

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра комп'ютерної інженерії

КОСТЕЦЬКИЙ Віталій Володимирович

Програмний модуль підрахунку змін стану елементів
людського обличчя /
Software module for counting changes in the state of
elements of the human face

спеціальність: 123 – Комп'ютерна інженерія
освітньо-професійна програма – Комп'ютерна інженерія

Кваліфікаційна робота

Виконав: студентка групи KI-41
КОСТЕЦЬКИЙ Віталій Володимирович

Науковий керівник
к.т.н. Піцун О.Й..

Тернопіль 2025

АНОТАЦІЯ

Костецький В.В. Програмний модуль підрахунку змін стану елементів людського обличчя. – Рукопис.

Кваліфікаційна робота на здобуття освітнього ступеня «бакалавр» за спеціальністю 123 «Комп'ютерна інженерія», освітньо-професійна програма. Західноукраїнський національний університет, Тернопіль, 2025.

Робота написана обсягом 46 сторінок і містить 7 рисунків, 7 таблиць, 3 додатки та 31 джерел за переліком посилань. Обсяг графічного матеріалу 2 аркуші формату А3.

Метою кваліфікаційної роботи є розробка програмного модуля для аналізу кількісних змін стану елементів людського обличчя на основі відеопотоку. Спроектовано архітектуру програмного модуля, яка реалізує поетапну обробку відеопотоку з виявленням облич, розпізнаванням ключових точок і аналізом динаміки змін елементів обличчя, що забезпечує модульність, масштабованість та інтегровану візуалізацію. Запропонована модель та алгоритм аналізу емоцій у часі забезпечують динамічне виявлення як стабільних емоційних станів, так і перехідних або змішаних виразів обличчя. Модель використовує вектор геометричних ознак для формування ймовірнісного розподілу емоцій і враховує зміну цих розподілів у часі, що дозволяє точно фіксувати моменти зміни настрою.

Ключові слова: АНАЛІЗ ОБЛИЧ, ТОЧКИ ОРІЄНТИРИ, АНАЛІЗ ЕМОЦІЙ, PYTHON.

ANNOTATION

Kostetskyi V.V. Software module for counting changes in the state of elements of the human face. – Manuscript.

Qualification work for the Bachelor degree in specialty 123 “Computer Engineering”, educational and professional program. Western Ukrainian National University, Ternopil, 2025.

The work is written in 46 pages and contains 7 figures, 7 tables, 3 appendices and 31 references. The volume of graphic material is 2 sheets of A3 format.

The purpose of the qualification work is to develop a software module for analyzing quantitative changes in the state of human facial elements based on a video stream. The architecture of the software module is designed, which implements step-by-step processing of the video stream with face detection, keypoint recognition and analysis of the dynamics of changes in facial elements, which provides modularity, scalability and integrated visualization. The proposed model and algorithm for analyzing emotions in time provide dynamic detection of both stable emotional states and transient or mixed facial expressions. The model uses a vector of geometric features to form a probability distribution of emotions and takes into account the change in these distributions over time, which allows you to accurately record moments of mood change

Keywords: FACIAL ANALYSIS, LANDMARK POINTS, EMOTIONS ANALYSIS, PYTHON.

ЗМІСТ

Перелік умовних скорочень	9
Вступ.....	10
1 Аналіз стану елементів людського обличчя.....	12
1.1 Аналіз алгоритмів розпізнавання людських облич	12
1.2 Алгоритми визначення точок-орієнтирів	16
1.3 Набори тестових зображень	18
1.4 Постановка задач кваліфікаційної роботи.....	20
2 Алгоритми визначення змін стану елементів обличчя	23
2.1 Аналіз емоцій за геометрією обличчя.....	23
2.2 Модель змін емоцій у часі.....	25
2.3 Алгоритм аналізу зміни стану елементів людського обличчя	30
3 Реалізація та дослідження програмного модуля.....	33
3.1 Реалізація програмного модуля	33
3.2 Структуру коду програмного модуля	38
3.2 Тестування програмного модуля	42
Висновки	46
Список використаних джерел	47
Додаток А Техніко-економічне обґрунтування розробки	51
Додаток Б Вихідний текст програмного засобу.....	57
Додаток В Світлокопії виданих публікацій.....	60

					КР.КІ.10660768.00.00.000 ПЗ			
Змн.	Лист	№ докум.	Підпис	Дата				
Розробив	Костецький В.В				Програмний модуль підрахунку змін стану елементів людського обличчя	Літ.	Арк.	Акрушів
Перевір.	Піцун О.Й.						7	
Консульт.	Савка Н.Я.					ЗУНУ,ФКІТ, КІ-42		
Н. Контр.	Дубчак Л.О.							
Затвердив	Дубчак Л.О.							

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

FAN	–	Facial Alignment Network
SVM	–	Support Vector Machines
CNN	–	Convolutional Neural Networks
RNN	–	Recurrent Neural Networks
AU		Action Units

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

ВСТУП

У сучасних комп'ютеризованих системах взаємодії людини з машиною ключову роль відіграє здатність алгоритмів розпізнавати та інтерпретувати невербальні сигнали, зокрема зміни міміки обличчя. Вивчення динаміки елементів людського обличчя, таких як положення брів, очей, рота, є фундаментальним для низки прикладних завдань. Зокрема завдана медичної діагностики психоемоційного стану до адаптивних інтерфейсів у сфері освіти, маркетингу, ігор і безпеки. Актуальність розробки програмного модуля, що дозволяє не лише фіксувати стани, але й відслідковувати їхню зміну в часі, зумовлена потребою у точному розумінні емоційних і когнітивних реакцій користувачів.

У багатьох галузях, зокрема в освіті, зростає інтерес до вбудованих інтелектуальних систем, які здатні реагувати на рівень зацікавленості та емоційного комфорту студентів. У медицині детекція мікровиразів або атипових змін у симетрії обличчя дозволяє здійснювати ранню діагностику неврологічних порушень або емоційних розладів. У сфері безпеки автоматизований аналіз змін міміки допомагає виявляти ознаки тривожності або агресії. Таким чином, інструмент для відстеження змін станів елементів обличчя є критично важливим компонентом сучасних біометричних та поведінкових аналітичних систем.

Разом з тим, розробка подібного програмного модуля супроводжується низкою складних інженерних і методологічних проблем. Зокрема, актуальними залишаються завдання забезпечення стійкості алгоритмів до варіативності освітлення, положення голови, індивідуальних анатомічних особливостей та культурних відмінностей у вираженні емоцій. Значним викликом є також створення моделей, здатних не лише фіксувати моментальні вирази, але й інтерпретувати динамічну зміну станів, включно з перехідними та змішаними емоціями. Особливої уваги набуває необхідність використання математичних підходів, які дозволяють формалізувати часову еволюцію обличчя, зберігаючи при цьому обчислювальну ефективність і високу точність.

Об'єкт дослідження: процес зміни стану елементів людського обличчя.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						10
Змн.	Арк.	№ докум.	Підпис	Дата		

Предмет дослідження: алгоритми аналізу емоцій людського обличчя у відеопотоці. Метою кваліфікаційної роботи є розробка програмного модуля для аналізу кількісних змін стану елементів людського обличчя на основі відеопотоку. Для досягнення мети потрібно розв'язати наступні задачі:

- проаналізувати існуючі алгоритми розпізнавання та аналізу обличчя;
- розробити моделі для визначення змін стану елементів людського обличчя;
- створити архітектуру програмного модуля та вибір відповідних бібліотек і технологій;
- реалізувати програмний модуль із можливістю обробки відео;
- провести тестування розроблених алгоритмів.

