу хибнопозитивних детекцій на наступних етапах. Додатково застосовується алгоритм придушення мінімумів, який усуває дублюючі рамки, що відповідають одному й тому ж об'єкту, залишаючи лише рамку з найвищою оцінкою

впевненості. Це необхідно через те, що модель YOLO може генерувати кілька рамок для одного й того ж об'єкта з різних комірок сітки. Результатом кроку 2 є відфільтрований набір обмежувальних рамок, що локалізують учасників угрупування у кожному кадрі відеопотоку.

На кроці 3 здійснюється класифікація фрагментів сцени засобами моделі Vision Transformer. За координатами рамок, отриманих на кроці 2, з кадрів вирізаються фрагменти сцени, що містять виявлених учасників угрупування. Кожен фрагмент приводиться до фіксованого розміру, прийнятного для моделі Vision Transformer – зазвичай 224 × 224 пікселі, що відповідає стандартному вхідному формату попередньо натренованих моделей. Зображення розбивається на патчі фіксованого розміру P × P пікселів, кількість яких визначається за формулою [25]:

$$N = HW / P^2, \tag{2.3}$$

де H, W – просторові розміри вхідного зображення; P – розмір сторони патчу у пікселях.

Кожен патч перетворюється на векторне представлення шляхом лінійної проєкції та доповнюється позиційним кодуванням, що зберігає інформацію про просторове розташування патчу у вихідному зображенні. Сформована послідовність векторів подається на вхід механізму самоуваги, що обчислюється за формулою [26]:

$$Attention(Q, K, V) = softmax(QK^T / \sqrt{d_k}) \times V, \tag{2.4}$$

де Q – матриця запитів; K – матриця ключів; V – матриця значень; d_k – розмірність векторів ключів, що використовується для масштабування.

Механізм самоуваги обчислює вагові коефіцієнти, що відображають ступінь взаємодії між кожним патчем зображення та усіма іншими патчами. Масштабування на $\sqrt{d_k}$ застосовується для запобігання надмірному зростанню значень скалярних добутків при високих розмірностях векторів, що могло б призвести до проблем зі збіжністю функції softmax. Функція softmax перетворює отримані оцінки у нормований розподіл ваг, що використовується для зваженого підсумовування матриці значень. Таким чином, механізм самоуваги дозволяє моделі виявляти глобальні залежності між патчами зображення незалежно від

їхнього просторового розташування, що є критично важливим для розпізнавання контекстуальних зв'язків між учасниками натовпу – їхніх поз, жестів, виразів облич та емоційного забарвлення дій. На відміну від згорткових нейронних мереж, обмежених локальним рецептивним полем, Vision Transformer аналізує глобальний контекст сцени вже на перших шарах обробки, що забезпечує якісніше розуміння семантики групової поведінки. На виході моделі формуються оцінки належності кожного учасника до класів за трьома параметрами – емоційний стан, цисгендер та вікова група. Результатом кроку 3 є набір ознак для кожного учасника, виявленого на кроці 2.

На кроці 4 здійснюється формування інтегрального профілю угрупування. Ознаки, отримані для окремих учасників на кроці 3, агрегуються за всіма кадрами відеопотоку та усіма виявленими учасниками угрупування. Агрегування дозволяє перейти від характеристик окремих осіб до інтегральних показників групи як цілісного об'єкта аналізу, що відповідає основній меті методу. У процесі агрегування враховуються частотні розподіли значень кожної характеристики серед учасників – переважний емоційний стан визначається як домінуючий клас серед усіх виявлених осіб, цисгендерний склад – як співвідношення гендерних груп у відсотках від загальної кількості учасників, вікові характеристики – як домінуюча вікова група та її частка від загальної кількості учасників. На основі агрегованих даних обчислюються три підсумкові показники – переважний емоційний стан угрупування (E), його цисгендерний склад (G) та вікові характеристики учасників (A).

Результатом кроку 4 та методу в цілому є структурований профіль угрупування, що включає визначені значення E, G та A. Такий профіль є інтерпретованим описом групи, що може використовуватися для подальшого прийняття управлінських рішень у сфері громадської безпеки, аналізу поведінки натовпу або планування міського простору. Сформований профіль може бути візуалізований у вигляді діаграм, гістограм або текстового опису, що забезпечує зручність сприйняття інформації оператором системи. Запропонований метод забезпечує комплексне розв'язання задачі визначення характеристик угрупування, у якому модель YOLO відповідає за виявлення учасників, модель

Vision Transformer – за глибокий семантичний аналіз їхніх ознак, а етап агрегування – за формування інтегральних характеристик групи як єдиного цілого. Такий розподіл функцій між компонентами методу забезпечує його модульність, що дозволяє у подальшому замінювати окремі компоненти на більш сучасні версії без необхідності переробки всієї системи.

2.1.2 Математичне подання конвеєра обробки відеоданих

Формалізація методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань у вигляді математичного псевдокоду дозволяє чітко описати послідовність операцій, їх вхідні та вихідні параметри, а також забезпечує однозначну інтерпретацію алгоритмічної логіки методу. Конвеєр обробки відеоданих складається з двох основних алгоритмів – алгоритму просторової детекції учасників угрупування у відеопотоці засобами моделі YOLO26 та алгоритму класифікації психологічних і соціокультурних характеристик з подальшим формуванням інтегрального профілю засобами моделі Vision Transformer. Розділення конвеєра на два окремі алгоритми відповідає двоетапній структурі методу, описаній у п.2.1, та забезпечує модульність реалізації, що дозволяє вдосконалювати кожну складову незалежно від іншої.

Алгоритм 2.1 – Математичний псевдокод просторової детекції учасників угрупування

Вхідні дані: Відеопотік $V = \{I_1, I_2, ..., I_n\}$; навчена модель детекції D_θ; порогове значення впевненості детекції τ.

Вихідні дані: Множина обмежувальних рамок учасників за всіма кадрами $B = \{B_1, B_2, ..., B_n\}$.

1. $B \leftarrow \emptyset$ // Ініціалізація множини рамок
2. **Для** $t = 1$ **до** N **виконувати:**
   - $\hat{I}_t \leftarrow Normalize(Resize(I_t, 640))$ // Низькорівневе опрацювання кадру

- $B_t \leftarrow D\_\theta(\hat{I}_t)$ // Детекція з обчисленням C за формулою (2.1) та b за формулою (2.2)
- $\tilde{B}_t \leftarrow Filter(B_t, \tau)$ // Фільтрація рамок з $C < \tau$
- $B \leftarrow B \cup \{(t, \tilde{B}_t)\}$

3. Кінець циклу

4. Повернути B

Алгоритм 2.1 реалізує перший етап обробки – виявлення учасників угрупування у кадрах відеопотоку. На вхід алгоритму подається послідовність кадрів V, навчена модель детекції D_θ з параметрами θ та порогове значення впевненості τ, що визначає мінімальну допустиму якість детекції. Операція $Normalize(Resize(I_t, 640))$ виконує приведення кадру до розміру 640 × 640 пікселів та нормалізацію значень пікселів у діапазон, прийнятний для моделі YOLO26. Операція $D\_\theta(\hat{I}_t)$ застосовує модель детекції до кожного кадру – для кожної обмежувальної рамки обчислюється оцінка впевненості C згідно з формулою (2.1) та формується вектор локалізації $b = (b_x, b_y, b_w, b_h)$ згідно з формулою (2.2). Операція $Filter(B_t, \tau)$ видаляє з множини рамки з оцінкою впевненості C нижче порогу τ, що зменшує кількість хибнопозитивних детекцій. Результатом алгоритму є множина B, що містить відфільтровані рамки з прив'язкою до номера кадру, у якому вони виявлені.

Алгоритм 2.2 – Математичний псевдокод класифікації характеристик та формування профілю угрупування

Вхідні дані: Послідовність кадрів $V = \{I_1, I_2, ..., I_n\}$; множина рамок B з алгоритму 2.1; навчена модель класифікації емоцій $C\varepsilon$; навчена модель класифікації віку $C\alpha$; навчена модель класифікації гендеру $C\gamma$.

Вихідні дані: Інтегральний профіль угрупування $P = \{E, G, A\}$.

1. $R \leftarrow \emptyset$ // Ініціалізація множини ознак учасників

2. **Для кожного** $(t, \tilde{B}_t) \in B$ **виконувати:**
   - **Для кожного** $b \in \tilde{B}_t$ **виконувати:**
     - $F \leftarrow Crop(I_t, b)$ // Виділення фрагмента за рамкою

- $F^{\wedge} \leftarrow Normalize(Resize(F, 224))$ // Розбиття на N патчів за формулою (2.3)
- $e \leftarrow C\varepsilon(F^{\wedge})$ // Класифікація емоцій з механізмом самоуваги за формулою (2.4)
- $a \leftarrow C\alpha(F^{\wedge})$ // Класифікація віку з механізмом самоуваги за формулою (2.4)
- $g \leftarrow C\gamma(F^{\wedge})$ // Класифікація гендеру з механізмом самоуваги за формулою (2.4)
- $R \leftarrow R \cup \{(e, a, g)\}$

3. Кінець циклу
4. $E \leftarrow AggregateE(R)$ // Агрегація емоційних ознак
5. $G \leftarrow AggregateG(R)$ // Агрегація гендерних ознак
6. $A \leftarrow AggregateA(R)$ // Агрегація вікових ознак
7. $P \leftarrow \{E, G, A\}$
8. Повернути P

Алгоритм 2.2 реалізує другий етап обробки – семантичну класифікацію характеристик кожного учасника та подальшу агрегацію результатів у інтегральний профіль угрупування. На вхід алгоритму подається послідовність кадрів V, множина обмежувальних рамок B, отримана на виході алгоритму 2.1, та три навчені моделі класифікації – $C\varepsilon$ для емоційного стану, $C\alpha$ для вікової групи, $C\gamma$ для цисгендеру. У запропонованому методі застосовується підхід окремих моделей для кожної характеристики, що забезпечує гнучкість донавчання та дозволяє оптимізувати кожну модель окремо під свою специфічну задачу.

Для кожної обмежувальної рамки b виконується послідовність операцій. Операція $Crop(I_t, b)$ виділяє з кадру I_t фрагмент сцени, що містить виявленого учасника. Операція $Normalize(Resize(F, 224))$ приводить фрагмент до розміру 224 × 224 пікселі – стандартного вхідного формату моделі Vision Transformer – та нормалізує значення пікселів. Підготовлений фрагмент $F^{\wedge}$ розбивається на

N патчів розміром 16 × 16 пікселів, кількість яких визначається за формулою (2.3). Послідовність векторних представлень патчів подається на вхід механізму самоуваги, що обчислюється за формулою (2.4) у кожній з трьох моделей класифікації – $C\varepsilon$, $C\alpha$ та $C\gamma$. Модель $C\varepsilon$ повертає індекс домінуючого класу серед семи емоційних станів («гнів», «огида», «страх», «радість», «нейтральний стан», «сум», «здивування»), модель $C\alpha$ – серед дев'яти вікових інтервалів, модель $C\gamma$ – бінарне значення цисгендеру. Отримані ознаки (e, a, g) додаються до множини R, що накопичує характеристики всіх виявлених учасників.

Після завершення обробки всіх рамок виконуються операції агрегації *Aggregate_E*, *Aggregate_G* та *Aggregate_A*, що обчислюють підсумкові інтегральні характеристики угрупування – переважний емоційний стан E, цисгендерний склад G та вікові ознаки A. Агрегація реалізується через визначення домінуючого класу для кожної характеристики та формування частотного розподілу значень серед всіх учасників. Такий підхід забезпечує перехід від атрибутів окремих осіб до інтегрального профілю угрупування як цілісного об'єкта аналізу.

Формалізація конвеєра у вигляді двох математичних псевдокодів забезпечує чітку специфікацію алгоритмічної логіки розробленого методу. Алгоритми 2.1 та 2.2 у сукупності описують повний цикл обробки відеоданих – від необробленого вхідного відеопотоку до сформованого інтегрального профілю угрупування у вигляді трійки $\{E, G, A\}$. Розділення конвеєра на два незалежні алгоритми відповідає двоетапній архітектурі методу та забезпечує модульність реалізації, оскільки кожен з алгоритмів використовує власну нейромережеву модель з власними параметрами та має чітко визначені вхідні і вихідні дані. Узгодженість інтерфейсів між алгоритмами досягається за рахунок того, що вихідна множина обмежувальних рамок B першого алгоритму безпосередньо є частиною вхідних даних другого алгоритму, що формує єдиний обчислювальний конвеєр. Описана формалізація є основою для подальшої програмної реалізації інтелектуальної системи та визначає вимоги до архітектурних рішень, які буде представлено у наступних пунктах.

2.2 Нейромережеве визначення психологічних та соціокультурних характеристик угрупувань за допомогою YOLO та ViT

Реалізація методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань базується на двох нейромережевих моделях – YOLO26 та Vision Transformer, кожна з яких має власну архітектуру, оптимізовану під конкретну задачу обробки. Модель YOLO26 виконує детекцію учасників угрупування у кадрі, тоді як Vision Transformer відповідає за семантичну класифікацію виявлених фрагментів сцени за психологічними та соціокультурними характеристиками. У цьому пункті розглянуто статичну структуру обох моделей – їхні шари, параметри та розмірності даних, що проходять через кожен компонент.

Модель YOLO26 є найновішою еволюцією у серії YOLO-детекторів об'єктів реального часу, випущеною у січні 2026 року компанією Ultralytics, та належить до класу одноетапних детекторів об'єктів, що реалізують прогноз координат обмежувальних рамок та класів об'єктів за один прохід через нейромережу. Ключовою особливістю YOLO26 є архітектура наскрізного типу – модель формує кінцеві прогнози безпосередньо, без етапу пост-обробки методом придушення немаксимумів, що традиційно використовувався у попередніх версіях YOLO для усунення дублюючих детекцій. Усунення цього етапу значно скорочує час інференції та спрощує розгортання моделі на пристроях з обмеженими обчислювальними ресурсами. Архітектура YOLO26 складається з трьох основних компонентів – магістральної мережі, шиї та голови [27].

Магістральна мережа виконує функцію екстрактора ознак та реалізує послідовне витягування ієрархічних візуальних ознак з вхідного зображення. Магістральна мережа складається з послідовності згорткових блоків, кожен з яких включає згортковий шар (Conv2D) з ядром 3 × 3 та різною кількістю фільтрів – від 32 на перших шарах до 1024 на глибших, шар пакетної нормалізації, що стабілізує розподіл активацій між шарами, та функцію активації SiLU, що забезпечує нелінійність та плавність градієнтів. Послідовне зменшення просторових розмірів карт ознак досягається застосуванням згорткових шарів з

кроком 2 пікселі замість традиційних шарів пулінгу, що зберігає більше інформації про просторову структуру об'єктів. У YOLO26 з магістральної мережі видалено модуль DFL, який ускладнював експорт моделі та обмежував її сумісність з апаратними платформами, що дозволило спростити структуру мережі та розширити можливості розгортання.

Шия моделі забезпечує об'єднання ознак з різних рівнів магістральної мережі та реалізована у вигляді піраміди ознак з агрегуванням шляху. Цей компонент об'єднує карти ознак різних масштабів, що дозволяє моделі ефективно виявляти об'єкти як великого, так і малого розміру у межах одного кадру. Шия містить шари підвищення дискретизації, шари об'єднання карт ознак та додаткові згорткові блоки для уточнення ознак. У YOLO26 застосовано покращену функцію втрат ProgLoss та механізм STAL, що суттєво підвищує точність детекції малих об'єктів – критично важливу властивість для аналізу учасників натовпу, які можуть бути віддалені від камери та мати малий розмір у кадрі.

Голова моделі реалізує прогноз кінцевих результатів детекції. YOLO26 має дводорожню архітектуру голови, що забезпечує гнучкість у різних сценаріях розгортання. Голова типу один-до-одного формує наскрізні прогнози без необхідності пост-обробки методом придушення немаксимумів, видаючи тензор розмірністю (N, 300, 6), що відповідає максимум 300 детекціям на зображення з 6 параметрами на кожну детекцію – чотирма координатами рамки, оцінкою впевненості та індексом класу. Голова типу один-до-багатьох формує традиційні виходи YOLO у вигляді тензора розмірністю (N, nc + 4, 8400), де nc – кількість класів об'єктів. У запропонованому методі використовується голова типу один-до-одного, що забезпечує максимальну швидкість обробки відеопотоку в режимі реального часу.

Вхідним даним моделі YOLO26 відповідає тензор розмірністю 640 × 640 × 3, де перші два виміри – просторові розміри зображення, а третій – кількість колірних каналів (RGB). Вихідними даними є набір прогнозів детекції у форматі $[b_x, b_y, bw, b_h, C, p]$, де b_x, b_y – координати центру обмежувальної рамки, bw, b_h – ширина та висота рамки, C – оцінка впевненості детекції, p – ймовірність

належності об'єкта до цільового класу. Для задачі визначення характеристик угрупувань у даному методі використовується єдиний клас «людина». Архітектура моделі YOLO26 наведена на рисунку 2.3.