Результати дослідження апробовано на II Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених «Інтелектуальні комп'ютерні системи та мережі» [1] (додаток В).

У першому розділі проведено огляд і класифікацію методів, які використовуються в системах комп'ютерного зору для аналізу облич. Розгляд охоплює історичні підходи, сучасні згорткові нейронні мережі, а також гібридні моделі з використанням геометричних ознак.

Другий розділ присвячено побудові власної моделі та алгоритму для виявлення змін у міміці обличчя. Тут зосереджено увагу на геометричному аналізі векторів зміщення ключових точок, оцінці емоцій та динаміці їх зміни в часі.

У третьому розділі наведено практичну реалізацію запропонованого алгоритму у вигляді програмного модуля з відкритою архітектурою. Проведено експерименти та тестування на відеоданих з реальними мімічними змінами.

1 АНАЛІЗ СТАНУ ЕЛЕМЕНТІВ ЛЮДСЬКОГО ОБЛИЧЧЯ

1.1 Аналіз алгоритмів розпізнавання людських облич

Детекція та локалізація людського обличчя на зображеннях та у відеопотоці є однією з ключових задач комп'ютерного зору, що знаходить широке застосування в різних галузях, таких як біометрична ідентифікація, аналіз поведінки користувачів, моніторинг стану людини та системи відеоспостереження. У зв'язку з цим постійно зростає науковий інтерес до розробки ефективних методів, що поєднують високу точність, швидкість та стійкість до шумів і змін умов навколишнього середовища. Схему типового процесу аналізу зміни стану елементів людського обличчя у відеопотоці зображено на рисунку 1.1

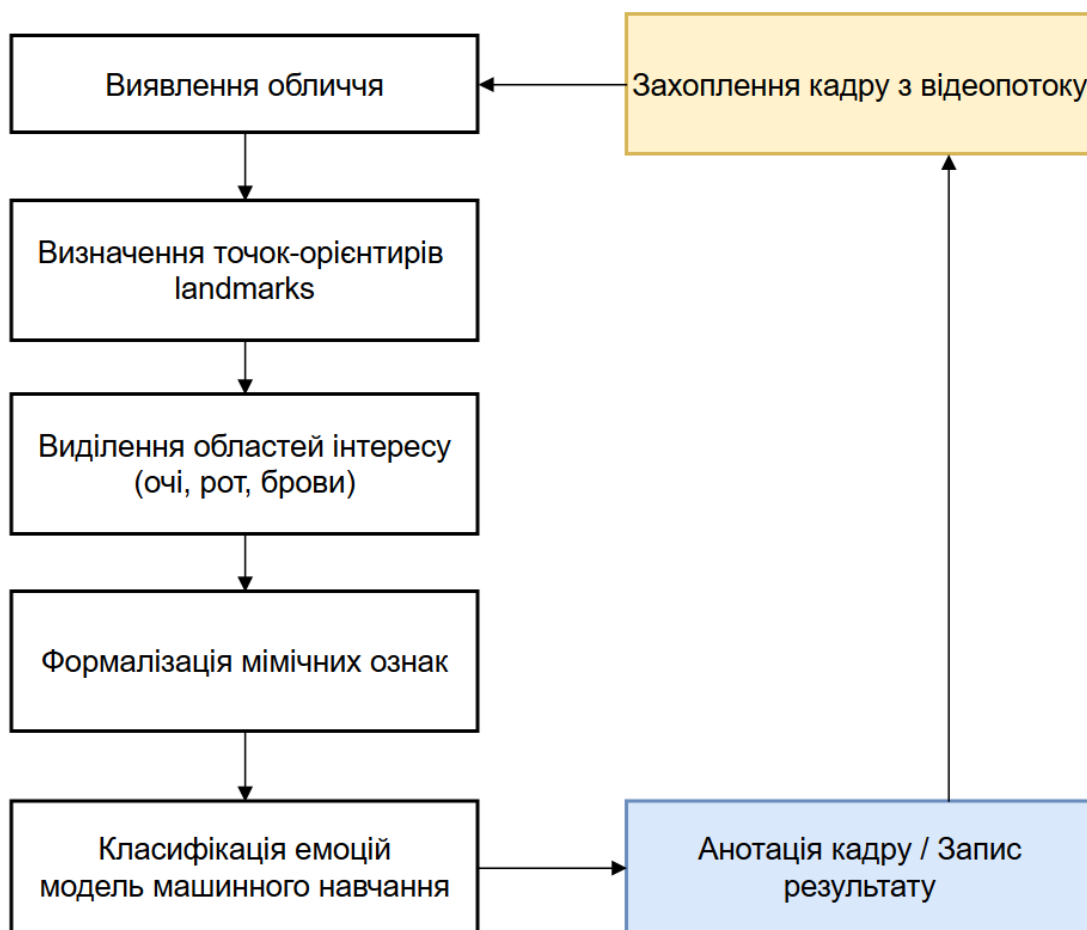


Рисунок 1.1 - Процес аналізу зміни стану елементів людського обличчя

Перші розробки у сфері виявлення або детекції облич базувалися на класичних алгоритмах, які передбачали пошук характерних рис і геометричних співвідношень окремих елементів обличчя. Такі методи здебільшого реалізовували підхід, заснований на аналізі ознак (feature-based methods), наприклад, використовуючи алгоритми Viola-Jones, які застосовують каскад класифікаторів на основі ознак Хаара (Haar features). Цей підхід відзначається високою швидкістю обробки, що дозволяє ефективно використовувати його в системах реального часу, однак характеризується недостатньою точністю та нестійкістю до змін освітлення, ракурсів і часткових перекриттів.

Наступним етапом розвитку стали методи, що використовують глибоке навчання, зокрема згорткові нейронні мережі (CNN). Підхід на основі CNN реалізує безпосередньо пошук обличчя як окремого класу об'єкта за допомогою навчених нейромережових моделей, які автоматично виділяють складні ознаки на різних рівнях абстракції. До найбільш відомих методів цієї групи належать моделі SSD (Single Shot MultiBox Detector), YOLO (You Only Look Once) [2], Faster R-CNN (Region-based CNN). Дані методи дозволяють досягти високої точності завдяки своїй здатності до узагальнення та навчанню на великих масивах зображень, однак їхньою основною проблемою залишається висока обчислювальна складність, що може обмежувати використання в ресурсно-обмежених пристроях.

Окремо слід виділити методи, спеціально призначені для аналізу відеопотоків, які враховують часову кореляцію кадрів. Тому ефективними виявляються моделі на основі рекурентних нейронних мереж (RNN) або мереж із механізмом уваги (attention-based networks), що здатні до послідовного відстеження обличчя в умовах його руху та зміни ракурсів. Подібні підходи дозволяють покращити стабільність детекції у відео за рахунок аналізу не лише просторових, але й часових характеристик обличчя, проте вони потребують додаткових ресурсів пам'яті й обчислювальних потужностей.

Для задач локалізації обличчя застосовуються додаткові етапи, що включають визначення координат та меж обличчя у вигляді прямокутних рамок (bounding boxes) або ключових точок (face landmarks). Сучасні методи, такі як

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

Multi-task Cascaded Convolutional Networks (MTCNN), вирішують задачу детекції та локалізації обличчя комплексно, поєднуючи ці етапи в єдину систему. MTCNN є багатокаскадною CNN-моделлю, яка ефективно здійснює як пошук обличчя, так і визначення точок прив'язки (очей, носа, рота тощо), забезпечуючи достатньо високу точність у поєднанні з прийнятною швидкістю роботи.

Окремий науковий напрямок становлять гібридні методи, що поєднують класичні підходи з сучасними нейромережевими технологіями. Наприклад, використання попереднього етапу на основі алгоритму Viola-Jones для швидкого відсіювання потенційних регіонів обличчя з подальшим уточненням їх за допомогою згорткових нейромереж дозволяє досягти компромісу між точністю та швидкодією.

Система кодування лицьових рухів FACS.

Система кодування лицьових рухів (Facial Action Coding System, FACS) є вичерпною, анатомічно обґрунтованою системою для опису всіх візуально помітних рухів обличчя людини. Розроблена психологами Полом Екманом та Воллесом Фрізенем у 1970-х роках, FACS розбиває вирази обличчя на окремі компоненти м'язового руху, які називаються "одинацями дії" (Action Units, AU). Кожна одиниця дії відповідає скороченню або розслабленню певного м'яза або групи м'язів обличчя.

FACS надає детальний фреймворк для розуміння того, як рухи лицьових м'язів формують різні вирази. Наприклад, посмішка може включати одиниці дії, пов'язані з підняттям щік та витягуванням куточків рота вгору. Шляхом аналізу цих рухів, навчені фахівці можуть описати будь-який вираз обличчя з точки зору інтенсивності, тривалості та симетрії задіяних одиниць дії.

Система FACS є потужним інструментом, що дозволяє об'єктивно і детально аналізувати міміку, не покладаючись на суб'єктивну інтерпретацію емоцій. Вона використовується в різноманітних галузях, включаючи психологію, дослідження емоцій, анімацію та клінічні дослідження.

В галузі інформаційних технологій та комп'ютерних наук FACS відіграє ключову роль у розробці та вдосконаленні систем комп'ютерного зору та штучного

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		14

інтелекту, пов'язаних з аналізом людської міміки. Основні напрямки використання FACS описані нижче.

Автоматичне розпізнавання емоцій з допомогою FACS. FACS є основою для багатьох алгоритмів автоматичного розпізнавання емоцій. Навчивши комп'ютерні моделі ідентифікувати та кодувати одиниці дії на зображеннях або відео облич, системи можуть потім класифікувати ці комбінації AU відповідно до відомих патернів виразів, пов'язаних з базовими емоціями (такими як радість, сум, злість, страх, здивування, відраза). Це використовується в таких додатках, як аналіз настрою користувачів, персоналізація контенту та моніторинг емоційного стану.

Взаємодія людини з комп'ютером (HCI). Розуміння виразів обличчя користувачів може покращити взаємодію з комп'ютерними системами. Наприклад, системи можуть адаптувати свою поведінку або надавати зворотний зв'язок на основі емоційного стану користувача, виявленого за допомогою аналізу FACS. Це може бути корисно в навчальних системах, іграх або адаптивних інтерфейсах.

Детекція неправди та аналіз поведінки. Хоча використання FACS для детекції неправди є дискусійним і не є абсолютно надійним, система дозволяє аналізувати мікроекспресії – дуже короткі та мимовільні вирази обличчя, які можуть свідчити про приховані емоції. Комп'ютерні системи, що використовують FACS, можуть допомагати дослідникам та аналітикам виявляти та вивчати такі тонкі рухи.

Медицина та охорона здоров'я. В IT-рішеннях для медицини FACS може застосовуватися для моніторингу стану пацієнтів, аналізу больових виразів обличчя або оцінки ефективності лікування психічних розладів, що впливають на міміку.

Маркетинг та дослідження споживачів. Аналіз виразів обличчя споживачів під час взаємодії з продуктами, рекламою або контентом за допомогою автоматизованих систем на основі FACS може надати цінну інформацію про їхню емоційну реакцію та залученість.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		15

1.2 Алгоритми визначення точок-орієнтирів

Точки-орієнтири або ключові точки (landmarks) — це координати основних анатомічних точок на обличчі (очі, ніс, рот тощо) (рисунок 1.2). Для їх знаходження використовуються моделі, натреновані на спеціальних датасетах (наприклад, 68-point або 5-point models):

- Dlib shape predictor (68 landmarks) [3] — класичний метод, який на основі координат обличчя прогнозує 68 точок;
- MediaPipe Face Mesh — сучасна модель, що знаходить понад 468 ключових точок обличчя; Дуже точний інструмент для детального аналізу;
- Facial Alignment Network (FAN) — спеціалізовані нейронні мережі для детекції ключових точок навіть на складних зображеннях.

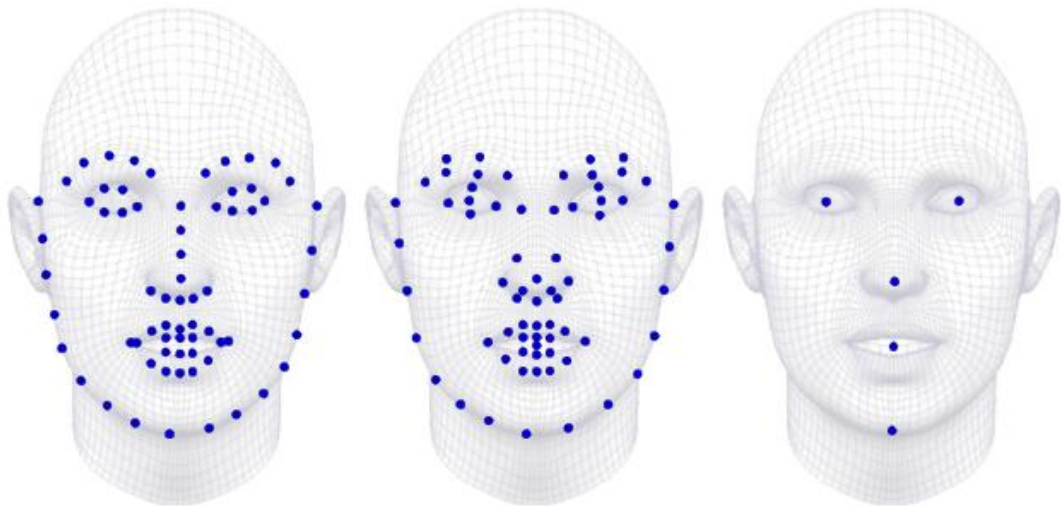


Рисунок 1.2 - Точки-орієнтири або landmarks

Вектор ознак обличчя (наприклад, з 128 або 512 елементів) отримується в результаті пропускання зображення обличчя через нейронну мережу-енкодер. Цей вектор:

- зберігає інформативні характеристики обличчя;
- дозволяє порівнювати різні обличчя за допомогою метрик відстані (наприклад, косинусної).

Вектор ознак обличчя застосовуються для таких задач:

- face verification чи належать два обличчя одній людині
- face identification знаходження людини в базі;
- emotion recognition — аналіз векторів або їх похідних для класифікації емоцій (іноді у поєднанні з окремим класифікатором);
- face clustering / search — групування схожих облич.