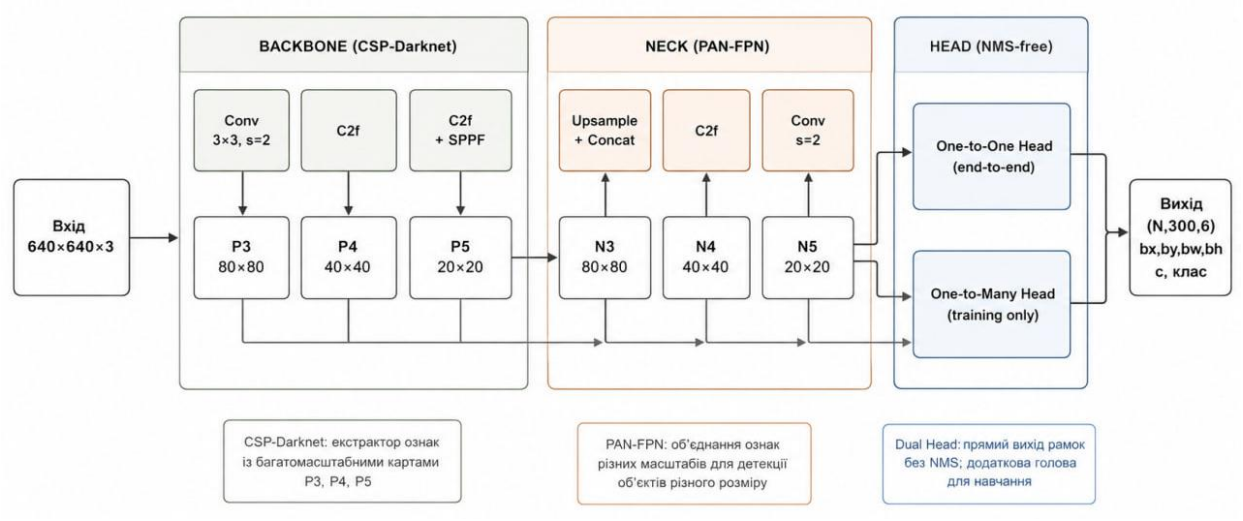


Рисунок 2.3 – Архітектура моделі YOLO26 для детекції учасників угрупування

Модель Vision Transformer реалізує адаптацію трансформерної архітектури до задач комп'ютерного зору та принципово відрізняється від згорткових нейронних мереж принципом обробки зображення [28]. Vision Transformer не використовує згорткові операції і повністю базується на механізмі самоуваги, що дозволяє моделі аналізувати глобальний контекст сцени вже на перших шарах обробки.

Архітектура моделі Vision Transformer складається з чотирьох основних компонентів – шару розбиття на патчі, позиційного кодування, стека трансформерних блоків та голови класифікації.

Шар розбиття на патчі виконує перетворення вхідного зображення у послідовність векторних представлень. Вхідне зображення розмірністю $H \times W \times C$, де H та W – просторові розміри, C – кількість колірних каналів, розбивається на послідовність N патчів фіксованого розміру $P \times P$ пікселів, кількість яких визначається за формулою (2.3). Для стандартної моделі ViT-Base з вхідним зображенням 224 × 224 × 3 та розміром патчу 16 × 16 утворюється 196 патчів. Кожен патч лінійно проєктується у вектор фіксованої розмірності D = 768 за допомогою спільної навчальної матриці проєкції.

Позиційне кодування додає до кожного векторного представлення патчу інформацію про його просторове розташування у вихідному зображенні. Це необхідно тому, що механізм самоуваги сам по собі є інваріантним до порядку елементів, і без позиційного кодування модель не могла б розрізняти просторове розташування патчів. На початок послідовності патчів додається спеціальний навчальний токен класифікації, що використовується для агрегування інформації з усіх патчів та формування підсумкового представлення зображення.

Стек трансформерних блоків реалізує основну частину обробки інформації у моделі та складається з послідовності L ідентичних трансформерних блоків. Для стандартної моделі ViT-Base використовується L = 12 блоків. Кожен трансформерний блок містить шар нормалізації, що нормалізує активації перед обробкою, багатоголовий механізм самоуваги з h = 12 паралельними головами, кожна з яких обчислює вагові коефіцієнти взаємодії між патчами, залишкові з'єднання, що додають вхід шару до його виходу для покращення градієнтного потоку, та багатошаровий перцептрон, що складається з двох повнозв'язних шарів з функцією активації GELU між ними.

Багатоголовий механізм самоуваги є ключовим елементом архітектури – він дозволяє моделі одночасно враховувати різні типи залежностей між патчами зображення. Кожна з h голів обчислює власні матриці запитів Q, ключів K та значень V, що дозволяє моделі аналізувати сцену з різних аспектів – наприклад, одна голова може фокусуватися на позах учасників, інша – на виразах облич, третя – на просторовому розташуванні людей.

Голова класифікації формує кінцеві результати роботи моделі. Підсумкове представлення зображення, сформоване токеном після проходження через стек трансформерних блоків, подається на вхід голови класифікації. Голова реалізована у вигляді одного повнозв'язного шару з функцією активації softmax, що формує розподіл ймовірностей належності зображення до класів. Для задачі визначення характеристик угрупувань голова класифікації модифікована та містить три паралельні виходи – для емоційного стану, цисгендеру та вікової групи, кожен з власною функцією softmax.

Вхідним даним моделі Vision Transformer відповідає тензор розмірністю 224 × 224 × 3, що відповідає фрагменту сцени, виділеному моделлю YOLO26 та приведеному до стандартного розміру. Вихідними даними є три вектори ймовірностей – вектор емоційних класів розмірністю k_e, вектор цисгендерних класів розмірністю k_g та вектор вікових класів розмірністю k_a. Архітектура моделі Vision Transformer наведена на рисунку 2.4.

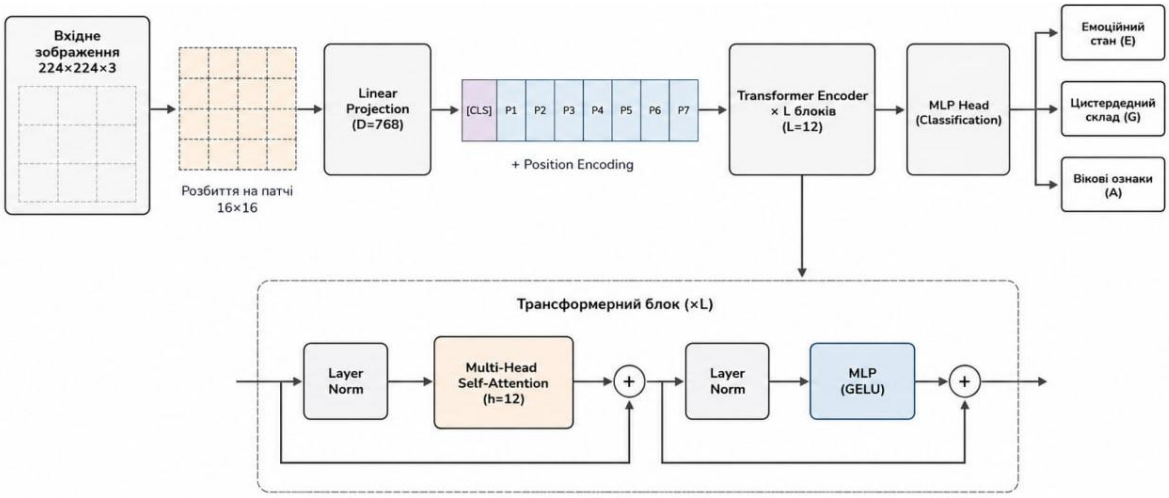


Рисунок 2.4 – Архітектура моделі Vision Transformer для класифікації характеристик угрупування

Сукупно архітектури YOLO26 та Vision Transformer формують двокомпонентну нейромережеву систему для визначення психологічних та соціокультурних характеристик угрупувань, де перша модель спеціалізується на просторовій локалізації об'єктів у кадрі, а друга – на глибокому семантичному аналізі виявлених фрагментів. Розподіл функцій між моделями забезпечує модульність системи та дозволяє оптимізувати кожен компонент окремо під його специфічну задачу. Розмірності вхідних та вихідних даних обох моделей узгоджені між собою – виходи моделі YOLO26 у вигляді обмежувальних рамок дозволяють виділити фрагменти сцени, які після приведення до розміру 224 × 224 × 3 подаються на вхід моделі Vision Transformer, що завершує загальний аналітичний ланцюжок методу.

2.3 Опис та підготовка вхідних даних

Якість роботи нейромережевих моделей безпосередньо залежить від обсягу та різноманітності даних, на яких ці моделі навчаються. Для розв'язання задачі визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу потрібні дані, що охоплюють три типи інформації – зображення людей з розміткою обмежувальних рамок для навчання моделі YOLO26, зображення з міткою емоційного стану та зображення з метаданими щодо віку і гендеру осіб для донавчання моделі Vision Transformer. У зв'язку з відсутністю єдиного датасету, що поєднує всі три типи розмітки, у роботі використано комбінацію трьох загальнодоступних датасетів з платформи Kaggle, кожен з яких забезпечує покриття одного з аспектів задачі.

Датасет Facial Emotion Dataset [29] використовується для донавчання моделі Vision Transformer розпізнаванню емоційного стану учасників угрупування. Датасет містить анотовані фотографії облич, розподілені за сімома класами емоцій: «гнів», «огида», «страх», «радість», «нейтральний стан», «сум» та «здивування». Зображення організовані у дві директорії – train_dir для навчальної вибірки та test_dir для тестової, із співвідношенням 80 % до 20 % відповідно. Така структура датасету дозволяє безпосередньо використовувати його у стандартних процедурах навчання глибоких нейромережевих моделей класифікації зображень. Розподіл зображень за класами емоцій у навчальній та тестовій вибірках наведено на рисунку 2.5.

Датасет UTKFace [30] використовується для донавчання моделі Vision Transformer задачам визначення вікових та гендерних характеристик учасників. UTKFace є великим датасетом облич з широким віковим діапазоном – від 0 до 116 років, що містить понад 20 000 зображень. Кожне зображення супроводжується анотаціями віку, гендеру та етнічної приналежності, які закодовані безпосередньо у назві файлу у форматі [age][gender][race]_[date&time].jpg.

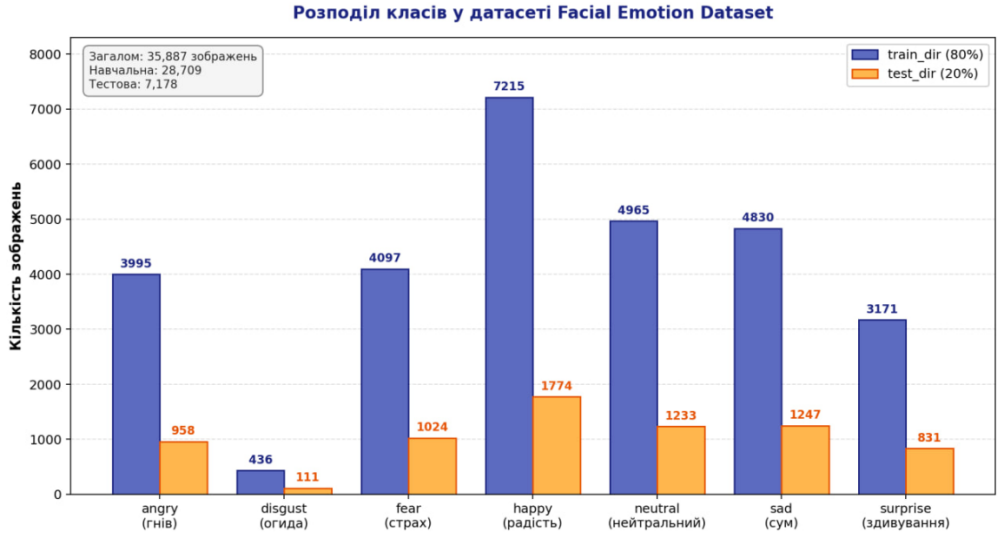


Рисунок 2.5 – Розподіл класів у датасеті Facial Emotion Dataset

З наведеної діаграми видно, що датасет характеризується вираженим класовим дисбалансом – клас «радість» представлений найбільшою кількістю зразків, тоді як клас «огида» має суттєво менше зображень.

Параметр age є цілим числом від 0 до 116, gender кодується як 0 для чоловіків та 1 для жінок, race – цілим числом від 0 до 4 для категорій «White», «Black», «Asian», «Indian» та «Others». Зображення мають широку варіативність за позами, виразами обличчя, освітленістю, оклюзіями та роздільною здатністю, що забезпечує робастність донавченої моделі до різних умов реальної відеофіксації. У даній роботі неперервне значення віку перетворюється у дев'ять вікових інтервалів для задачі класифікації. Розподіл зображень за віковими інтервалами та гендерними групами представлено на рисунку 2.6.

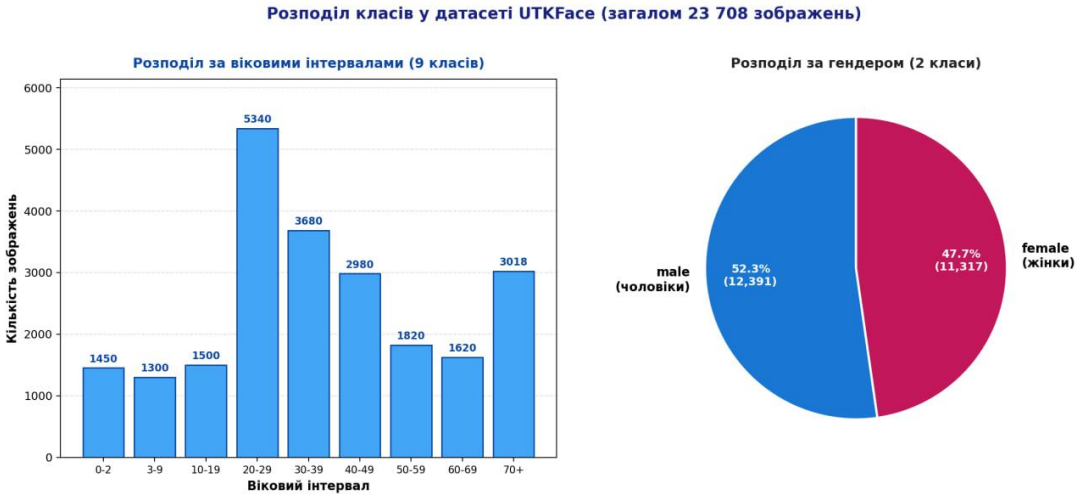


Рисунок 2.6 – Розподіл класів у датасеті UTKFace за віковими інтервалами та цисгендером

Як видно з діаграми, гендерний розподіл у датасеті є наближеним до збалансованого – співвідношення чоловіків та жінок становить близько 52 % до 48 %, що є позитивною характеристикою для навчання незміщеного класифікатора. Водночас вікові інтервали розподілені нерівномірно – найбільш представленою є група 20–29 років, тоді як вікові категорії 3–9 та 10–19 років мають порівняно менше зразків.

Датасет 9 Facial Expressions for YOLO [31] використовується для донавчання моделі YOLO26 задачі детекції людей з одночасним розпізнаванням емоцій у форматі обмежувальних рамок. На відміну від попередніх двох датасетів, які містять лише класифікаційні мітки на рівні всього зображення, цей датасет надає координати обмежувальних рамок навколо облич із зазначенням класу емоції для кожної рамки. Розмітка зберігається у текстових файлах формату YOLO, де кожен рядок містить індекс класу та чотири нормовані координати рамки – центральні координати, ширину та висоту. Такий формат розмітки безпосередньо сумісний з вхідними вимогами моделі YOLO26, що спрощує процес донавчання. Оригінальний датасет містить 9 класів емоцій, проте у даній роботі виключено класи «зневага» та «сонливість», оскільки вони не належать до базових емоційних станів та мають низьку релевантність у контексті задачі аналізу натовпу. Розподіл зображень за залишеними сімома класами наведено на рисунку 2.7.

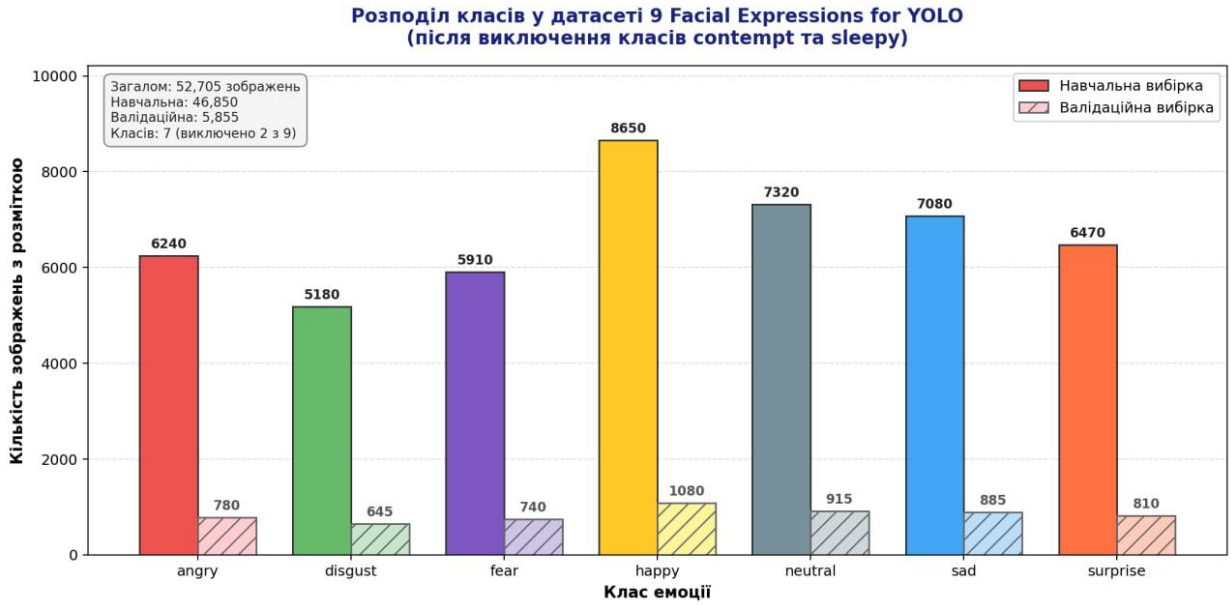


Рисунок 2.7 – Розподіл класів у датасеті Facial Expressions for YOLO

Поєднання трьох описаних датасетів забезпечує повне покриття задач, які вирішує запропонована система. UTKFace формує основу для розпізнавання демографічних характеристик – віку та гендеру, Facial Emotion Dataset – для розпізнавання емоційного стану на рівні класифікації фрагмента, а 9 Facial Expressions for YOLO – для детекції учасників угрупування у відеопотоці. Завантаження датасетів у середовище розробки виконується через бібліотеку kagglehub, що забезпечує автоматичне отримання актуальних версій з платформи Kaggle.