Таблиця 1.1 – Методи побудови векторів облич

Назва моделі	Тип	Особливості використання
FaceNet [4]	CNN-ембедінг	Повертає 128- або 512-розмірний вектор. Тренується на triplet loss. Висока точність.
OpenFace [5]	Легка модель	Відкрите рішення, менше обчислювальних витрат.
ArcFace	Advanced head	Використовує angular margin в loss-функції для кращої роздільності між класами.
CosFace	Advanced head	Вводить косинусну відстань у loss-функцію.
SphereFace	Advanced head	Модифікує простір ознак на сферичну поверхню для кращої класифікації.
ResNet / IRSE	Архітектура	Бекбони, які використовуються для побудови ембедінгів у багатьох моделях. IRSE (Improved Residual SE) — варіант із блоками squeeze-and-excitation.

Щоб визначити емоції на обличчі в кожному кадрі, використовуються класифікаційні моделі, які приймають вхідні дані (зазвичай кадр обличчя або вектор ознак) і прогнозують категорію емоції:

- класичні CNN — прості моделі на основі LeNet, VGG, ResNet.
- Pre-trained networks on AffectNet / FER2013 / RAF-DB — моделі, навчені на великих датасетах, які розпізнають базові емоції: радість, сум, страх, злість, здивування, відраза, нейтральний стан.
- Temporal models (LSTM, GRU) — для аналізу змін емоцій у часі, застосовують рекурентні мережі, які враховують послідовності кадрів.
- Transformer-based models (ViT, TimeSformer) — сучасні моделі, що можуть враховувати як просторову, так і часову інформацію для прогнозування емоцій.

1.3 Набори тестових зображень

Розглянемо датасети (рисунок 1.3) для аналізу емоцій і мімічної поведінки обличчя, які використовуються в комп'ютерному зорі для тренування та оцінювання моделей розпізнавання [6] емоційних станів у зображеннях або відео (зокрема, у відеопотоці). Вони охоплюють різні емоції, міміку, афективні стани, природні умови, динаміку та реалістичність.



Рисунок 1.3 – Приклад датасету [6]

AffectNet [7, 8] — це один із найбільших датасетів для аналізу емоцій, що містить близько мільйона реальних зображень облич, зібраних із вебу. Зображення анотовано за сімома основними емоціями та двома континуальними афективними змінними: валентністю (емоційна насиченість) і активацією (рівень збудження). Його використовують для тренування як моделей класифікації емоцій, так і регресійних моделей, здатних оцінювати інтенсивність емоційного стану. Завдяки

масштабності та різноманітності цей набір даних забезпечує хорошу генералізацію моделей у реальних умовах.

RAF-DB (Real-world Affective Faces Database) [9] включає близько 30 тисяч зображень облич з анотаціями емоцій, створеними кількома людьми. Він відзначається підвищеною реалістичністю зображень — включає варіативність у позах, освітленні та виразах. Унікальною особливістю є підтримка багатокласових та мультиміткових анотацій, що дозволяє враховувати змішані емоційні стани. RAF-DB широко застосовується для навчання моделей точного емоційного розпізнавання в умовах "дикої природи".

EmotiW [10] — це щорічне змагання, яке базується на використанні датасету AFEW. Воно орієнтоване на розпізнавання емоцій у відео, знятих у реалістичних умовах, таких як сцени з фільмів або телешоу. Основна мета EmotiW — оцінити здатність алгоритмів працювати з відеопотоками, де емоційні вирази змінюються у часі, супроводжуються шумами, змінним освітленням і неідеальними ракурсами. Це змагання стимулює розробку моделей, здатних до ефективного аналізу динамічних станів у складних середовищах.

AFEW (Acted Facial Expressions in the Wild) — це відеодатасет, зібраний із фільмів, що містить приблизно 1800 кліпів із акторськими відтвореннями семи основних емоцій. Хоча емоції тут є інсценованими, набору притаманна різноманітність у віці, статі, національності, а також реалізм у сценах, які моделюють реальні життєві ситуації. AFEW активно використовується для розробки моделей, здатних аналізувати послідовність змін міміки протягом часу, зокрема із залученням рекурентних нейронних мереж.

CREMA-D (Crowd-sourced Emotional Multimodal Actors Dataset) [11] пропонує відео- та аудіозаписи з емоційними виразами, виконаними 91 актором. Датасет є мультимодальним і включає зображення, звук та текст, що дозволяє досліджувати взаємозв'язки між мімікою та вокальними характеристиками. Завдяки високій якості запису і ретельній анотації, CREMA-D активно застосовується для навчання мультимодальних моделей, які враховують одночасно візуальні й звукові сигнали емоційного стану.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						19
Змн.	Арк.	№ докум.	Підпис	Дата		

DAiSEE (Dataset for Affective States in E-Environments) [12] орієнтований на вивчення афективних станів під час взаємодії користувачів із навчальним контентом. Він складається з тривалих відеозаписів, на яких анотовано рівні зацікавленості, нудьги, сплутаності та фрустрації. Цей датасет розроблений спеціально для моделювання емоційної взаємодії в освітніх або цифрових середовищах. DAiSEE є особливо цінним для створення систем адаптивного навчання, що можуть динамічно реагувати на зміни емоцій користувача у процесі перегляду відео.

Таблиця 1.3 – Датасети для тренування алгоритмів

Датасет	Тип даних	Застосування у відеоаналізі станів
AffectNet	зображення	Навчання фрейм-аналізу (кадр за кадром)
RAF-DB	зображення	Багатокласова класифікація емоцій у кадрах
AFEW	відео	Тренування LSTM/GRU для відстеження змін у часі
EmotiW	відео	Оцінка роботи систем на послідовності
CREMA-D	відео + аудіо	Мультимодальне розпізнавання у динаміці
DAiSEE	відео (довгі)	Виявлення стійких станів у веб-інтерфейсах

1.4 Постановка задач кваліфікаційної роботи

Узагальнимо етапи переходу від визначення ключових точок обличчя до розпізнавання емоцій [13–17].

Спочатку для визначення ключових точок (Landmark Detection). використовуються алгоритми комп'ютерного зору (наприклад, моделі глибокого навчання, такі як DLIB, MediaPipe Face Mesh, або інші спеціалізовані моделі) для виявлення та локалізації ключових точок на обличчі. Ці точки (зазвичай від десятків до сотень) відповідають важливим анатомічним областям, таким як куточки очей, брови, ніс, куточки рота, контур губ тощо. Точність визначення цих точок є критичною для подальших кроків.

Далі йде витягнення ознак (Feature Extraction), тобто обчислюються відповідні ознаки

- геометричні ознаки, тобто відстані між точками, кути, співвідношення між різними областями обличчя; Наприклад, відстань між бровами (нахмуреність), висота підняття куточків рота (посмішка), площа відкриття рота, і так далі;
- відносні зміни, якщо аналізується відео, важливі зміни положення точок з плином часу; швидкість руху, амплітуда коливань, напрямок руху;
- параметри деформації, тобто обчислення деформацій обличчя, наприклад, використовуючи метод «активних моделей зовнішнього вигляду» (Active Appearance Models, AAM) або подібні техніки.

Етап класифікації емоцій (Emotion Classification). Витягнені ознаки використовуються як вхідні дані для класифікатора. Класифікатор - це алгоритм машинного навчання, який навчено зіставляти певні комбінації ознак з конкретними емоціями. Типові методи класифікації:

- традиційні методи машинного навчання: SVM (Support Vector Machines), Random Forest, класифікатори на основі логістичної регресії.
- глибоке навчання для якого найчастіше використовуються згорткові нейронні мережі (Convolutional Neural Networks, CNN) або рекурентні нейронні мережі (Recurrent Neural Networks, RNN), особливо для аналізу відеопослідовностей. Часто використовують комбінації CNN для вилучення ознак з окремих кадрів та RNN для моделювання часової динаміки.

Етап розпізнавання переходів між емоціями. Тут аналізується часова послідовність вилучених ознак або вихідних даних класифікатора емоцій. Використовуються методи, які враховують контекст та часові залежності:

- приховані марковські моделі (Hidden Markov Models, HMM) Моделюють переходи між емоційними станами як стохастичний процес.
- рекурентні нейронні мережі (RNN) LSTM, GRU здатні запам'ятовувати попередні стани та використовувати цю інформацію для прогнозування поточних.
- часові згорткові мережі (Temporal Convolutional Networks, TCN) які ефективні для моделювання довготривалих залежностей у часових даних.

Часто як проміжний етап або як основа для розробки класифікаторів емоцій використовується Facial Action Coding System (FACS). Замість безпосереднього

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						21
Змн.	Арк.	№ докум.	Підпис	Дата		

класифікування емоцій, системи можуть спочатку визначати одиниці дії (Action Units, AU) за допомогою класифікаторів, навчених на даних, закодованих за FACS. Далі, комбінації AU зіставляються з емоціями. Переваги використання FACS:

- більш гранулярне представлення виразів обличчя;
- стійкість до варіацій у вираженні емоцій;
- можливість розпізнавання складних та змішаних емоцій.

Приклади використання FACS:

- аналіз настрою, системи оцінюють емоції людей у різних випадках: під час перегляду реклами, використання програмного забезпечення);
- взаємодія людини з комп'ютером (HCI), адаптація інтерфейсу на основі емоційного стану користувача;
- медицина: діагностика та моніторинг психічних розладів;
- робототехніка: створення емоційно-інтелектуальних роботів.

Проблеми та виклики використання FACS:

- варіативність: людські обличчя та вирази дуже різноманітні;
- шум та освітлення: якість зображення може сильно впливати на результати;
- інтерпретація емоцій може бути суб'єктивною;
- обчислювальна складність, оскільки обробка відео в реальному часі вимагає значних обчислювальних ресурсів.

Метою є створення програмного модуля, який автоматично аналізує відеопотік, виявляє обличчя, відстежує ключові елементи (landmarks), і виконує кількісний підрахунок змін їхнього стану (наприклад, зміни положення брів, губ, повік). Визначимо функції програми для реалізації в роботі:

- обробка відео або послідовності зображень;
- детекція обличчя;
- визначення точок-орєнтирів;
- кількісна оцінка змін координат точок-орєнтирів у часі;
- візуалізація результатів .

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		22

2 АЛГОРИТМИ ВИЗНАЧЕННЯ ЗМІН СТАНУ ЕЛЕМЕНТІВ ОБЛИЧЧЯ

2.1 Аналіз емоцій за геометрією обличчя

Аналіз емоцій починається з вибору ключових точок та параметрів для аналізу. Основою геометричного аналізу міміки обличчя є виявлення 68 стандартних маркерів, які використовуються у більшості сучасних систем комп'ютерного зору [18]. Ці точки відповідають основним анатомічним орієнтирам і розташовуються на роті, очах, бровах, носі та лінії щелепи. Згідно з поширеною схемою маркування:

- рот: точки 48–67 охоплюють як зовнішній, так і внутрішній контури губ;
- очі: ліва пара ока представлена точками 36–41, права — 42–47;
- брови: точки 17–21 відповідають лівій брові, 22–26 — правій;
- ніс: область носа охоплюється точками 27–35;
- щелепа: нижній контур обличчя позначено точками 0–16.

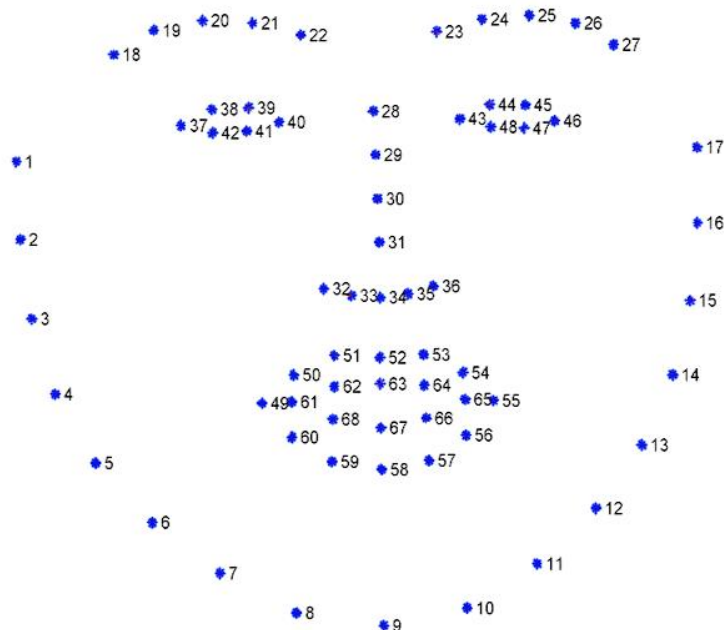


Рисунок 2.1 – 68 точок орієнтирів на обличчі

Ця система дозволяє візуалізувати топологію обличчя та забезпечує базу для подальшого чисельного аналізу емоційного стану людини.

Першим кроком у процедурі розпізнавання емоцій є отримання координат

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		23

точок обличчя зображення або кадру відео. Зазвичай для цього використовуються попередньо навчені моделі, такі як `dlib.shape_predictor`, які повертають координати точок у вигляді масиву. Після цього з набору всіх 68 точок виділяються окремі ділянки, які найбільш інформативні для емоційного аналізу — ротова щілина, брови, очі тощо.

На основі координат точок розраховуються метричні та кутові ознаки, що відображають характерні зміни міміки. Наприклад:

- відстань між внутрішніми кінцями брів: `np.linalg.norm(points[21] - points[22])` - використовується для виявлення ознак гніву;
- відкритість рота: `np.linalg.norm(points[66] - points[62])` — характеризує ступінь відкриття рота при здивуванні або сміху;
- ширина посмішки: `np.linalg.norm(points[48] - points[54])` — служить індикатором позитивних емоцій;
- стиснутість очей (eye aspect ratio): середня вертикальна відстань між точками очного контуру, наприклад, `(points[37], points[41])`, що може сигналізувати про втому або роздратування;
- кут нахилу брів — обчислюється за координатами відповідних точок (наприклад, між точками 17–19 або 22–24).

Приклад функції для обчислення відкритості рота:

```
def mouth_opening(points):  
    return np.linalg.norm(np.array(points[66]) -  
np.array(points[62]))
```

У загальному випадку існує два підходи до класифікації емоцій на основі цих ознак:

1) Евристичні (логічні) правила. Цей підхід базується на фіксованих порогових значеннях для конкретних ознак. Наприклад:

```
if mouth_opening(points) > threshold and smile_width(points) >  
threshold2:  
    emotion = "Happy"  
elif brow_furrow(points) < low_threshold:  
    emotion = "Angry"
```

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						24
Змн.	Арк.	№ докум.	Підпис	Дата		

Такий підхід є простим у реалізації та інтерпретації, проте має обмежену гнучкість і точність, особливо при варіаціях міміки та освітлення.