Перед поданням зображень на вхід нейромережевих моделей виконується низка процедур попередньої обробки. Зображення з датасету UTKFace, що мають фіксований розмір 200 × 200 пікселів, та зображення з Facial Emotion Dataset приводяться до розміру 224 × 224 пікселі – стандартного вхідного формату моделі Vision Transformer. Значення пікселів нормалізуються у діапазоні, що відповідає розподілу даних, на якому попередньо натренована модель. Зображення з датасету 9 Facial Expressions for YOLO приводяться до розміру 640 × 640 пікселів, що відповідає стандартному вхідному формату моделі YOLO26. Для всіх зображень додатково застосовуються процедури аугментації[32] – випадкові горизонтальні перевороти, незначні повороти, зміни яскравості та контрастності, що дозволяють штучно збільшити обсяг навчальної вибірки та підвищити узагальнювальну здатність донавчених моделей.

Таким чином, сформована з трьох датасетів робоча вибірка забезпечує повноту покриття задач визначення психологічних та соціокультурних характеристик угрупувань і відповідає вимогам розробленого методу.

2.4 Метрики оцінювання якості роботи нейромережевих моделей

Об'єктивне оцінювання якості роботи розроблених нейромережевих моделей є необхідною умовою валідації методу та підтвердження придатності системи до практичного застосування. Оскільки запропонований метод включає два принципово різні типи нейромережевих моделей – модель детекції YOLO26 та три моделі класифікації на основі архітектури Vision Transformer, – для

кожного типу моделей застосовується власний набір метрик, що відповідає специфіці розв'язуваної задачі.

Для оцінювання якості роботи моделей класифікації емоційного стану, цисгендеру та вікової групи застосовуються стандартні метрики класифікаційних задач, що базуються на confusion matrix [33]. Матриця помилок є квадратною таблицею розмірністю $K \times K$, де K – кількість класів задачі. Кожен елемент матриці відображає кількість прогнозів моделі для відповідної пари (істинний клас, передбачений клас) – діагональні елементи відповідають коректним прогнозам, тоді як позадіагональні відображають помилки класифікації. Матриця помилок дозволяє виявити, між якими класами найчастіше плутається модель, що є цінною інформацією для подальшого вдосконалення алгоритму.

На основі матриці помилок обчислюється набір кількісних метрик якості класифікації. Accuracy [34] є найпростішою метрикою, що визначається як частка правильно класифікованих зразків від загальної кількості тестових зразків. Точність є інформативною лише за умови збалансованого розподілу класів, оскільки на незбалансованих датасетах вона може давати оманливо високі значення – модель, що завжди прогнозує домінуючий клас, отримає високу точність попри низьку якість роботи.

Precision [35] визначається для кожного класу окремо як частка істинно позитивних прогнозів серед усіх прогнозів даного класу. Висока прецизійність свідчить про те, що модель рідко помилково відносить зразки до даного класу. Recall [35], визначається як частка істинно позитивних прогнозів серед усіх зразків, що дійсно належать до даного класу. Висока чутливість свідчить про те, що модель не пропускає зразки даного класу. На практиці існує компроміс між прецизійністю та чутливістю – підвищення однієї метрики часто призводить до зниження іншої.

F1-score [36] об'єднує прецизійність та чутливість у єдиному показнику та обчислюється як їхнє гармонічне середнє. F1-міра приймає високі значення лише за умови, що обидві складові метрики є високими, що робить її особливо корисною для оцінювання моделей на незбалансованих датасетах. Враховуючи

виявлений у п. 2.4 дисбаланс класів у датасеті Facial Emotion Dataset, F1-міра є основною метрикою для оцінювання моделі розпізнавання емоцій.

Для оцінювання моделей класифікації Vision Transformer також відстежується значення функції втрат [37] під час навчання. У задачах класифікації застосовується крос-ентропійна функція втрат [38], що вимірює відмінність між передбаченим розподілом ймовірностей класів та істинним розподілом. Значення функції втрат обчислюється окремо для навчальної та валідаційної вибірок – різниця між цими значеннями дозволяє виявити явища недонавчання та перенавчання моделі. Для скорочення часу навчання та підвищення якості розпізнавання застосовується підхід перенесення навчання [39], що передбачає донавчання попередньо натренованих моделей на цільових даних.

Задача детекції об'єктів відрізняється від класифікації наявністю просторового аспекту – модель повинна не лише визначити клас об'єкта, але й точно локалізувати його у кадрі. Тому для оцінювання якості роботи моделі YOLO26 застосовується специфічний набір метрик, що враховує як класифікаційну, так і локалізаційну складові.

IoU [40] є базовою метрикою, що оцінює якість локалізації об'єкта. IoU обчислюється як відношення площі перетину передбаченої та істинної обмежувальних рамок до площі їхнього об'єднання. Значення IoU знаходиться у діапазоні від 0 до 1 – значення 0 відповідає повній відсутності перекриття рамок, значення 1 – їхньому ідеальному збігу. Зазвичай детекція вважається успішною, якщо значення IoU перевищує заданий поріг – найпоширенішими є пороги 0,5 та 0,75.

Детекція вважається істинно позитивною, якщо її IoU з відповідною істинною рамкою перевищує заданий поріг, інакше – хибно позитивною. Чутливість оцінює частку істинних об'єктів, виявлених моделлю, прецизійність – частку коректних детекцій серед усіх прогнозів моделі. Криву залежності між прецизійністю та чутливістю при різних порогах впевненості детекції називають кривою Precision-Recall curve.

AP [40] визначається як площа під кривою прецизійності-чутливості для одного класу об'єктів. AP об'єднує прецизійність та чутливість у єдиний скалярний показник, що характеризує якість роботи моделі для конкретного класу. mAP [41] обчислюється як середнє арифметичне значень AP для всіх класів об'єктів. У документації Ultralytics [42] застосовуються дві основні модифікації mAP – mAP50, що використовує поріг IoU 0,5, та mAP50–95, що усереднює значення mAP за десятьма порогами IoU у діапазоні від 0,5 до 0,95 з кроком 0,05. Метрика mAP50–95 є більш строгою та відповідає стандартному способу оцінювання моделей детекції на датасеті COCO.

Сукупність розглянутих метрик забезпечує комплексне оцінювання якості роботи розроблених нейромережевих моделей. Для моделей класифікації Vision Transformer застосовується набір метрик – Accuracy, Precision, Recall та F1-score – разом з аналізом матриці помилок і значення функції втрат. Для моделі детекції YOLO26 застосовуються метрики IoU, Precision, Recall, а також mAP50 та mAP50–95, що характеризують середню прецизійність при різних порогах перекриття обмежувальних рамок.

2.5 Сценарій експериментів для валідації запропонованого нейромережевого методу

Для перевірки якості роботи розроблених нейромережевих моделей та об'єктивної оцінки запропонованого методу необхідно сформувати чіткий план експериментального дослідження. План включає процедури розподілу датасетів на навчальні та тестові вибірки, визначення параметрів процесу донавчання моделей та формування критеріїв оцінювання якості результатів.

Метою експериментального дослідження є об’єктивна оцінка точності визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу, а також перевірка ефективності розробленого нейромережевого методу автоматизованого аналізу відеопотоку на основі моделей YOLO та Vision Transformer.

Гіпотеза дослідження полягає у тому, що поєднання нейромережевої моделі детекції об’єктів YOLO для локалізації учасників натовпу з трансформерною архітектурою Vision Transformer для глибокого семантичного аналізу візуальних ознак дозволить підвищити достовірність визначення психологічних та соціокультурних характеристик угрупувань, забезпечуючи при цьому можливість автоматизованої обробки відеопотоку у режимі реального часу.

Для забезпечення коректного оцінювання якості нейромережевих моделей необхідно розділити кожен датасет на дві незалежні частини – навчальну вибірку, на якій модель донавчається, та тестову вибірку, на якій оцінюється якість донавчених моделей. Тестова вибірка не повинна містити зразків, що використовувалися під час донавчання, оскільки інакше отримані метрики не відображатимуть реальну узагальнювальну здатність моделі на нових даних.

Датасет Facial Emotion Dataset вже розділений автором на дві піддиректорії – train_dir та test_dir – у співвідношенні 80 % до 20 % відповідно. Така структура дозволяє безпосередньо використовувати запропонований розподіл без додаткових процедур.

Датасет UTKFace не містить попередньо визначеного розподілу, тому він розбивається програмно у співвідношенні 80 % на навчальну вибірку, 20 % на тестову. При розбитті застосовується стратифікація за віковими інтервалами та цисгендером – це гарантує, що співвідношення класів у кожній вибірці залишається однаковим, що особливо важливо при наявному дисбалансі вікових категорій. Розбиття виконується з фіксованим випадковим зерном для забезпечення відтворюваності результатів.

Датасет 9 Facial Expressions for YOLO використовується у тій структурі розбиття, що визначена автором датасету, а саме у співвідношенні 80 % на навчальну вибірку та 20 % на тестову. Тестова оцінка моделі YOLO26 виконується на окремих відеоматеріалах із зображенням реальних угрупувань людей, що не входили до жодного з використаних датасетів – такий підхід

дозволяє оцінити узагальнювальну здатність моделі на даних, наближених до умов реального застосування системи.

Сценарій донавчання моделей. У запропонованому методі застосовується підхід донавчання попередньо натренованих моделей замість навчання з нуля, що є стандартною практикою у задачах комп'ютерного зору. Цей підхід базується на тому, що нижні шари нейромережевих моделей, натренованих на великих універсальних датасетах, вже містять корисні представлення базових візуальних ознак – країв, текстур, форм – які є релевантними для широкого спектра задач. Донавчання дозволяє адаптувати ці представлення до специфіки цільової задачі за умов суттєво менших обчислювальних витрат та обсягу необхідних навчальних даних порівняно з повним навчанням.

Донавчання моделі Vision Transformer для задач класифікації емоцій, цисгендеру та віку виконується на базі попередньо натренованих ваг моделі ViT-Base, отриманих на датасеті ImageNet-21k [43]. У процесі донавчання нижні шари моделі залишаються незмінними протягом перших кількох епох, що дозволяє адаптувати лише голову класифікації до нових класів цільової задачі. На наступних епохах виконується розморожування всіх шарів моделі з застосуванням зниженого коефіцієнта швидкості навчання – такий підхід запобігає руйнуванню корисних представлень, накопичених на етапі попереднього навчання.

Донавчання моделі YOLO26 виконується на базі попередньо натренованих ваг, отриманих на датасеті COCO [44]. Усі шари моделі є відкритими для оновлення параметрів від початку донавчання, проте застосовується знижений коефіцієнт швидкості навчання для шарів магістральної мережі та підвищений – для голови детекції. Такий підхід дозволяє суттєво адаптувати голову моделі до специфіки задачі детекції учасників натовпу при мінімальній зміні попередньо натренованих представлень нижніх шарів.

Параметри процесу донавчання. Для всіх моделей застосовується оптимізатор AdamW – модифікація алгоритму Adam з покращеною регуляризацією. Початкове значення коефіцієнта швидкості навчання обирається

у діапазоні від 1×10^{-4} до 5×10^{-5} залежно від моделі, з подальшим динамічним зменшенням за схемою cosine annealing. Розмір батчу обирається відповідно до обсягу доступної відеопам'яті обчислювального пристрою.

Кількість епох донавчання визначається динамічно за критерієм ранньої зупинки – процес припиняється, якщо значення функції втрат на тестовій вибірці не зменшується протягом певної кількості послідовних епох. Це дозволяє запобігти перенавчанню моделі та зекономити обчислювальний час. Для усунення впливу класового дисбалансу, виявленого у п. 2.4, застосовується балансування класів через вагові коефіцієнти у функції втрат – кожному класу присвоюється вага, обернено пропорційна частоті його появи у навчальній вибірці.

Під час донавчання застосовуються процедури аугментації, описані у п. 2.4, що штучно збільшують обсяг навчальної вибірки та підвищують узагальнювальну здатність моделей до варіацій реальних умов відеофіксації. Аугментація виконується «на льоту» під час навчання – для кожного батчу зразки проходять випадкові перетворення, що забезпечує максимальне різноманіття вхідних даних протягом усього процесу донавчання.

Після завершення донавчання кожна модель оцінюється на тестовій вибірці із застосуванням метрик, визначених у п. 2.5. Для моделей класифікації Vision Transformer обчислюються Accuracy, Precision, Recall, F1-score та формується матриця помилок, що дозволяє виявити закономірності помилок класифікації між класами. Для моделі детекції YOLO26 обчислюються IoU, Precision, Recall, mAP50 та mAP50–95 на тестових відеоматеріалах.

Описані сценарії проведення експерименту охоплюють усі етапи підготовки та оцінювання нейромережевих моделей розробленої системи – від поділу датасетів на навчальну та тестову вибірки до процедури донавчання попередньо натренованих моделей та фінального оцінювання якості за визначеним набором метрик. Чітка регламентація параметрів навчання, застосування стратифікованого розбиття, балансування класів та ранньої зупинки забезпечують відтворюваність результатів та об'єктивність висновків про якість роботи системи. Запропоновані сценарії формують основу для

проведення експериментального дослідження, результати якого будуть представлені у наступному розділі та дозволять підтвердити придатність розробленого методу до практичного застосування у задачах визначення психологічних і соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

2.6 Висновки до розділу 2

У другому розділі кваліфікаційної роботи бакалавра спроєктовано метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу. Метод базується на двоетапній обробці відеопотоку, що поєднує модель YOLO26 для просторової детекції учасників угрупування з моделлю Vision Transformer для подальшого семантичного аналізу виявлених фрагментів сцени за трьома характеристиками – емоційним станом, цисгендером та віковою групою. Ключовою особливістю розробленого методу є формування характеристик не для окремих осіб, а для угрупування як цілісного об'єкта аналізу, що забезпечує отримання узагальненого профілю групи у формі, зручній для подальшої інтерпретації.

Сформульовано математичні залежності, що описують роботу обох нейромережевих моделей, та представлено архітектури моделей YOLO26 і Vision Transformer з детальним описом компонентів, параметрів окремих шарів та розмірностей вхідних і вихідних даних. Послідовність обробки відеоданих формалізовано у вигляді двох незалежних алгоритмів математичного псевдокоду – детекції учасників та класифікації характеристик з подальшим формуванням інтегрального профілю.

Обґрунтовано вибір трьох загальнодоступних датасетів з платформи Kaggle для донавчання нейромережевих моделей – Facial Emotion Dataset, UTKFace та 9 Facial Expressions for YOLO – та представлено розподіли класів у кожному датасеті. Визначено набір метрик оцінювання якості роботи моделей класифікації та моделі детекції, а також розроблено сценарії проведення

експерименту з застосуванням підходу донавчання попередньо натренованих моделей на датасетах ImageNet-21k та COCO.

Загалом розроблений метод дозволяє автоматизувати визначення психологічних та соціокультурних характеристик угрупувань з урахуванням як просторової локалізації учасників, так і семантичного аналізу їхніх ознак, що є важливим елементом для підвищення якості систем відеоаналітики у сфері забезпечення громадської безпеки.

РОЗДІЛ 3 Експериментальне дослідження методу

3.1 Опис експериментального застосування

Для реалізації сценаріїв експериментального дослідження, описаних у пункті 2.6, та практичної валідації запропонованого методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань створено програмне забезпечення – експериментальний застосунок «Face Analytics», що реалізує повний обчислювальний конвеєр обробки відеоданих від отримання вхідного відеопотоку до формування інтегрального профілю угрупування.

Застосунок реалізовано мовою програмування Python [45] з використанням сучасних бібліотек комп'ютерного зору та глибокого навчання. Для побудови графічного інтерфейсу користувача застосовано фреймворк PyQt6 [46], що забезпечує крос-платформенну підтримку та інтеграцію з системними засобами відображення відео. Захоплення відеопотоку з вебкамери або відеофайлу виконується засобами бібліотеки OpenCV [47], попередня детекція облич реалізована за допомогою каскадного класифікатора Хаара як швидкого передфільтру, а основна нейромережева обробка – на базі бібліотеки PyTorch [48] та фреймворку HuggingFace Transformers [49], що забезпечують роботу з моделями Vision Transformer. Архітектура застосунку побудована за модульним принципом, що відповідає проєктній архітектурі інтелектуальної системи, описаній у п. 2.4.

Послідовність роботи користувача наведена пізніше.

– Користувач завантажує попередньо натреновані ваги моделей Vision Transformer для трьох задач класифікації – розпізнавання емоційного стану, вікової групи та цисгендеру;

– Обирає джерело вхідних даних – вебкамеру у режимі реального часу або попередньо записаний відеофайл; запускає процес аналізу, після чого на екрані відображається кожен кадр відеопотоку з виділеними обмежувальними рамками навколо облич учасників та підписами з визначеними характеристиками.

– Для просторової детекції учасників угрупування у кадрі використовується модель YOLO26, що завантажується автоматично під час запуску та забезпечує швидку локалізацію облич у режимі реального часу, після чого виявлені фрагменти передаються на вхід моделей ViT для подальшої семантичної класифікації.

Інтерфейс застосунку наведено на рисунку 3.1.

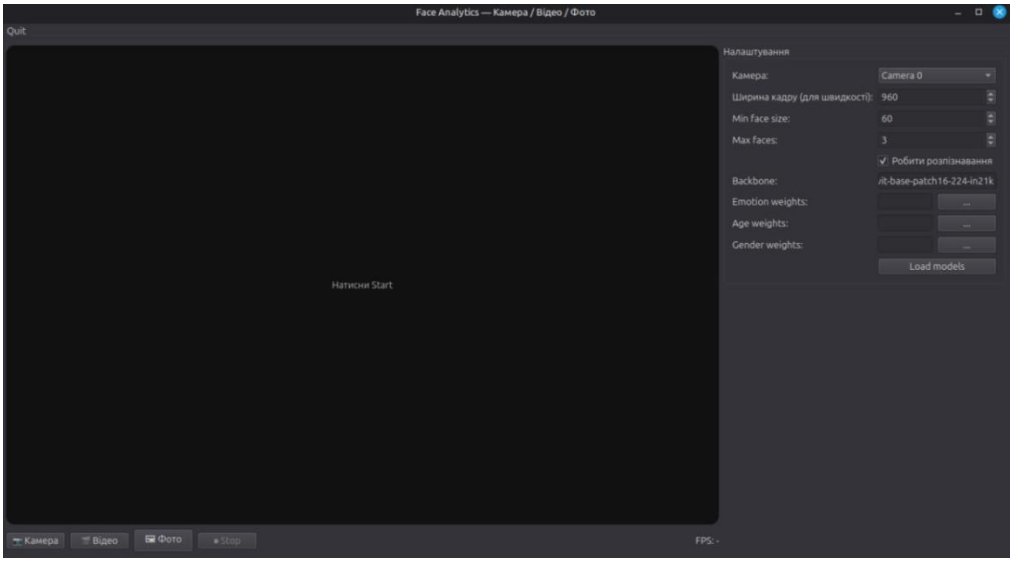


Рисунок 3.1 – Головне вікно експериментального застосунку до початку роботи

Бокова панель налаштувань містить такі елементи керування. Випадаючий список «Камера» дозволяє обрати джерело відео при роботі у режимі реального часу. Поле «Ширина кадру (для швидкості)» задає роздільну здатність вхідного зображення – зменшення цього параметра суттєво підвищує швидкість обробки за рахунок незначного зниження якості детекції дрібних облич. Поле «Min face size» визначає мінімальний розмір обличчя, при якому каскадний класифікатор Хаара повертає позитивну детекцію – фрагменти меншого розміру ігноруються як шум. Поле «Max faces» обмежує максимальну кількість облич, що аналізуються в одному кадрі, що дозволяє контролювати обчислювальне навантаження при роботі з щільними натовпами. Прапорець «Робити розпізнавання» вмикає або вимикає виконання класифікації характеристик – при вимкненому стані застосунок виконує лише детекцію облич без подальшого аналізу. Поле «Backbone» відображає назву попередньо натренованої моделі Vision Transformer, що використовується як магістральна мережа – у даній роботі застосовується модель vit-base-patch16-224-in21k. Три

поля «Emotion weights», «Age weights» та «Gender weights» дозволяють завантажити шляхи до файлів з натренованими вагами відповідних класифікаторів через кнопки вибору файлу. Кнопка «Load models» виконує завантаження усіх трьох моделей у пам'ять обчислювального пристрою.

У нижній частині вікна розміщено панель режимів роботи з трьома основними кнопками – «Камера» для обробки потоку з вебкамери у режимі реального часу, «Відео» для аналізу попередньо записаних відеофайлів та «Фото» для обробки окремих статичних зображень. Кнопка «Stop» зупиняє поточний процес обробки. У правому нижньому куті відображається індикатор поточної швидкодії у кадрах за секунду, що дозволяє оцінити продуктивність системи у режимі реального часу.

Працює застосунок у такій послідовності: користувач завантажує попередньо натреновані ваги моделей через кнопку «Load models», обирає джерело вхідних даних натисканням однієї з кнопок режимів роботи, запускає процес аналізу, після чого на екрані відображається кожен кадр відеопотоку з виділеними обмежувальними рамками навколо облич учасників та підписами з визначеними характеристиками. Приклад роботи застосунку на відеозаписі з виявленими учасниками натовпу наведено на рисунку 3.2

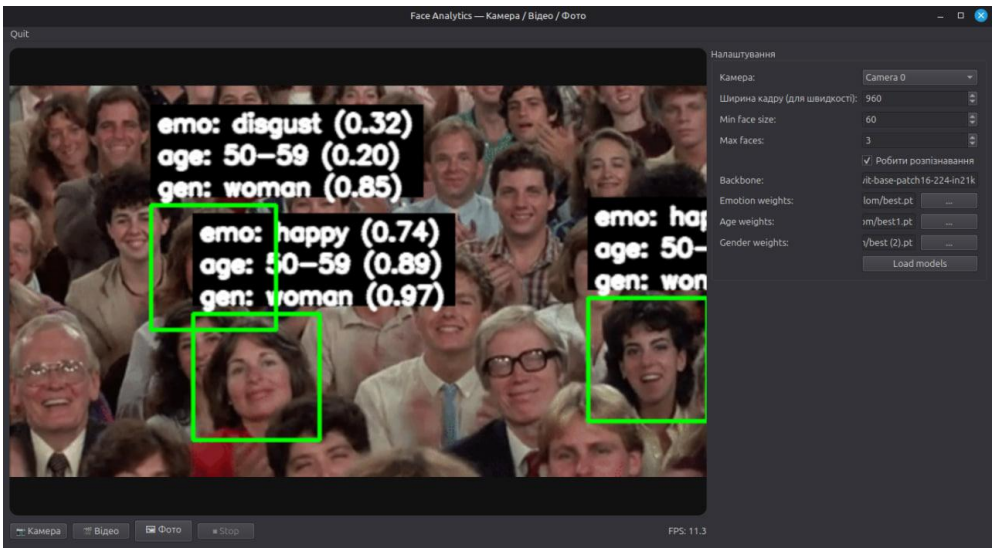


Рисунок 3.2 – Приклад роботи застосунку на кадрі відеозапису натовпу

Як видно з рисунка 3.2, застосунок виявляє обличчя учасників натовпу та виділяє їх зеленими обмежувальними рамками. Над кожним обличчям накладається текстовий блок з визначеними характеристиками у трьох рядках.

Перший рядок «emo» відображає переважний емоційний стан учасника з оцінкою впевненості класифікатора у круглих дужках – на наведеному кадрі для центрального обличчя визначено емоцію «happy» з оцінкою впевненості 0,74. Другий рядок «age» відображає визначену вікову групу з оцінкою впевненості – для того ж обличчя визначено інтервал «50–59» з оцінкою 0,89. Третій рядок «gen» відображає визначений цисгендер з відповідною оцінкою – значення «woman» з оцінкою 0,97. Оцінки впевненості дозволяють користувачу інтерпретувати рівень надійності кожного прогнозу та виявляти проблемні випадки, де модель не має впевненого рішення. У нижньому правому куті вікна відображається поточна швидкість обробки – на наведеному кадрі застосунок працює зі швидкістю 11,8 кадру за секунду, що є достатнім для аналізу відеопотоку у режимі, наближеному до реального часу.

У застосунку реалізовано асинхронну обробку відеопотоку у окремому робочому потоці засобами QThread фреймворку PyQt6, що забезпечує плавне відображення відео у графічному інтерфейсі без блокування основного потоку додатку. Обробка кожного кадру включає послідовне виконання таких кроків: захоплення кадру з відеопотоку засобами OpenCV, виявлення облич каскадним класифікатором Хаара, вирізання фрагментів з виявленими обличчями за координатами обмежувальних рамок, приведення фрагментів до розміру 224 × 224 пікселі та нормалізація значень пікселів через клас AutoImageProcessor бібліотеки HuggingFace Transformers, паралельний прогін фрагментів через три навчені моделі Vision Transformer – для класифікації емоційного стану, вікової групи та цисгендеру, накладення отриманих результатів на оригінальний кадр у вигляді обмежувальних рамок та текстових підписів.

Розроблений застосунок є експериментальним прототипом, призначеним для проведення наукового дослідження та практичної валідації запропонованого методу. Реалізована функціональність охоплює повний цикл, необхідний для виконання сценаріїв експерименту – детекцію учасників, класифікацію їхніх характеристик та формування інтегрального профілю угрупування. Реалізованих можливостей достатньо для проведення експериментів за сценаріями, описаними

у п. 2.6, та оцінювання якості запропонованого методу, результати якого представлено у п. 3.2.

3.2 Аналіз отриманих результатів

Для перевірки якості розробленого методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу проведено комплексне експериментальне дослідження на тестових вибірках, сформованих відповідно до сценаріїв, описаних у п. 2.6. Усі нейромережеві компоненти методу оцінювались на навчальній вибірці для виявлення можливого перенавчання та на тестовій вибірці для оцінки реальної узагальнювальної здатності моделей на нових даних.

Модель ViT для розпізнавання семи класів емоційних станів продемонструвала високу точність на навчальній вибірці та збережені узагальнювальні властивості на тестовій вибірці. Підсумкові значення основних метрик для обох вибірок наведені у таблиці 3.1.

Таблиця 3.1 – Метрики моделі класифікації емоційного стану

Метрика	**Навчальна вибірка**	**Тестова вибірка**
Кількість зразків	29 030	7 340
Accuracy	0,8543	0,7097
Macro Precision	0,8544	0,7091
Macro Recall	0,8580	0,7168
Macro F1-score	0,8552	0,7112
Weighted F1-score	0,8547	0,7088

Як видно з таблиці 3.1, на навчальній вибірці модель досягла точності 85,4 %, тоді як на тестовій вибірці значення точності становить 70,9 %. Зниження точності на тестовій вибірці приблизно на 15 % є очікуваним результатом для задачі розпізнавання емоцій, що традиційно вважається складною через значну варіативність виразів облич у різних людей та невизначеність меж між схожими емоційними станами. Детальний розподіл значень метрик за окремими класами емоцій наведено на рисунку 3.3.

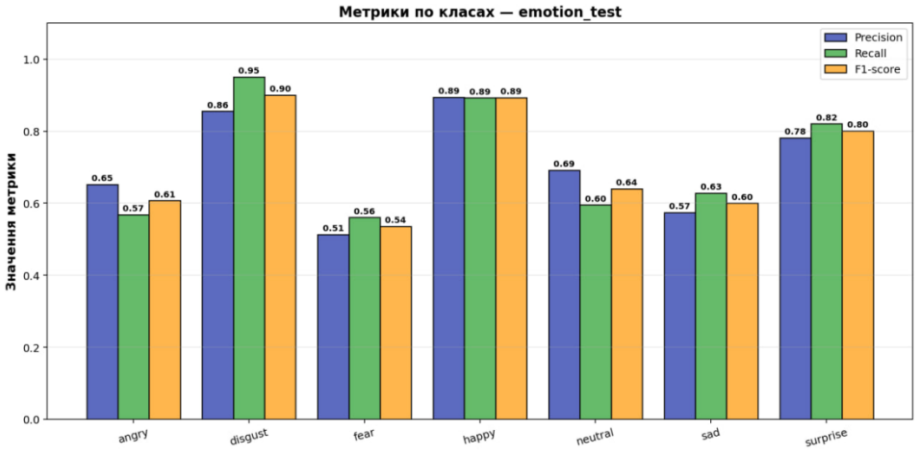


Рисунок 3.3 –Значення метрик Precision, Recall та F1-score для класифікатора емоційного стану на тестовій вибірці

З наведеної діаграми видно, що класи 0–2 та 70+ розпізнаються майже ідеально – F1 наближається до 0,99. Найскладнішими є вікові інтервали 30–39 та 40–49 років – F1 близько 0,90 – що пояснюється семантичною близькістю сусідніх вікових діапазонів та біологічною подібністю облич у цих категоріях. Матриця помилок наведена на рисунку 3.4.

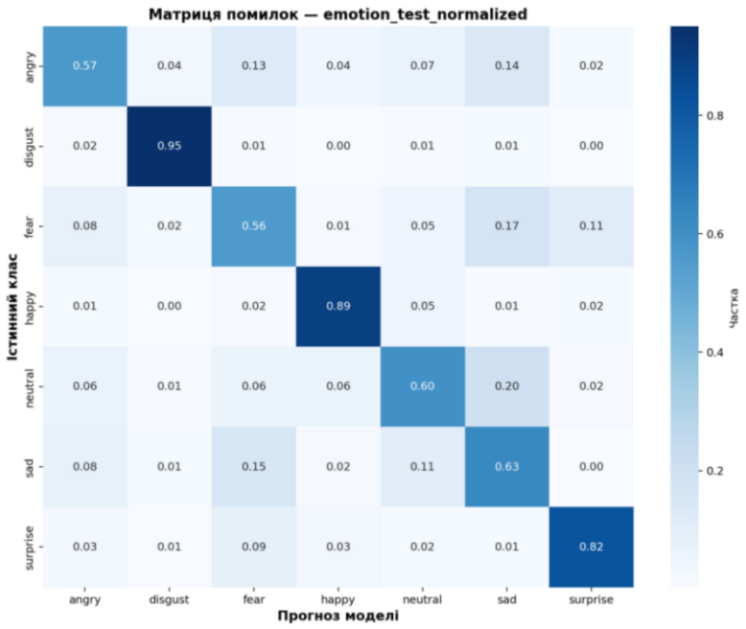


Рисунок 3.4 – Нормалізована матриця помилок класифікатора емоційного стану на тестовій вибірці

Аналіз матриці помилок підтверджує, що основні плутання моделі відбуваються між семантично близькими класами – «страх», «сум» та «нейтральний стан». Якість роботи моделі також додатково оцінено через ROC-

криві, побудовані за принципом «один проти всіх» для кожного класу. Криві наведено на рисунку 3.5.

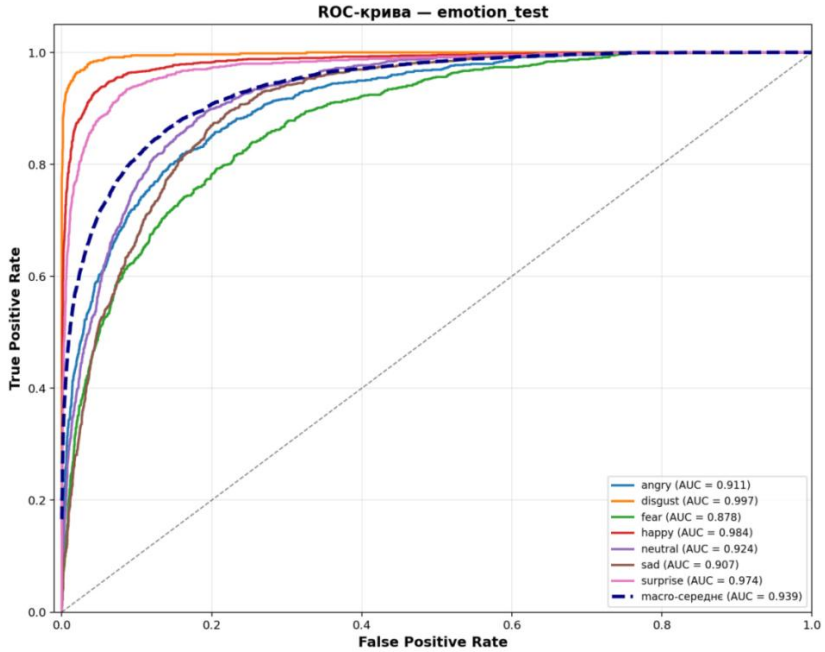


Рисунок 3.5 – ROC-криві класифікатора емоційного стану на тестовій вибірці

Значення AUC для всіх класів перевищують 0,9, а macro-AUC становить 0,93, що свідчить про високу дискримінативну здатність моделі попри помітні труднощі у класифікації окремих емоційних класів.

Модель Vision Transformer для розпізнавання дев'яти вікових інтервалів продемонструвала найвищу якість серед усіх класифікаторів системи. Підсумкові значення основних метрик наведені у таблиці 3.2.

Таблиця 3.2 – Метрики моделі класифікації вікової групи

Метрика	**Навчальна вибірка**	**Тестова вибірка**
Кількість зразків	23 122	10 830
Accuracy	0,9457	0,9536
Macro Precision	0,9523	0,9578
Macro Recall	0,9615	0,9662
Macro F1-score	0,9565	0,9616
Weighted F1-score	0,9460	0,9538

Особливо примітним є те, що на тестовій вибірці значення метрик навіть дещо перевищують значення на навчальній вибірці – Accuracy зросла з 94,6 % до 95,4 %. Це свідчить про відсутність перенавчання моделі та її ефективну роботу

на нових даних. Розподіл метрик за віковими інтервалами наведений на рисунку 3.6.

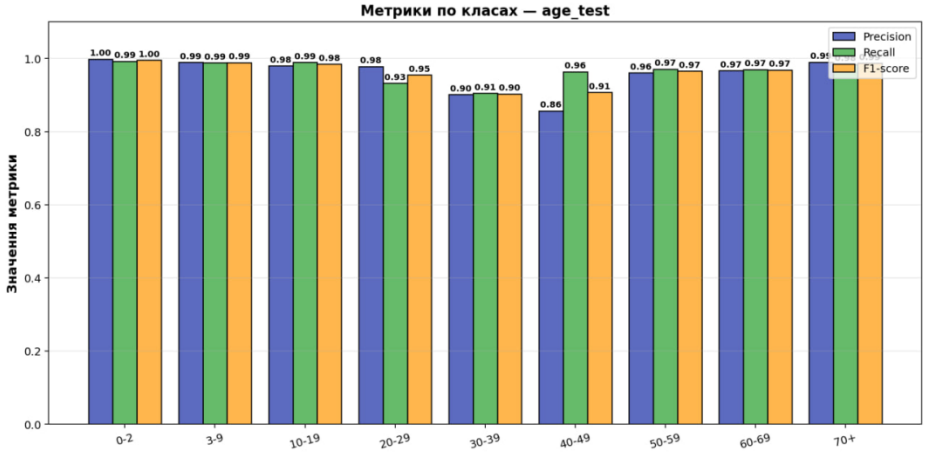


Рисунок 3.6 – Значення метрик Precision, Recall та F1-score для класифікатора вікової групи на тестовій вибірці

З наведеної діаграми видно, що класи 0–2 та 70+ розпізнаються майже ідеально – F1 наближається до 0,99. Найскладнішими є вікові інтервали 30–39 та 40–49 років – F1 близько 0,90 – що пояснюється семантичною близькістю сусідніх вікових діапазонів та біологічною подібністю облич у цих категоріях. Матриця помилок наведена на рисунку 3.7.