2) Моделі машинного навчання. У цьому випадку формується вектор ознак, який включає в себе набір геометричних параметрів (відстані, кути, співвідношення), після чого здійснюється навчання класифікаційної моделі. Алгоритм включає такі етапи:

1. векторизація параметрів (наприклад, 10–20 числових ознак для кожного зображення);
2. маркування набору даних за емоціями (радість, смуток, гнів тощо);
3. навчання моделі (наприклад, SVM, Random Forest, Logistic Regression).

Типовий приклад логіки класифікації емоції:

```
if mouth_opening(points) > 15 and smile_width(points) > 30:  
    print("Посмішка – можливо, радість")  
elif eye_aspect_ratio(points) < 0.2:  
    print("Закриті очі – можливо, сонливість або роздратування")
```

Для практичної реалізації аналізу емоцій за ключовими точками обличчя можуть використовуватись такі інструменти для мови Python:

- numpy — базова бібліотека для чисельних обчислень та роботи з векторами координат;
- pandas — використовується для структурування ознак у таблиці та підготовки датасету;
- scikit-learn — бібліотека для машинного навчання, що надає реалізації алгоритмів класифікації.

2.2 Модель змін емоцій у часі

Під час емоційної реакції людське обличчя змінюється поступово. Тобто перехід від «нейтрального» до «радісного» не відбувається миттєво — є проміжні, змішані вирази обличчя. Саме в ці моменти система, яка аналізує емоції, може

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		25

одночасно виявляти ознаки кількох емоцій або не виявляти жодної чітко [19, 19–22]. Наприклад:

- людина починає усміхатися, але ще не розплющила очі — модель може бачити частково "радість", частково "нейтральність".

- у перехідний момент модель не впевнена, яку саме емоцію показує обличчя. Замість одного чіткого результату вона дає кілька варіантів із певними ймовірностями.

Ми припускаємо, що саме такі “змішані” або “нейтральні” стани — це нормальна частина процесу зміни емоцій. Їх треба враховувати, а не викидати як помилки.

Розглянемо алгоритми визначення емоцій, які дають мультикласову відповідь. Зокрема, Бібліотека Py-Feat [23] дає ймовірності або інтенсивності для кожної базової емоції: радості, суму, злості, страху, подиву, відрази, нейтральності. Вивід — не одна емоція, а вектор значень. Наприклад, “радість: 0.65”, “подив: 0.25”, інше — близько нуля. Це дозволяє визначити змішані або перехідні стани, аналізуючи не максимум, а розподіл.

Навчені глибокі згорткові мережі, наприклад FER+ [24], зазвичай видають softmax-вектор ймовірностей на виході. Наприклад: [neutral: 0.1, happy: 0.7, surprise: 0.15]. Високі значення в кількох емоціях означають неоднозначність — система бачить ознаки кількох станів. Якщо модель дообладнана порогами або метриками (наприклад, ентропія), то вона може спеціально позначати «перехідний стан» чи «змішану емоцію».

Моделі на основі FACS визначають окремі рухи обличчя (action units), а вже потім на їх основі — емоції. Один і той самий рух може бути спільним для кількох емоцій (наприклад, підняття брів — і для подиву, і для страху). Може бути одночасна активація AUs, що відповідають різним емоціям — тому тут природно виникає мультикласовий опис.

Можна сформулювати модель кількісного аналізу змін емоційного стану в часі на основі гіпотези про перехідні стани. Така модель може враховувати як дискретність міток, так і безперервну динаміку геометричних ознак.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						26
Змн.	Арк.	№ докум.	Підпис	Дата		

Задамо вектор геометричних ознак обличчя (відстані, кути, співвідношення), що змінюється в часі t :

$$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_n(t)].$$

Мітка емоційного стану в момент часу t

$$y(t) \in \mathcal{E} \cup \{\text{neutral}\},$$

де клас емоції $\mathcal{E} = \{\text{happy}, \text{sad}, \text{angry}, \dots\}$.

Потрібно змоделювати залежність $y(t)$ як функцію від історії зміни ознак $\mathbf{x}(t - \Delta t), \dots, \mathbf{x}(t)$

Під час переходів між стабільними емоційними станами (наприклад, від "нейтрального" до "радість"), спостерігається мультикласова невизначеність, тобто в окремі моменти часу ймовірні одночасні прояви кількох емоцій або ж відсутність чіткої емоції (нейтральність).

Отже, В момент часу t модель генерує ймовірнісний вектор

$$\mathbf{p}(t) = [p_1(t), p_2(t), \dots, p_k(t)],$$

де $p_i(t) = P(y(t) = e_i \mid \mathbf{x}(t)), e_i \in \mathcal{E}$

Якщо $\max_i p_i(t) < \theta$, тобто жодна емоція не домінує, тоді вважається, що це перехідний стан або нейтральний вираз.

Міра зміни емоційного стану.

Для аналізу динаміки можна використати нормовану швидкість зміни вектора ознак:

$$\dot{\mathbf{x}}(t) = \frac{\|\mathbf{x}(t) - \mathbf{x}(t - \Delta t)\|}{\Delta t},$$

та швидкість зміни ймовірнісного розподілу емоцій:

$$\dot{\mathbf{p}}(t) = \frac{\|\mathbf{p}(t) - \mathbf{p}(t - \Delta t)\|}{\Delta t}.$$

Високі значення $\dot{\mathbf{p}}(t)$ можуть свідчити про переходи між станами, в той час як стабільний стан характеризується низькою варіативністю $\dot{\mathbf{p}}(t) \approx 0$.

Щоб згладити класифікацію та уникнути хибних сплесків, можна використовувати ковзаюче вікно та агрегувати значення:

$$\bar{\mathbf{p}}(t) = \frac{1}{T} \sum_{i=0}^{T-1} \mathbf{p}(t - i \cdot \delta t),$$

і класифікувати емоцію як:

- $\operatorname{argmax}_i \bar{p}_i(t)$, якщо $\max_i \bar{p}_i(t) > \theta$
- або як нейтральний / перехідний стан, якщо це не виконується.

Припущення про мультикласовість або нейтральність у перехідних моментах є цілком обґрунтованою з фізіологічної та психологічної точок зору. Така поведінка може бути змодельована як стохастичний перехід між станами, що проявляється через зміну геометричних ознак і ймовірнісного розподілу класифікатора. Запропонована модель дозволяє як розпізнавати сталі емоційні стани, так і інтерпретувати нестабільні переходи між ними.

На графіку візуалізовано модель динамічного розпізнавання емоцій у часі. Пояснення:

- криві нейтральної, радісної та здивованої емоцій показують, як змінюються ймовірності в часі.
- сіра пунктирна лінія — це поріг визначеності $\theta = 0,5$, що розділяє чітко виражені стани від нечітких.
- зелені області — періоди, де одна емоція домінує (тобто має ймовірність понад поріг). Це стабільні емоційні стани.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						28
Змн.	Арк.	№ докум.	Підпис	Дата		

— помаранчеві області — зони невизначеності або змішаних/нейтральних станів, де жодна емоція не домінує.

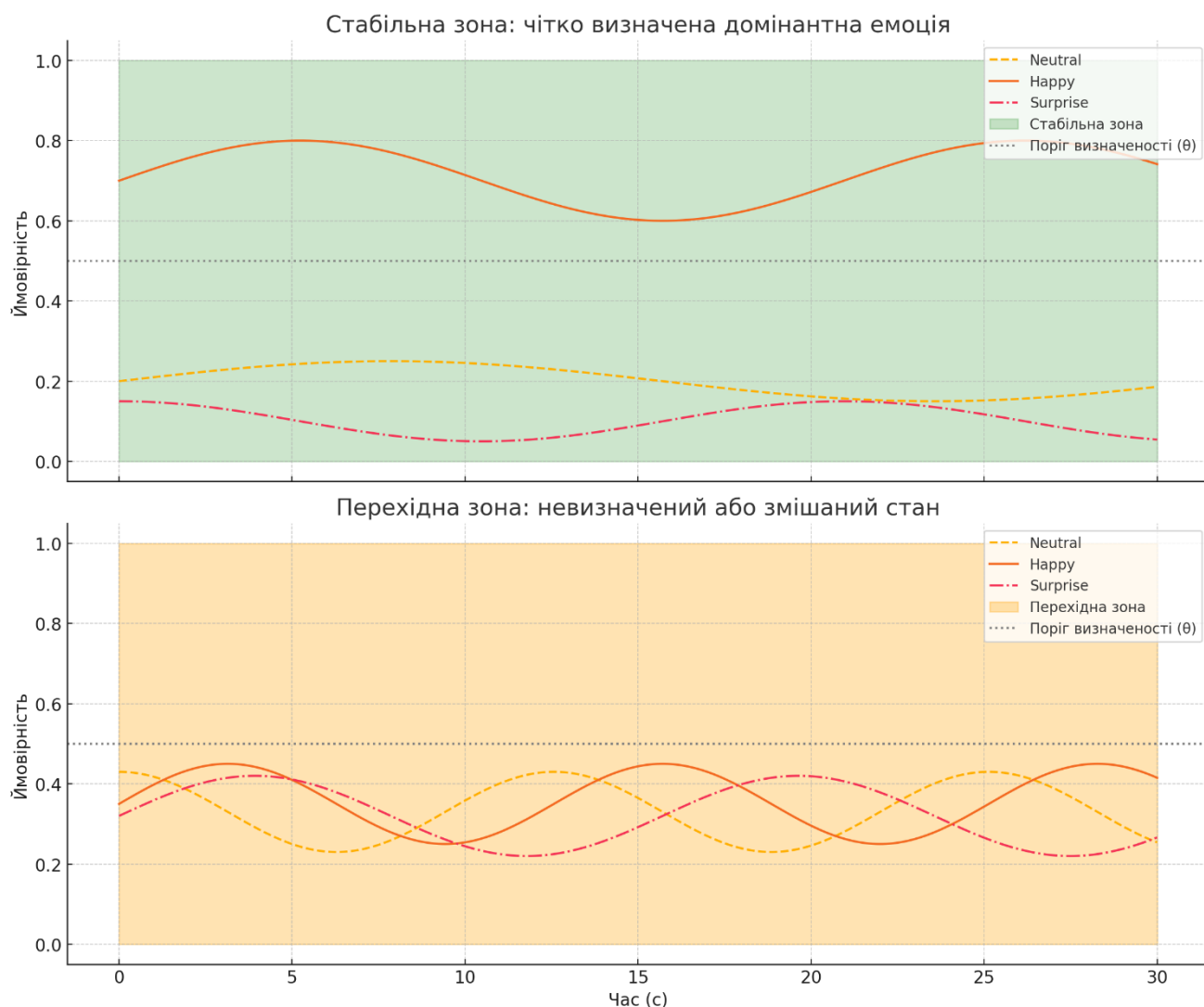


Рисунок 2.2 – Ілюстрація визначення періодів змін емоцій

Верхній графік (стабільна зона) — показує моменти, коли модель чітко визначає домінантну емоцію (ймовірність > порогу θ). Одна емоція (наприклад, "радість") стабільно домінує над іншими, ймовірності явно перевищують поріг визначеності.

Нижній графік (перехідна зона) — демонструє моменти, коли система не впевнена (жодна емоція не переважає), що свідчить про змішані або нейтральні стани. Жодна з емоцій чітко не домінує, ймовірності близькі одна до одної, що вказує на перехідні або змішані стани.

2.3 Алгоритм аналізу зміни стану елементів людського обличчя

Алгоритм реалізує детекцію змін емоційного стану особи у відеопотоці, базуючись на обчисленні динаміки геометричних ознак обличчя у часі та векторі ймовірностей емоцій. Модель включає аналіз перехідних станів, мультикласову класифікацію та механізм виявлення нечітких (нейтральних або змішаних) виразів на основі порогового аналізу.

Опишемо алгоритм динамічного аналізу емоційного стану особи

1. Вхідні дані:

- $\mathbf{X}(t) = [\mathbf{x}_t^1, \dots, \mathbf{x}_t^n]$ — послідовність векторів геометричних ознак обличчя у часі;
- Δt — фіксований крок між часовими кадрами;
- $\theta \in (0,1)$ — поріг визначеності для прийняття рішення про домінантну емоцію;
- T — довжина ковзаючого вікна згладжування.

Модель класифікації

$$\mathcal{M}: \mathbf{x}_t \mapsto \mathbf{p}_t$$

де $\mathbf{p}_t = [p_1(t), \dots, p_k(t)]$, $\sum p_i(t) = 1$, $p_i(t) = P(e_i | \mathbf{x}_t)$

Крок 1. Отримання геометричних ознак. На кожному кадрі відео із застосуванням, наприклад, dlib або mediapipe, визначаються 68 ключових точок обличчя. З них формується вектор ознак $\mathbf{x}_t$, що включає:

- відстані між бровами;
- ширину рота;
- ступінь відкритості очей;
- кути нахилу брів тощо.

Крок 2. Обчислення розподілу ймовірностей емоцій. Вектор ознак $\mathbf{x}_t$ передається до класифікатора $\mathcal{M}$, який повертає ймовірнісний розподіл $\mathbf{p}_t$ над

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						30
Змн.	Арк.	№ докум.	Підпис	Дата		

множиною емоцій $\mathcal{E}$.

Крок 3. Згладжування вектора ймовірностей. Усереднюється вектор $\mathbf{p}_t$ за ковзаючим вікном:

$$\bar{\mathbf{p}}_t = \frac{1}{T} \sum_{i=0}^{T-1} \mathbf{p}_{t-i \cdot \Delta t}.$$

Це дозволяє уникнути хибних короточасних реакцій.

Крок 4. Аналіз визначеності. Визначається максимальна ймовірність:

$$p_{\max}(t) = \max_i \bar{p}_i(t).$$

Правила:

- якщо $p_{\max}(t) > \theta$, то $y(t) = \arg \max_i \bar{p}_i(t)$ — визначено основну емоцію.
- якщо $p_{\max}(t) \leq \theta$, то $y(t)$ = перехідний або нейтральний стан.