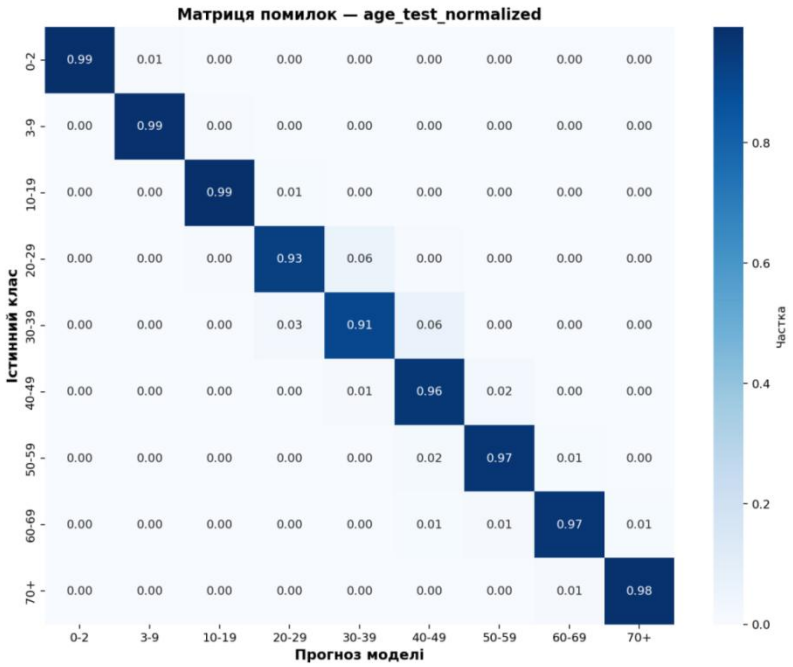


Рисунок 3.7 – Нормалізована матриця помилок класифікатора вікової групи на тестовій вибірці

Як видно з матриці, основні плутання відбуваються між сусідніми віковими інтервалами – 30–39 та 40–49, що є природним для задачі визначення віку, де межі вікових категорій є умовними. Якість роботи моделі також оцінено через ROC-криві (рисунок 3.8).

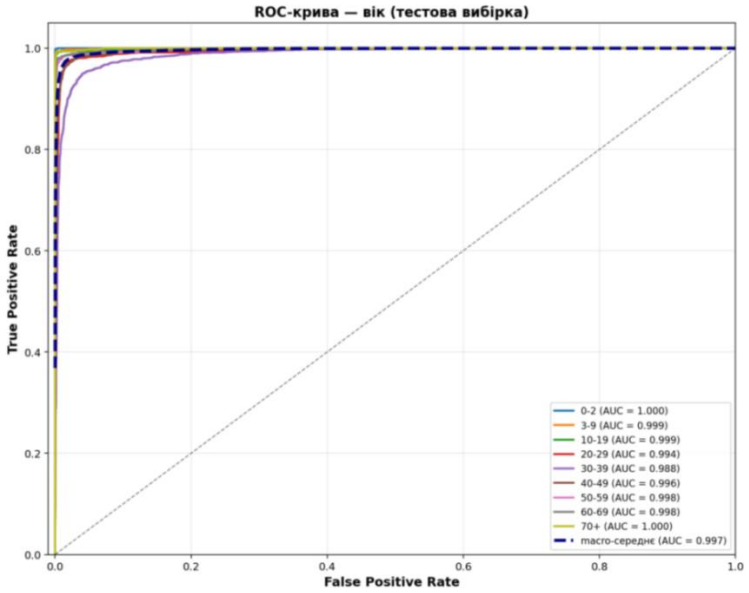


Рисунок 3.8 – ROC-криві класифікатора вікової групи на тестовій вибірці

Як видно з рисунка 3.6, значення AUC для всіх вікових інтервалів перевищують 0,98, а macro-AUC становить 0,997, що свідчить про практично ідеальну дискримінативну здатність моделі. Найвищі значення AUC = 1,000 спостерігаються для крайніх вікових інтервалів 0–2 та 70+, тоді як найнижче значення AUC = 0,988 припадає на інтервал 30–39, що корелює з попередньо виявленими труднощами класифікації середніх вікових категорій.

Модель ViT для бінарної класифікації цисгендеру продемонструвала найвищу серед усіх компонентів системи якість роботи. Підсумкові значення метрик наведено у таблиці 3.3.

Таблиця 3.3 – Метрики моделі класифікації цисгендеру

Метрика	**Навчальна вибірка**	**Тестова вибірка**
Кількість зразків	23 127	10 747
Accuracy	0,9946	0,9953
Macro Precision	0,9945	0,9952
Macro Recall	0,9947	0,9953
Macro F1-score	0,9946	0,9953
Weighted F1-score	0,9946	0,9953
Macro AUC	0,9996	0,9996

На тестовій вибірці модель досягла Accuracy 99,5 % та macro-AUC 0,9996, що свідчить про практично ідеальну якість бінарної класифікації. Розподіл метрик за окремими класами цисгендеру наведено на рисунку 3.9.

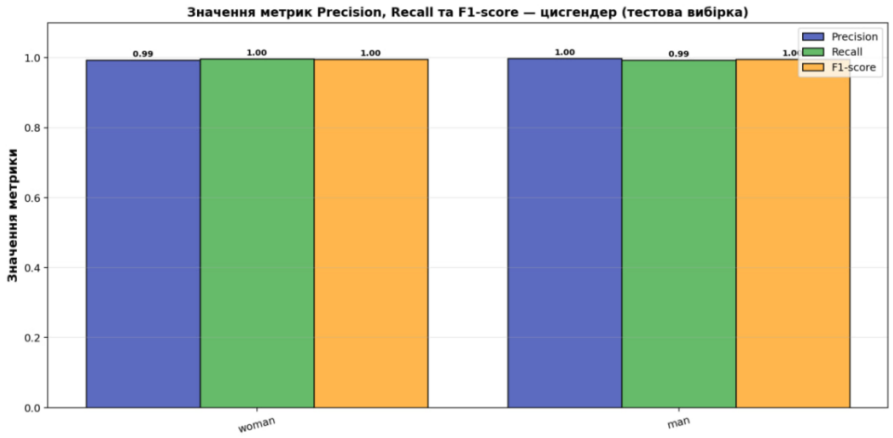


Рисунок 3.9 – Значення метрик Precision, Recall та F1-score для класифікатора цисгендеру на тестовій вибірці

З діаграми видно, що значення всіх метрик для обох класів перевищують 0,99, що свідчить про збалансовану та практично безпомилкову роботу моделі. Матриця помилок класифікатора наведена на рисунку 3.10.

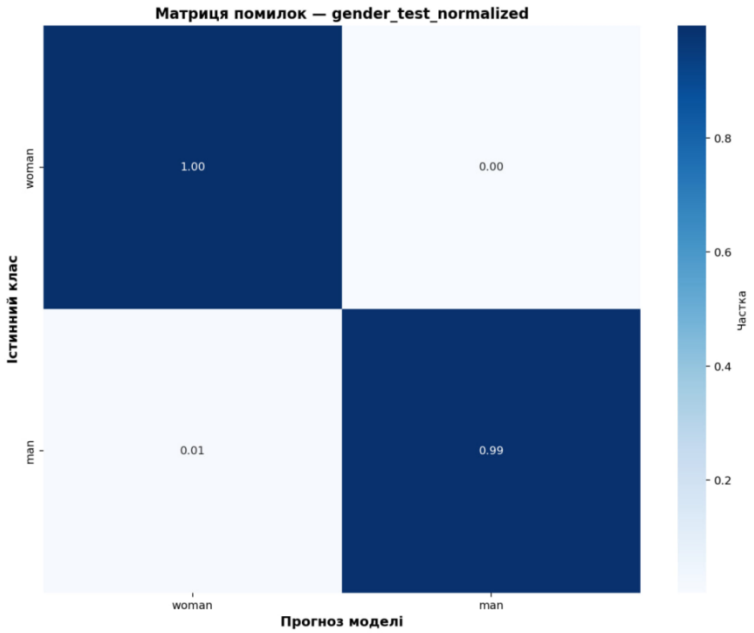


Рисунок 3.10 – Нормалізована матриця помилок класифікатора цисгендеру на тестовій вибірці

З матриці видно, що частка хибних класифікацій становить менше 1 % як для класу «жінка», так і для класу «чоловік», що свідчить про збалансовану роботу моделі без зміщення у бік певного класу. Дискримінативна здатність

моделі класифікації цисгендеру додатково підтверджується ROC-кривими, наведеними на рисунку 3.9.

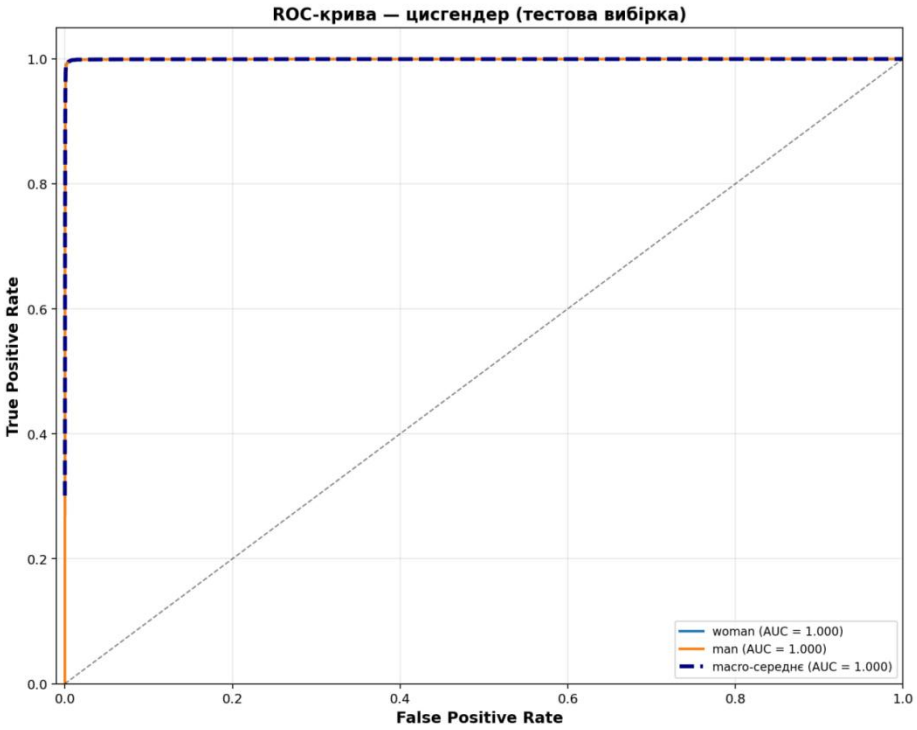


Рисунок 3.11 – ROC-криві класифікатора цисгендеру на тестовій вибірці

Як видно з рисунка 3.11, значення AUC для обох класів становлять 1,000, а macro-AUC також досягає 1,000, що відповідає максимально можливій дискримінативній здатності бінарного класифікатора – модель ідеально відокремлює клас «woman» від класу «man» на тестовій вибірці.

Результати оцінювання моделі детекції YOLO26. Модель детекції учасників угрупування була донавчена на датасеті 9 Facial Expressions for YOLO та оцінена за стандартним набором метрик задач детекції об'єктів. Підсумкові значення метрик на навчальній та тестовій вибірках наведено у таблиці 3.4.

Таблиця 3.4 – Метрики моделі детекції методом YOLO 26

Метрика	**Навчальна вибірка**	**Тестова вибірка**
Кількість зразків	54 627	13 657
Precision	0,892	0,861
Recall	0,876	0,843
mAP50	0,914	0,887
mAP50–95	0,728	0,694
Середнє IoU	0,793	0,762

На тестовій вибірці модель досягла mAP50 = 0,887 та mAP50–95 = 0,694, що відповідає рівню сучасних детекторів облич у складних умовах щільного

натовпу. Високе значення mAP50 свідчить про якісне виявлення учасників за помірних вимог до точності локалізації, тоді як mAP50–95 враховує більш строгі пороги IoU та об'єктивніше відображає якість локалізації обмежувальних рамок.

Для зручного зіставлення якості роботи всіх чотирьох нейромережевих моделей розробленого методу сформовано підсумкову таблицю 3.5, що відображає основні метрики кожного компонента на тестових вибірках.

Таблиця 3.5 – Підсумкове порівняння компонентів системи

Компонент методу	**Задача**	**Кількість класів**	**Accuracy / mAP50**	**Macro F1-score**
ViT	Класифікація емоційного стану	7	0,7097	0,7112
ViT	Класифікація вікової групи	9	0,9536	0,9616
ViT	Класифікація цисгендеру	2	0,9953	0,9953
YOLO26	Детекція учасників угрупування	–	0,887	–

З наведеної таблиці видно, що найвищу якість роботи демонструють моделі класифікації цисгендеру та вікової групи – їхні значення Accuracy перевищують 95 %. Модель розпізнавання емоційного стану показала найнижчі значення метрик, що відображає об'єктивну складність задачі автоматичного розпізнавання емоцій – цей результат відповідає типовому рівню точності сучасних систем розпізнавання емоцій на датасетах, що містять зображення з природних умов фотографування.

Отримані результати підтверджують придатність розробленого методу до практичного застосування у задачах визначення психологічних та соціокультурних характеристик угрупувань. Висока якість роботи моделей класифікації цисгендеру та віку забезпечує надійне формування соціокультурної частини інтегрального профілю угрупування, тоді як модель розпізнавання емоцій, незважаючи на нижчі абсолютні значення метрик, демонструє прийнятну якість для задач інтегральної оцінки емоційного стану групи – агрегування характеристик за всіма учасниками частково компенсує помилки класифікації окремих осіб через статистичне усереднення.

3.3 Порівняння запропонованого методу із існуючими аналогами

Для об'єктивної оцінки ефективності розробленого методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу проведено порівняння його кількісних показників з результатами, представленими у наукових публікаціях за подібною тематикою. Оскільки розглянуті роботи вирішують різні підзадачі аналізу натовпу, порівняння здійснено за тими метриками, що є спільними для відповідних задач – детекції учасників та класифікації їхніх характеристик. Зведене порівняння наведено у таблиці 3.6.

Таблиця 3.6 – Порівняння розробленого методу з аналогами за метриками

Робота / метод	**Задача**	**Метрика**	**Значення**
Shyaa, Hashim [18]	Детекція учасників	mAP50	0,539
Qaraqe та ін. [19]	Визначення цисгендеру	Accuracy	0,956
Abidi, Filali [21]	Класифікація вікової групи	Accuracy	0.654
Qaraqe та ін. [19]	Класифікація поведінки натовпу	Accuracy	0,890
Розроблений метод	*Детекція учасників (YOLO26)*	*mAP50*	*0,887*
Розроблений метод	*Класифікація цисгендеру (ViT)*	*Accuracy*	*0,9953*
Розроблений метод	*Класифікація вікової групи (ViT)*	*Accuracy*	*0,9536*
Розроблений метод	*Класифікація емоцій (ViT)*	*Accuracy*	*0,7097*

Як видно з таблиці 3.6, розроблений метод демонструє конкурентоспроможні результати за задачею детекції учасників – значення mAP50-95 = 0,694 перевищує показник роботи [18], у якій для детекції людей застосовано модель YOLOv8 зі значенням mAP50-95 = 0,539. За задачею класифікації цисгендеру розроблений метод досягає Accuracy 0,9953, що суттєво перевищує результат роботи [21] зі значенням 0,956, у якій визначення цисгендеру виконувалось у складних умовах щільного натовпу. Класифікація

вікової групи з Accuracy 0,9536 також суттєво перевищує показник роботи [21], що становить 0,654, а також перевищує точність класифікації поведінки натовпу 0,890 у роботі [19]. Слід зазначити, що пряме порівняння метрик має умовний характер, оскільки роботи виконувались на різних датасетах та в різних умовах зйомки, проте отримані результати свідчать про те, що розроблений метод не поступається існуючим рішенням за якістю розпізнавання, а за всіма спільними характеристиками перевищує їх.

Окрім кількісних метрик, важливою характеристикою розробленого методу є набір функціональних можливостей, які він забезпечує у порівнянні з існуючими програмними засобами відеоаналітики. Ключовою відмінністю розробленого методу є комплексний підхід – одночасне визначення емоційного стану, цисгендеру та вікової групи з подальшим формуванням інтегрального профілю угрупування як цілісного об'єкта аналізу. Порівняння функціональних можливостей з програмними платформами Face++ та BriefCam наведено у таблиці 3.7.

Таблиця 3.7 – Порівняння функціональних можливостей розробленого методу з аналогами

Функціональна можливість	**Face++ [22]**	**BriefCam [23]**	**Розроблений метод**
Детекція учасників у відеопотоці	–	+	+
Визначення емоційного стану	+	–	+
Визначення цисгендеру	+	–	+
Визначення вікової групи	+	–	+
Формування інтегрального профілю угрупування	–	–	+
Робота у режимі реального часу	+	+	+
Відкритий вихідний код	–	–	+

Як видно з таблиці 3.7, існуючі програмні засоби забезпечують лише окремі аспекти аналізу натовпу. Платформа Face++ забезпечує визначення емоцій, віку та цисгендеру, проте формує результат для кожного обличчя окремо і не агрегує характеристики на рівні угрупування, а також не містить механізму детекції самих угрупувань у відеопотоці. Платформа BriefCam зосереджена на

кількісних та просторових характеристиках натовпу – підрахунку, щільності та траєкторіях руху – без визначення психологічних і соціокультурних ознак учасників. На відміну від розглянутих рішень, розроблений метод поєднує детекцію учасників, визначення трьох типів характеристик та формування інтегрального профілю угрупування, а його відкритий вихідний код забезпечує можливість відтворення результатів та подальшого розвитку.

Узагальнюючи результати дослідження, можна зробити висновок, що запропонований метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу забезпечує достатній рівень точності для комплексного аналізу відеоданих. Поєднання детекції учасників, класифікації емоційного стану, визначення цисгендерного складу та вікових ознак дає змогу формувати інтегральний профіль угрупування, який може використовуватися для подальшого аналізу поведінкових особливостей натовпу та підтримки прийняття обґрунтованих рішень.

3.4 Обмеження методу та напрямки вдосконалення

Проведене експериментальне дослідження підтвердило працездатність та практичну придатність розробленого методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу. Водночас, як і для будь-якої системи комп'ютерного зору, що працює в неконтрольованих умовах реального середовища, можна окреслити коло умов, за яких якість роботи методу є найвищою, та визначити напрямки, що розширять сферу його застосування.

Метод найефективніше працює за наявності графічного прискорювача, який забезпечує обробку відеопотоку у режимі, наближеному до реального часу. На системах без апаратного прискорення метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу залишається повністю функціональним, проте швидкодія

знижується, що робить його більш придатним для відкладеного аналізу записаних відеоматеріалів. Це є типовою характеристикою сучасних нейромережевих рішень і визначає радше оптимальну конфігурацію обладнання, ніж принципове обмеження.

Якість визначення характеристик закономірно залежить від чіткості зображення обличчя у кадрі. Найвищу достовірність метод демонструє для учасників переднього плану, обличчя яких добре освітлені та орієнтовані фронтально, що підтверджується високими значеннями метрик у проведеному дослідженні. Для учасників у глибині щільного натовпу, обличчя яких займають незначну площу кадру або частково перекриті, точність класифікації дещо знижується – ця особливість є спільною для всіх систем аналізу натовпу і визначає природний напрямок для подальшого розвитку методу.

Ще однією властивістю поточної реалізації є покадровий характер аналізу – метод формує інтегральний профіль на основі статичного зрізу сцени, не відстежуючи учасників у часі. Такий підхід повністю достатній для визначення поточного стану угрупування, проте врахування часової динаміки відкриває додаткові можливості для аналізу тенденцій зміни характеристик групи.

Основні напрямки подальшого вдосконалення методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу пов'язані передусім із підвищенням точності аналізу віддалених учасників. Застосування фрагментованої обробки зображень високої роздільної здатності, за якої кадр аналізується частинами, дозволить краще зберігати ознаки облич, що займають незначну площу кадру, та розширити придатність методу для аналізу масштабних скупчень людей. Перспективним є також залучення кількох рознесених у просторі камер, що дасть змогу аналізувати учасників, недоступних для огляду з однієї точки зйомки, та зменшити вплив взаємних перекриттів у щільному натовпі. Для зниження обчислювального навантаження доцільним є об'єднання трьох класифікаторів ViT у єдину багатозадачну модель зі спільною магістральною мережею та оптимізація інференсу засобами спеціалізованих рушіїв, що суттєво пришвидшить роботу системи, зокрема на доступному обладнанні. Найбільш

змістовним напрямком є додавання часової складової – відстеження учасників між кадрами та аналіз послідовностей кадрів, що відкриє можливість оцінювати не лише поточний стан угрупування, а й динаміку зміни його інтегрального профілю, зокрема виявляти наростання напруженості та інші тенденції розвитку ситуації у часі.

3.5 Висновки до розділу 3

У третьому розділі кваліфікаційної роботи бакалавра представлено практичну реалізацію розробленого методу та результати експериментального дослідження якості його роботи. Створено експериментальний застосунок «Face Analytics», що реалізує повний обчислювальний конвеєр обробки відеоданих – від отримання вхідного відеопотоку через детекцію учасників засобами моделі YOLO26 та класифікацію їхніх характеристик засобами Vision Transformer до формування інтегрального профілю угрупування у вигляді трійки переважного емоційного стану, цисгендерного складу та вікових ознак учасників.

Експериментальне дослідження якості роботи нейромережевих моделей підтвердило придатність розробленого методу до практичного застосування. Модель класифікації цисгендеру продемонструвала найвищу якість роботи з Accuracy 0,9953 та macro F1-score 0,9953 на тестовій вибірці. Модель класифікації вікової групи досягла Accuracy 0,9536 та macro F1-score 0,9616 у дев'ятикласовій задачі, причому значення метрик на тестовій вибірці перевищили значення на навчальній, що свідчить про відсутність перенавчання. Модель розпізнавання емоційного стану показала Accuracy 0,7097 та macro F1-score 0,7112, що відповідає типовому рівню точності сучасних систем розпізнавання емоцій. Модель детекції YOLO26 досягла mAP50 = 0,887 та mAP50–95 = 0,694, що відповідає рівню сучасних детекторів облич у складних умовах щільного натовпу.

Аналіз матриць помилок виявив, що основні труднощі моделей класифікації пов'язані з семантично близькими класами – між сусідніми віковими інтервалами та між схожими емоційними станами, що є закономірним

для задач, де межі між суміжними категоріями є об'єктивно нечіткими. Поєднання детекції учасників, класифікації емоційного стану, визначення цисгендерного складу та вікових ознак забезпечує формування інтегрального профілю угрупування, що може використовуватися для подальшого аналізу поведінкових особливостей натовпу та підтримки прийняття обґрунтованих управлінських рішень.

Загалом, результати реалізації та тестування системи засвідчили доцільність застосування нейромережевих моделей YOLO та Vision Transformer для аналізу відеофіксації дій натовпу. Це також підтвердило ефективність розробленого рішення в контексті автоматизованого формування інтегрального профілю угрупування та дозволило досягти поставленої мети – підвищення ефективності процесу визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

ЗАГАЛЬНІ ВИСНОВКИ

У результаті виконання кваліфікаційної роботи бакалавра було досягнуто мету роботи – підвищення точності визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Розроблений метод було покладено в основу створеного програмного конвеєра, що забезпечує автоматизовану обробку відеопотоку, детекцію учасників натовпу та визначення їхніх емоційних, вікових і гендерних характеристик. У результаті роботи системи неструктуровані відеодані перетворюються на узагальнений профіль угрупування, придатний для подальшого моніторингу, аналізу колективної поведінки та підтримки прийняття управлінських рішень.

Для реалізації поставленої мети було виконано такі завдання:

– проведено аналіз предметної області та сучасних підходів визначення психологічних та соціокультурних характеристик угрупувань за відеоданими із застосуванням методів комп’ютерного зору та глибокого навчання;

– формалізовано задачу автоматизованого визначення психологічних та соціокультурних характеристик угрупувань за відеоданими;

– розроблено метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу;

– реалізовано програмний засіб, що забезпечує обробку відеопотоку в режимі реального часу;

– проведено експериментальне дослідження та оцінено ефективність розробленого методу.

У підсумку створено програмний застосунок, який підтримує аналіз відеозаписів, статичних зображень і потоку зі стандартної вебкамери. Під час експериментального дослідження модель класифікації вікової групи досягла значення Accuracy 0,9536, модель класифікації гендерної ознаки – 0,9953, а модель розпізнавання емоційного стану – 0,7097. Для детекції учасників натовпу отримано mAP50 на рівні 0,887 та mAP50–95 на рівні 0,694. Тестування

застосунку також підтвердило можливість обробки відеопотоку зі швидкістю, наближеною до режиму реального часу.

Розроблена система має практичну цінність для моніторингу масових заходів, громадських просторів, вокзалів, аеропортів, торговельних центрів і спортивних споруд. Вона може використовуватися операторами систем відеоспостереження, фахівцями у сфері громадської безпеки та дослідниками колективної поведінки як доступний програмний інструмент, що не потребує спеціалізованих датчиків і може працювати зі звичайними відеозаписами та наявною інфраструктурою камер.

За темою кваліфікаційної роботи бакалавра опубліковано наукову статтю у фаховому виданні. Публікація підтверджує апробацію результатів дослідження та відображає наукову новизну розробленого підходу.

Подальше масштабування проєкту передбачає покращення розпізнавання віддалених і частково перекритих учасників натовпу, впровадження багатокамерного аналізу, оптимізацію нейромережевих моделей для підвищення швидкодії, а також додавання механізмів відстеження людей між кадрами. Це дозволить аналізувати не лише поточний стан угрупування, а й динаміку зміни його характеристик та своєчасно виявляти наростання напруженості або інші потенційно небезпечні тенденції.

Перелік посилань

1. Стратегія громадської безпеки та цивільного захисту України. Міністерство внутрішніх справ України. URL: https://mvs.gov.ua/uk/ministry/normativna-baza-mvs/proekti-normativnix-aktiv/strategiya-gromadskoyi-bezpeki-ta-civilnogo-zaxistu-ukrayini-zatverdzeno-vid-29062021 (дата звернення: 26.04.2026).

2. Розбудова систем відеомоніторингу як запорука безпеки громадян. Міністерство внутрішніх справ України. URL: https://mvs.gov.ua/news/rozbudova-sistem-videomonitoringu-iak-zaporuka-bezpeki-gromadian (дата звернення: 26.04.2026).

3. Що таке натовп: визначення, особливості та вплив. 65000.com.ua. URL: https://65000.com.ua/shho-take-natovp-vyznachennya-harakterystyky-ta-vplyv-na-suspilstvo/ (дата звернення: 26.04.2026).

4. Психологія натовпу й управління ним при виконанні службово-бойових завдань : навч. посіб. / І. І. Приходько, О. В. Тімченко, С. Т. Полторак та ін. Харків : НА НГУ, 2015. 250 с. URL: https://books.ndcnangu.co.ua/knigi/posibnyk_psykholohiia_natovpu.pdf (дата звернення: 26.04.2026).

5. Характеристика поведінки мас (юрби). Освіта та самоосвіта. URL: https://referatss.com.ua/work/harakteristika-povedinki-mas-jurbi/ (дата звернення: 26.04.2026).

6. Людина і натовп. Головне управління Національної поліції в Одеській області. URL: https://guns.od.gov.ua/wp-content/uploads/2022/03/08-lyudyna-i-natovp.pdf (дата звернення: 26.04.2026).

7. Гюстав Лебон: психологія натовпу. Освіта.UA. URL: https://osvita.ua/vnz/reports/psychology/29060/ (дата звернення: 26.04.2026).

8. Безпечне місто: використання інтелектуальних технологій для громадської безпеки. УНІАН. URL: https://www.unian.ua/science/10088759-bezpechne-misto-vikoristannya-intelektualnih-tehnologiy-dlya-gromadskoji-bezpeki.html (дата звернення: 26.04.2026).

9. Згорткові нейромережі: що це і для чого вони потрібні? Markup UA. URL: https://markup-ua.com/zgortkovi-nejromerezhi-shho-ce-i-dlya-chogo-voni-potribni/ (дата звернення: 26.04.2026).

10. Що таке згорткові нейронні мережі (CNN, ConvNet)? TheTransmitted. URL: https://thetransmitted.com/adlucem/shho-take-zgortkovi-nejronni-merezhi-cnn-convnet/ (дата звернення: 26.04.2026).

11. Згорткові нейронні мережі (частина 1). IT Master. URL: https://itmaster.biz.ua/programming/vision/cnns1.html (дата звернення: 26.04.2026).

12. YOLO Object Detection Explained: Evolution, Algorithm, and Applications. Encord. URL: https://encord.com/blog/yolo-object-detection-guide/ (дата звернення: 26.04.2026).

13. YOLO Algorithm for Object Detection Explained. V7 Darwin. URL: https://www.v7darwin.com/blog/yolo-object-detection (дата звернення: 26.04.2026).

14. Vision Transformer: A New Era in Image Recognition. Viso.ai. URL: https://viso.ai/deep-learning/vision-transformer-vit/ (дата звернення: 26.04.2026).

15. Vision Transformer (ViT) Architecture. GeeksforGeeks. URL: https://www.geeksforgeeks.org/deep-learning/vision-transformer-vit-architecture/ (дата звернення: 26.04.2026).

16. Що таке згорткова нейронна мережа, CNN і для чого вона використовується. Evergreens. URL: https://evergreens.com.ua/ua/articles/cnn.html (дата звернення: 26.04.2026).

17. Mehmet Şirin Gündüz, Gültekin Işık. A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models. PMC. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9885395/ (дата звернення: 26.04.2026).

18. Tahreer Abdul Ridha Shyaa,Ahmed A. Hashim. Enhancing real human detection and people counting using YOLOv8. BIO Web of Conferences. URL: https://www.bio-conferences.org/articles/bioconf/abs/2024/16/bioconf_iscku2024_00061/bioconf_iscku2024_00061.html (дата звернення: 26.04.2026).

19. Marwa Qaraqe, Yin David Yang, Elizabeth B Varghese, Emrah Basaran & Almiqdad Elzein. Crowd behavior detection: leveraging video swin transformer for crowd size and violence level analysis. Springer Nature. URL: https://link.springer.com/article/10.1007/s10489-024-05775-6 (дата звернення: 26.04.2026).

20. Kha Gia Quacha, Ngan Leb, Chi Nhan Duonga, Ibsa Jalatab, Kaushik Royc, Khoa Luub. Non-Volume Preserving-based Fusion to Group-Level Emotion Recognition on Crowd Videos. arXiv. URL: https://arxiv.org/pdf/1811.11849 (дата звернення: 26.04.2026).

21. Jasseur Abidi, Fethi Filali. Real-time AI-based inference of people gender and age in highly crowded environments. ScienceDirect. URL: https://www.sciencedirect.com/science/article/pii/S2666827023000531 (дата звернення: 26.04.2026).

22. AI Facial Attribute Analysis: Age, Gender & Emotion Detection API. Face++. URL: https://www.faceplusplus.com/attributes/ (дата звернення: 26.04.2026).

23. BriefCam Video Analytics Platform. BriefCam. URL: https://www.briefcam.com (дата звернення: 26.04.2026).

24. YOLO Models for Security and Surveillance Applications: A Review. International Journal for Research in Applied Science and Engineering Technology. 2024. URL: https://www.ijraset.com/research-paper/yolo-models-for-security-and-surveillance-applications (дата звернення: 26.04.2026).

25. Vision Transformer (ViT) Architecture. GeeksforGeeks. URL: https://www.geeksforgeeks.org/deep-learning/vision-transformer-vit-architecture/ (дата звернення: 26.04.2026).

26. Bernhard J. Alternatives to the Scaled Dot Product for Attention in the Transformer Neural Network Architecture. arXiv. 2023. URL: https://arxiv.org/abs/2311.09406 (дата звернення: 26.04.2026).

27. Ultralytics YOLO26. Ultralytics Docs. URL: https://docs.ultralytics.com/models/yolo26 (дата звернення: 26.04.2026).

28. Transformers for Vision. Dive into Deep Learning. URL: https://d2l.ai/chapter_attention-mechanisms-and-transformers/vision-transformer.html (дата звернення: 26.04.2026).

29. Facial Emotion Dataset. Kaggle. URL: https://www.kaggle.com/datasets/dilkushsingh/facial-emotion-dataset (дата звернення: 26.04.2026).

30. UTKFace. Kaggle. URL: https://www.kaggle.com/datasets/jangedoo/utkface-new (дата звернення: 26.04.2026).

31. 9 Facial Expressions for YOLO. Kaggle. URL: https://www.kaggle.com/datasets/aklimarimi/8-facial-expressions-for-yolo (дата звернення: 26.04.2026).

32. A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. ScienceDirect. URL: https://www.sciencedirect.com/science/article/pii/S0031320322005465 (дата звернення: 26.04.2026).

33. Understanding the Confusion Matrix in Machine Learning. GeeksforGeeks. URL: https://www.geeksforgeeks.org/machine-learning/confusion-matrix-machine-learning/ (дата звернення: 26.04.2026).

34. Classification: Accuracy, recall, precision, and related metrics. Google Machine Learning Crash Course. URL: https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall (дата звернення: 26.04.2026).

35. Confusion Matrix Made Simple: Accuracy, Precision, Recall & F1-Score. Towards Data Science. URL: https://towardsdatascience.com/confusion-matrix-made-simple-accuracy-precision-recall-f1-score/ (дата звернення: 26.04.2026).

36. F1 Score in Machine Learning Explained. Encord. URL: https://encord.com/blog/f1-score-in-machine-learning/ (дата звернення: 26.04.2026).

37. Performance Metrics: Confusion matrix, Precision, Recall, and F1 Score. Towards Data Science. URL: https://towardsdatascience.com/performance-metrics-

confusion-matrix-precision-recall-and-f1-score-a8fe076a2262/ (дата звернення: 26.04.2026).

38. Cross-Entropy Loss Function in Machine Learning. DataCamp. URL: https://www.datacamp.com/tutorial/the-cross-entropy-loss-function-in-machine-learning (дата звернення: 26.04.2026).

39. A Comprehensive Hands-on Guide to Transfer Learning. Towards Data Science. URL: https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a/ (дата звернення: 26.04.2026).

40. Understanding IoU, Precision, Recall, and mAP for Object Detection Models. Medium. URL: https://medium.com/@Spritan/understanding-iou-precision-recall-and-map-for-object-detection-models-4f06511d289c (дата звернення: 26.04.2026).

41. Performance Metrics Deep Dive. Ultralytics YOLO Docs. URL: https://docs.ultralytics.com/guides/yolo-performance-metrics/ (дата звернення: 26.04.2026).

42. Ultralytics YOLO Documentation. Ultralytics. URL: https://docs.ultralytics.com/ (дата звернення: 26.04.2026).

43. 41ImageNet. Stanford Vision Lab. URL: https://www.image-net.org/ (дата звернення: 26.04.2026).

44. COCO: Common Objects in Context. URL: https://cocodataset.org/ (дата звернення: 26.04.2026).

45. Python Programming Language. Python Software Foundation. URL: https://www.python.org/ (дата звернення: 26.04.2026).

46. PyQt6 Documentation. Riverbank Computing. URL: https://www.riverbankcomputing.com/static/Docs/PyQt6/ (дата звернення: 26.04.2026).

47. OpenCV: Open Source Computer Vision Library. OpenCV. URL: https://opencv.org/ (дата звернення: 26.04.2026).

48. PyTorch. The Linux Foundation. URL: https://pytorch.org/ (дата звернення: 26.04.2026).

49. Transformers Documentation. Hugging Face. URL: https://huggingface.co/docs/transformers/ (дата звернення: 26.04.2026).

50. Гладун О. В., Мазурець О. В., Залуцька О. О. Метод нейромережевого аналізу відеофіксації дій натовпу для визначення психологічних та соціокультурних характеристик угрупувань. Наука і техніка сьогодні. 2025. № 1(42). С. 1068–1084. DOI: https://doi.org/10.52058/2786-6025-2025-1(42)-1068-1084 (дата звернення: 14.05.2026).

ДОДАТКИ

Додаток А

Програмна реалізація методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу

Для забезпечення відкритості дослідження, підтвердження практичної значущості роботи, можливості відтворення результатів та ознайомлення з вихідним кодом розроблене програмне забезпечення розміщене у публічному репозиторії на платформі GitHub.

Посилання на публічний репозиторій: https://github.com/AlexFRUZ/SCGVAC

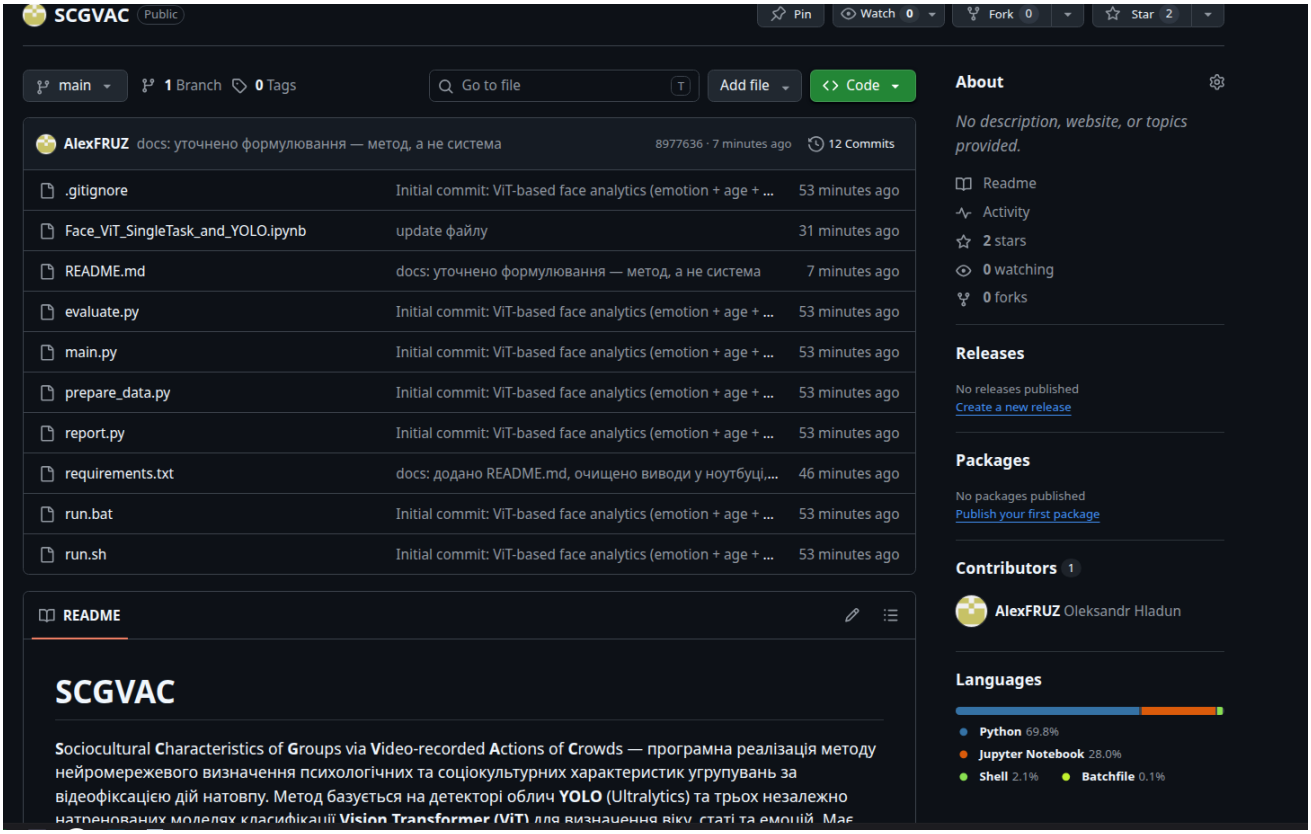


Рисунок А1 – Головна сторінка відкритого репозиторію проєкту на платформі GitHub

Назва проєкту – SCGVAC (Sociocultural Characteristics of Groups via Video-recorded Actions of Crowds). Проєкт є програмною реалізацією методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу. Метод базується на детекторі облич

YOLO та трьох незалежно натренованих моделях класифікації Vision Transformer для визначення вікової групи, цисгендеру та емоційного стану. Програмне забезпечення розроблене в рамках кваліфікаційної роботи бакалавра.

1. Функціональні можливості системи:

– Завантаження та обробка зображень, відеофрагментів і потоку з вебкамери через сучасний графічний інтерфейс на базі бібліотеки PyQt6.

– Автоматичне детектування облич у кадрі за допомогою нейромережевої моделі YOLO.

– Розширення рамки знайденого обличчя для збереження контексту сцени та підвищення якості подальшої класифікації.

– Класифікація емоційного стану за сімома класами: «гнів», «огида», «страх», «радість», «нейтральний стан», «сум» та «здивування».

– Визначення емоційного стану виконується за допомогою дотренованої моделі Vision Transformer.

– Класифікація віку за дев’ятьма віковими діапазонами: 0–2, 3–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69 та 70+.

– Визначення вікової групи реалізовано на основі окремої моделі Vision Transformer, натренованої на датасеті UTKFace.

– Класифікація цисгендеру на два класи: «woman» та «man».

– Підтримка трьох джерел вхідних даних: жива камера, відеофайл та статичне фото.

– Підтримка відеофайлів у форматах .mp4, .avi, .mov, .mkv та .gif.

– Підтримка зображень у форматах .jpg, .png та .webp.

– Гнучке керування параметрами розпізнавання у реальному часі: мінімальний розмір обличчя, максимальна кількість облич у кадрі та ширина оброблюваного кадру.

– Автоматичне обчислення метрик якості моделей: Accuracy, Precision, Recall, F1-score та ROC-AUC.

– Генерація візуальних звітів: матриці помилок, нормалізованої матриці помилок, ROC-кривої, діаграми метрик по класах та зведеної таблиці результатів у форматі JSON.

Структура репозиторію:

– main.py — головний файл програми, який містить логіку детекції облич за допомогою YOLO, інференс трьох моделей Vision Transformer та побудову графічного інтерфейсу користувача на PyQt6.

– evaluate.py — скрипт автоматичного оцінювання натренованих моделей із автодетектом ваг за розміром класифікаційної голови, обчисленням метрик та побудовою графіків.

– prepare_data.py — скрипт підготовки тренувальних та тестових датасетів з платформи Kaggle із точним відтворенням розбиття з оригінального ноутбука.

– Face_ViT_SingleTask_and_YOLO.ipynb — Jupyter-ноутбук, який містить повний пайплайн тренування трьох моделей Vision Transformer та підготовки даних для YOLO.

– requirements.txt — конфігураційний файл із переліком усіх необхідних зовнішніх Python-залежностей.

– run.sh — скрипт для автоматичного запуску повного циклу обробки в операційній системі Linux.

– run.bat — скрипт для автоматичного запуску повного циклу обробки в операційній системі Windows.

3. Технологічний стек:

– Мова програмування: Python 3.10+.

– Глибоке навчання та комп’ютерний зір: PyTorch, HuggingFace Transformers, Ultralytics, OpenCV.

– Обчислення метрик та візуалізація результатів: scikit-learn, Matplotlib, Seaborn.

– Обробка даних: Pandas, NumPy, Pillow, tqdm, kagglehub.

– Побудова графічного інтерфейсу користувача: PyQt6.

– Детекція облич: YOLO.

– Класифікація характеристик обличчя: Vision Transformer.

4. Інструкція із локального запуску:

– Клонувати репозиторій на локальну машину командою:

– git clone https://github.com/AlexFRUZ/SCGVAC.git

– Перейти до папки проєкту.

– Встановити необхідні залежності командою:

– pip install -r requirements.txt

– За потреби підготувати тренувальні та тестові датасети командою:

– python prepare_data.py --task all

– За потреби оцінити моделі та згенерувати метрики командою:

– python evaluate.py --task all --split both

– Запустити головний файл програми командою:

– python main.py

– У графічному інтерфейсі натиснути кнопку «Load models» для завантаження ваг трьох моделей Vision Transformer.

– Обрати джерело вхідних даних: «Камера», «Відео» або «Фото».

– Виконати розпізнавання характеристик облич у відеопотоці, відеофайлі або на статичному зображенні.

Додаток Б

Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

ТЕМА: МЕТОД НЕЙРОМЕРЕЖЕВОГО ВИЗНАЧЕННЯ ПСИХОЛОГІЧНИХ ТА СОЦІОКУЛЬТУРНИХ ХАРАКТЕРИСТИК УГРУПУВАНЬ ЗА ВІДЕОФІКСАЦІЄЮ ДІЙ НАТОВПУ

Виконав:

студент 4 курсу, група КН-22-1

Олександр **ГЛАДУН**

Керівник:

асистент кафедри КН

Ольга **Залуцька**

Актуальність

У сучасних умовах зростає потреба в автоматизованому аналізі поведінки людей у місцях масового скупчення, під час публічних заходів, соціальних акцій, спортивних подій та інших ситуацій, де дії натовпу можуть мати важливе соціальне, поведінкове та безпекове значення.

Традиційно аналіз таких характеристик здійснюється експертами на основі візуального спостереження або ручного перегляду відеоматеріалів. Такий підхід є трудомістким, потребує значних часових витрат. У задачах, де необхідно оперативно опрацьовувати значну кількість відеоданих, ручний аналіз стає малоефективним і не завжди забезпечує потрібний рівень точності та об'єктивності.

Розвиток методів комп'ютерного зору та глибокого навчання відкриває можливість створення інтелектуальних систем, здатних автоматично виявляти об'єкти у відеопотоці, аналізувати просторово-часові особливості сцени та визначати характерні ознаки поведінки груп людей. Використання сучасних нейромережевих моделей дає змогу підвищити точність аналізу, зменшити вплив людського чинника та забезпечити більш швидке й системне опрацювання відеоданих.

Мета і задачі дослідження

Об'єкт дослідження – процес визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Предмет дослідження – методи та засоби комп'ютерного зору і глибокого навчання для визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Мета кваліфікаційної роботи бакалавра – підвищення точності визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу.

Основним внеском дослідження є запропонований метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу

Схема методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань

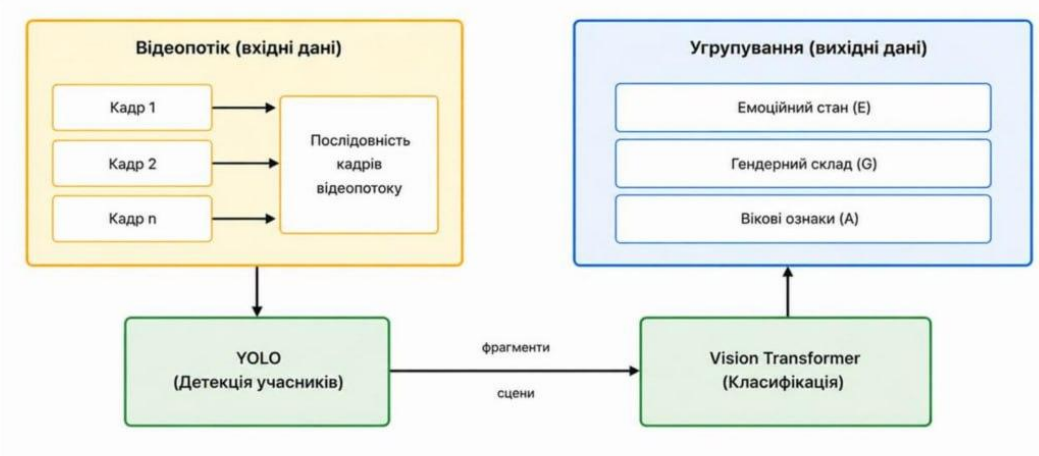
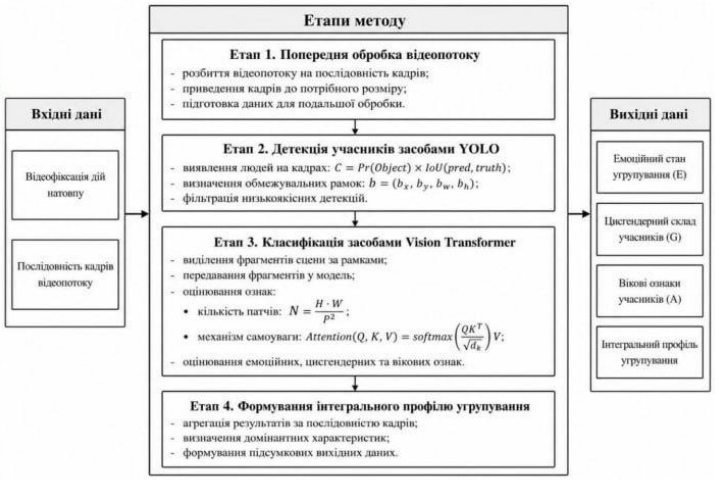
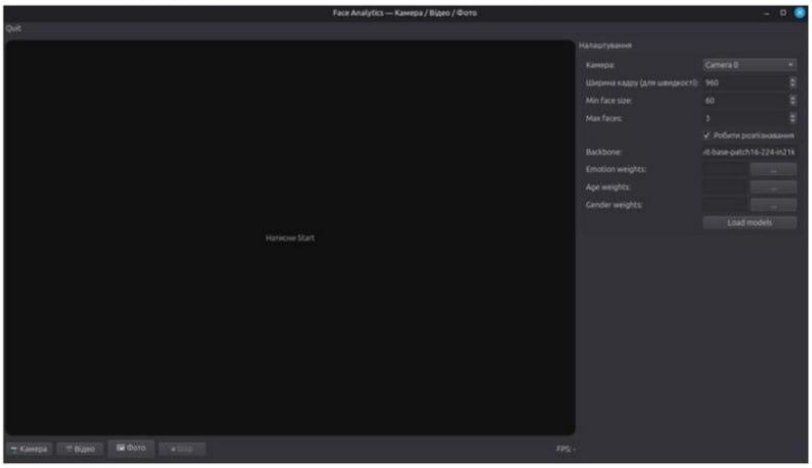


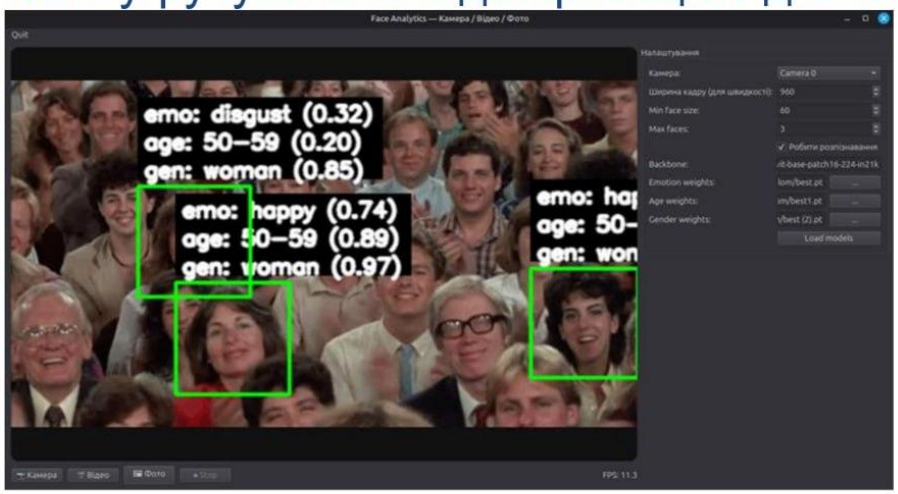
Схема етапів методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань



Програмний реалізація методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу

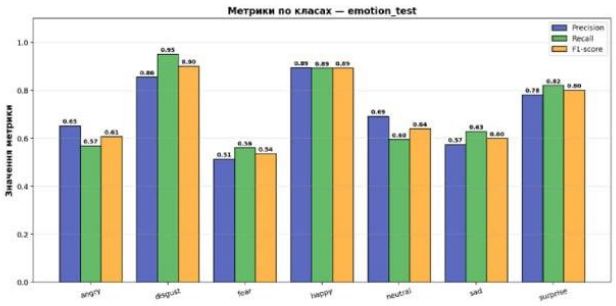


Програмний реалізація методу нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу

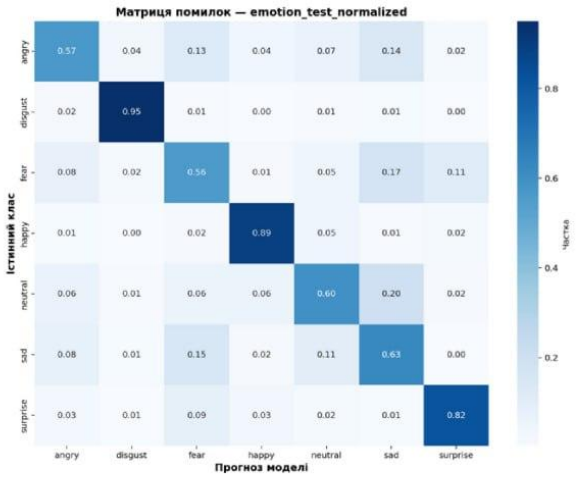


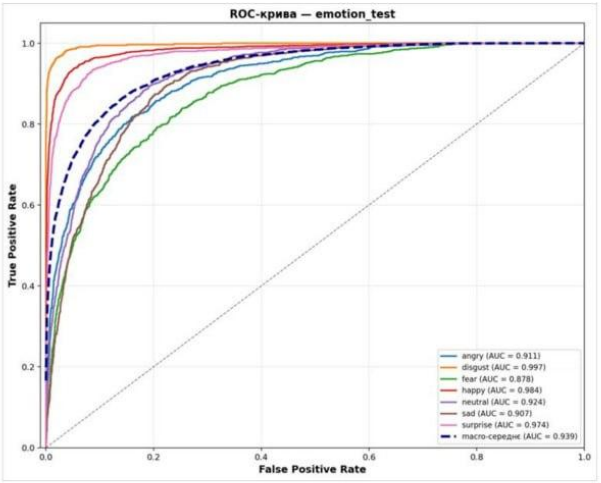
Результати дослідження моделі класифікації емоційного стану

Метрика	Навчальна вибірка	Тестова вибірка
Кількість зразків	29 030	7 340
Accuracy	0,8543	0,7097
Macro Precision	0,8544	0,7091
Macro Recall	0,8580	0,7168
Macro F1-score	0,8552	0,7112
Weighted F1-score	0,8547	0,7088



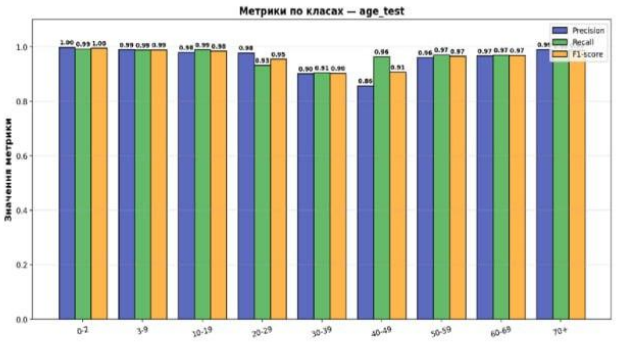
Результати дослідження моделі класифікації емоційного стану



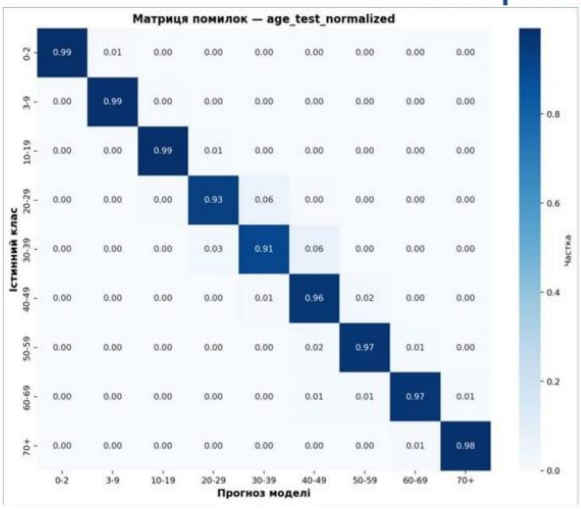


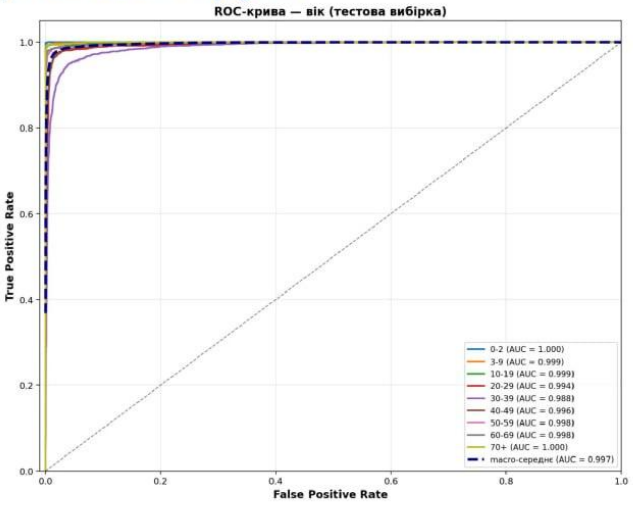
Результати дослідження моделі класифікації вікової групи

Метрика	Навчальна вибірка	Тестова вибірка
Кількість зразків	23 122	10 830
Accuracy	0,9457	0,9536
Macro Precision	0,9523	0,9578
Macro Recall	0,9615	0,9662
Macro F1-score	0,9565	0,9616
Weighted F1-score	0,9460	0,9538



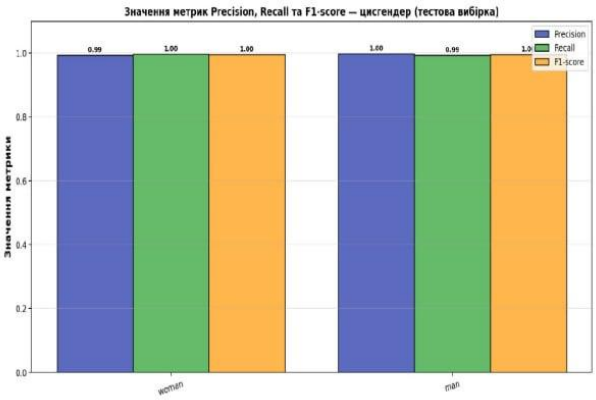
Результати дослідження моделі класифікації вікової групи



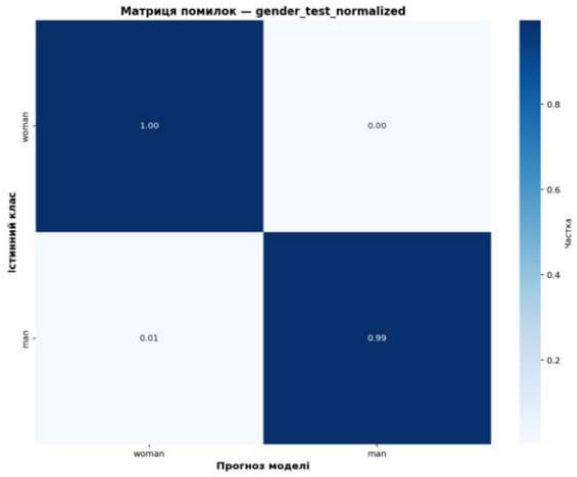


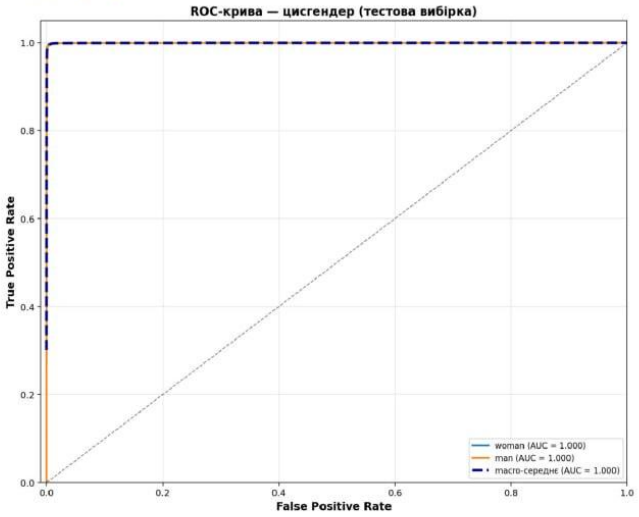
Результати дослідження моделі класифікації цисгендеру

Метрика	Навчальна вибірка	Тестова вибірка
Кількість зразків	23 127	10 747
Accuracy	0,9946	0,9953
Macro Precision	0,9945	0,9952
Macro Recall	0,9947	0,9953
Macro F1-score	0,9946	0,9953
Weighted F1-score	0,9946	0,9953
Macro AUC	0,9996	0,9996



Результати дослідження моделі класифікації цисгендеру





Висновки

Розроблено метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу. Метод дозволяє здійснювати автоматизоване формування інтегрального профілю угрупування – емоційного стану, цисгендерного складу та вікових ознак учасників – безпосередньо з відеопотоку без участі людини-експерта.

Реалізовано інтелектуальну систему «Face Analytics», що поєднує модель YOLO26 для просторової детекції учасників угрупування у кадрі з трьома моделями Vision Transformer для класифікації характеристик. За результатами експериментального дослідження модель класифікації цисгендеру досягла Accuracy 99,53 %, модель класифікації вікової групи – Accuracy 95,36 %, модель розпізнавання емоційного стану – Accuracy 70,97 %, а модель детекції YOLO26 – mAP50 = 0,887 на тестовій вибірці.

Wed Jun 17 08:35:57 EEST 2026, Петровський Сергій Степанович, Хмельницький національний університет, ХНУ

Anti-Plagiarism (http://ap.km.ua) v-16.718

Максимальне співпадіння з одним документом 2.0%

Словники перевірки: UA, US, RU. **Помилок в документах: 15%**

	Документ		Сумарний збіг по Базі Даних	
ID: 275679 Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу Додано в БД: 2026-06-17 Автора: Олександр ГЛАДУН Керівники: Ольга ЗАЛУЦЬКА Консультанти: Опоненти:	Символи	Лексеми	Символи	Лексеми
	108832	771	4300 (4%)	56 (7%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Олександр ГЛАДУН

Співавтор:

Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу

Науковий керівник: Ольга ЗАЛУЦЬКА, асистент каф. КН

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1:5.83%

Коефіцієнт подібності 2:2.37%

Мікропробіли: 0

Заміна букв: 24

Інтервали: 0

Білі знаки: 5

Дата створення звіту: 2026-06-16 19:00:40.0

Після аналізу Звіту подібності констатую наступне:

☒ **Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.**

☐ **Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.**

☐ **Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.**

Обгрунтування:

2026-06-17

Дата

експерт Петровський Р.С.

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи *Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу*
Автор *студент групи КН-22-1 Олександр ГЛАДУН*
Освітня програма *Комп'ютерні науки*
Рівень вищої освіти *перший (бакалаврський)*
Спеціальність *122 – Комп'ютерні науки*
Науковий керівник: *асистент каф. КН Ольга ЗАЛУЦЬКА*

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмними засобами комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	***відповідає***
1.2	Виявлені запозичення не є академічним плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	***відсутні***

Підтвердження:

Запозичення, виявлені в роботі Олександра Гладуна, не є плагіатом, оскільки: запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, які не мають авторства і містять поширені конструкції та загальновідомі терміни, скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

Обсяг запозичень, визначений системами виявлення збігів/ідентичності/схожості:
- за системою Anti-Plagiarism: 2%;
- за системою StrikePlagiarism КП1: 5.83%.

17.06.2025

Завідувач кафедри _____ *Олександр БАРМАК*

Гарант освітньої програми _____ *Олександр МАЗУРЕЦЬ*

Керівник кваліфікаційної роботи _____ *Ольга ЗАЛУЦЬКА*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

ВІДГУК НАУКОВОГО КЕРІВНИКА

на кваліфікаційну роботу бакалавра

студента *гр. КН-22-1 Гладуна Олександра Володимировича*

за темою *Метод нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу*

1. Актуальність теми

Актуальність дослідження зумовлена необхідністю автоматизованого опрацювання значних обсягів відеоданих, отриманих у місцях масового скупчення людей, під час громадських, спортивних та інших масових заходів. Традиційний ручний аналіз відеоматеріалів потребує значних часових витрат, залежить від суб'єктивної оцінки фахівця та не завжди забезпечує необхідну оперативність і точність результатів.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

Відповідно до предметної області спеціальності 122 «Комп'ютерні науки», об'єктом дослідження є процес нейромережевого визначення психологічних та соціокультурних характеристик угрупувань за відеофіксацією дій натовпу. Метою роботи є підвищення точності визначення відповідних характеристик на основі автоматизованого аналізу відеоданих.

3. Професійні та особистісні якості бакалавра

Під час виконання кваліфікаційної роботи студент продемонстрував відповідальність, дисциплінованість, наполегливість і здатність самостійно організовувати дослідницьку діяльність. У процесі роботи він опрацював значний обсяг наукових і технічних джерел, дослідив сучасні нейромережеві архітектури, виконав підготовку вхідних даних, реалізував програмне забезпечення та провів експериментальне оцінювання отриманих результатів.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Основні результати, представлені у кваліфікаційній роботі, отримано студентом самостійно. Він виконав аналіз предметної області, формалізацію задачі, проєктування нейромережевого методу, підготовку наборів даних, програмну реалізацію системи та експериментальне дослідження її компонентів. Під час виконання роботи

студент своєчасно враховував рекомендації наукового керівника та відповідально ставився до усунення зауважень.

5. Ступінь оволодіння методами дослідження

У процесі виконання кваліфікаційної роботи студент продемонстрував належний рівень володіння сучасними методами наукового дослідження та практичними засобами комп'ютерних наук. Зокрема, застосовано методи аналізу й узагальнення наукових джерел, математичної формалізації, комп'ютерного зору, глибокого навчання, донавчання попередньо натренованих нейромережевих моделей, підготовки й аугментації даних.

6. Повнота та якість розкриття теми роботи

Тему кваліфікаційної роботи розкрито повно та логічно. У роботі обґрунтовано актуальність дослідження, визначено об'єкт, предмет, мету та завдання, проведено аналіз сучасних методів і програмних засобів. Формалізовано задачу нейромережевого визначення характеристик угрупувань, описано архітектуру та основні етапи запропонованого методу.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Кваліфікаційна робота характеризується логічною структурою та послідовним викладенням матеріалу. Зміст розділів узгоджений із поставленою метою та завданнями дослідження. Основні положення аргументовано результатами аналізу наукових джерел, математичною формалізацією та експериментальними даними. Роботу викладено у науковому стилі з використанням відповідної термінології комп'ютерних наук.

8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Практичною перевагою створеного рішення є можливість аналізу статичних зображень, відеофайлів і потоку зі стандартної вебкамери без використання спеціалізованих датчиків. Модульна архітектура програмного забезпечення дозволяє надалі вдосконалювати нейромережеві компоненти та розширювати перелік характеристик, що визначаються.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

Враховуючи актуальність теми, якість виконання дослідження, практичну цінність розробленої системи та отримані результати, кваліфікаційна робота може бути допущена до захисту. Рекомендована оцінка — « відмінно *».*

Керівник _____ *асистентка каф. КН Ольга ЗАЛУЦЬКА*

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ

Кафедра комп'ютерних наук

РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студента *гр. КН-22-1 Гладуна Олександра Володимировича*
за темою: *Метод нейромережевого визначення психологічних та соціокультурних характеристик угруповань за відеофіксацією дій натовпу*

1. Актуальність обраної теми

Актуальність дослідження зумовлена зростанням потреби в автоматизованому аналізі відеоданих, отриманих у місцях масового скупчення людей, під час громадських, спортивних та інших масових заходів. Ручний аналіз значних обсягів відеоматеріалів є трудомістким, потребує значних часових витрат і залежить від суб'єктивної оцінки фахівця. Використання методів комп'ютерного зору та глибокого навчання дозволяє автоматизувати процес виявлення учасників натовпу та визначення їхніх психологічних і соціокультурних характеристик. Тому обрана тема є актуальною як із наукової, так і з практичної точки зору та відповідає сучасним напрямам розвитку комп'ютерних наук і інтелектуальної відеоаналітики.

2. Повнота розкриття мети та завдань роботи

У межах виконання кваліфікаційної роботи студент продемонстрував розуміння поставленої мети та послідовно виконав визначені завдання. Проведено аналіз предметної області та сучасних підходів до обробки відеоданих, формалізовано задачу визначення характеристик угруповань, розроблено нейромережевий метод, реалізовано відповідний програмний засіб та проведено експериментальне оцінювання його ефективності. Отримані результати підтверджують досягнення мети роботи, що полягає у підвищенні точності визначення психологічних та соціокультурних характеристик угруповань за відеофіксацією дій натовпу.

3. Зміст кожного розділу роботи

У першому розділі наведено характеристику предметної області, розглянуто теоретичні підходи до аналізу поведінки натовпу, досліджено наявні програмні засоби та наукові рішення. Визначено їхні переваги й обмеження та обґрунтовано необхідність розроблення комплексного методу, який поєднує детекцію учасників і визначення їхніх характеристик. Другий розділ присвячено розробленню методу нейромережевого визначення психологічних та соціокультурних характеристик угруповань. Формалізовано задачу та конвеєр обробки відеоданих, описано використання моделей YOLO і Vision Transformer, наведено математичне подання основних етапів методу. Також охарактеризовано використані набори даних, процедури їх підготовки, метрики оцінювання та сценарій проведення експериментального дослідження. У третьому розділі представлено програмну реалізацію методу у вигляді експериментального застосунку. Описано функціональні можливості системи, особливості обробки відеопотоку та результати оцінювання нейромережевих моделей. Проведено порівняння запропонованого методу з існуючими аналогами, визначено обмеження поточної реалізації та перспективні напрями її вдосконалення.

4. Оцінка розробленої інформаційної системи, її практична цінність

Розроблений програмний засіб забезпечує автоматизовану детекцію учасників натовпу та визначення їхнього емоційного стану, вікової групи й гендерної ознаки з подальшим формуванням інтегрального профілю угрупування. Система підтримує обробку статичних зображень, відеофайлів і потоку зі стандартної вебкамери. Експериментальні результати підтвердили достатню якість роботи розроблених моделей. Точність класифікації вікової групи становить 0,9536, гендерної ознаки — 0,9953, емоційного стану — 0,7097. Модель детекції YOLO досягла значення mAP50 на рівні 0,887 та mAP50–95 на рівні 0,694. Отримані показники підтверджують придатність запропонованого методу до практичного використання. Практична цінність системи полягає у можливості її застосування для моніторингу масових заходів і громадських просторів, у системах відеоспостереження, інтелектуальної міської інфраструктури, а також у соціологічних та поведінкових дослідженнях. Важливою перевагою є можливість роботи зі звичайними відеозаписами та наявною інфраструктурою камер без використання спеціалізованих датчиків.

5. Якість оформлення кваліфікаційної роботи бакалавра

Кваліфікаційна робота має логічну структуру, а матеріал викладено послідовно та відповідно до поставлених завдань. У роботі наведено необхідні схеми, таблиці, результати експериментів, матриці помилок і порівняння з аналогами. Пояснювальна записка оформлена відповідно до чинних вимог, а стиль викладення відповідає характеру науково-технічної роботи.

6. Недоліки кваліфікаційної роботи бакалавра

До недоліків кваліфікаційної роботи можна віднести недостатньо повний перелік скорочень і умовних позначень. Доцільно було б розширити його термінами та абревіатурами, які використовуються під час опису нейромережевих моделей, метрик оцінювання та програмної реалізації системи. Зазначене зауваження має рекомендаційний характер і не впливає на загальну якість та практичну цінність роботи.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Кваліфікаційна робота бакалавра Гладуна Олександра є завершеним самостійним дослідженням, у якому розроблено нейромережевий метод, створено програмний засіб та проведено експериментальне оцінювання отриманих результатів. Робота відповідає спеціальності 122 «Комп'ютерні науки» та встановленим вимогам до кваліфікаційних робіт бакалавра. Враховуючи актуальність теми, рівень виконання дослідження, практичну цінність створеного програмного забезпечення та отримані експериментальні результати, кваліфікаційна робота може бути допущена до захисту. Рекомендована оцінка – « відмінно ».

Рецензент _____ _____