Крок 5. Оцінка динаміки. Потрібно оцінити швидкість зміни:

$$\dot{\mathbf{p}}(t) = \frac{\| \bar{\mathbf{p}}(t) - \bar{\mathbf{p}}(t - \Delta t) \|}{\Delta t}.$$

Якщо $\dot{\mathbf{p}}(t)$ перевищує певний динамічний поріг δ , фіксується перехідний момент.

В результаті на кожному кроці часу t модель повертає:

- основну емоцію (якщо визначена);
- позначку "нейтральна" або "змішана", якщо розподіл не має домінанти;
- позначку "перехід", якщо зафіксована значна зміна.

Діаграму алгоритму зображено на рисунку 2.3

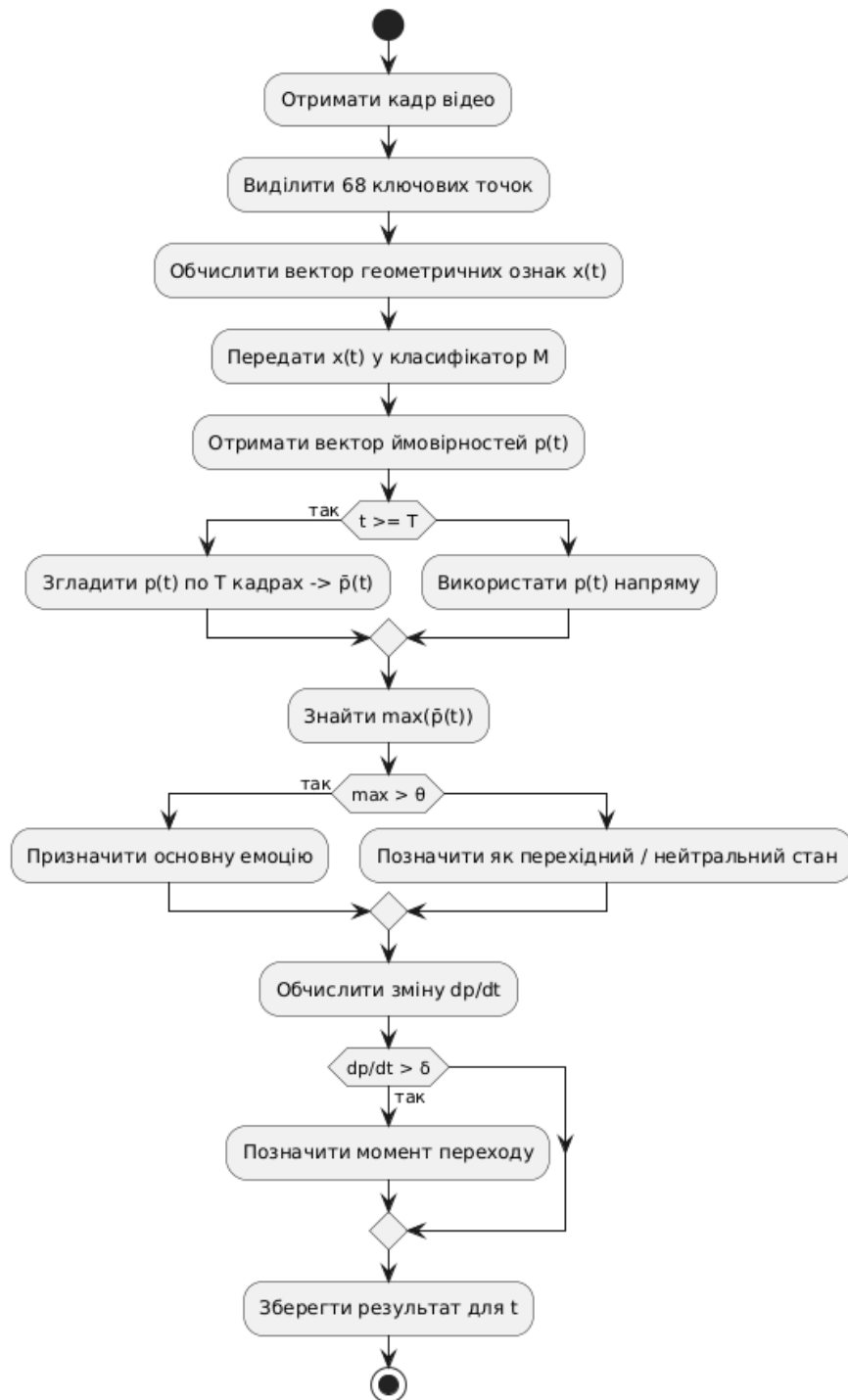


Рисунок 2.3 - UML діаграма алгоритму аналізу змін станів

Запропонована модель та алгоритм аналізу емоцій у часі забезпечують динамічне виявлення як стабільних емоційних станів, так і перехідних або змішаних виразів обличчя. Модель використовує вектор геометричних ознак для формування ймовірнісного розподілу емоцій і враховує зміну цих розподілів у часі, що дозволяє точно фіксувати моменти зміни настрою.

3 РЕАЛІЗАЦІЯ ТА ДОСЛІДЖЕННЯ ПРОГРАМНОГО МОДУЛЯ

3.1 Реалізація програмного модуля

Запропонований алгоритм аналізу емоцій за відео із урахуванням переходів і змішаних станів знаходить застосування в кількох технологічних та соціогуманітарних доменах. Її адаптивність і здатність працювати в реальному часі дозволяє інтегрувати таку систему в сучасні мультимодальні середовища, де важливо враховувати невербальну поведінку користувача або суб'єкта.

Таблиця 3.1 - Систематизація сфер застосування

Сфера застосування	Ключові цілі	Особливості функціонування
1. Освітні технології	Моніторинг уваги та залучення студентів під час онлайн-уроків	Працює з веб-камерою, вимоги до реального часу, підтримка групової аналітики
2. Людино-машинна взаємодія (HCI)	Адаптація інтерфейсу залежно від емоційного стану користувача	Вбудовування у GUI/UX-системи, низька затримка, приватність
3. Підтримка психічного здоров'я	Рання діагностика емоційних порушень (стрес, тривога, апатія)	Інтеграція з мобільними додатками, висока чутливість і специфічність
4. Відео-спостереження та безпека	Виявлення агресивної або деструктивної поведінки у публічному просторі	Працює з потоковим відео, витримує шум і різне освітлення
5. Маркетингові дослідження	Аналіз реакцій користувачів на рекламу, продукти, інтерфейси	Пост-обробка відео, вимагає високої точності, облік демографії
6. Віртуальні помічники та боти	Надання емоційно релевантних відповідей	Синхронізація емоційного стану з діалоговою системою
7. Реабілітація та логопедія	Моніторинг міміки у пацієнтів після інсульту або з розладами аутистичного спектра	Потрібна точна локалізація і чутливість до мікрорухів

Для формування технічного завдання на впровадження такої системи слід структурувати вимоги за кількома категоріями.

Таблиця 3.2 – Технічне завдання

Вимоги	Деталізація
Функціональні вимоги	Виявлення ключових точок обличчя у відеопотоці (68 або більше). Обчислення геометричних ознак та формування вектора ознак у реальному часі. Класифікація емоцій з використанням ймовірнісної моделі. Виявлення перехідних / змішаних станів на основі порогового аналізу. Згладжування результатів у часовому вікні (наприклад, 3–5 секунд).
Нефункціональні вимоги	Продуктивність: обробка не менше 10 кадрів/сек на одне обличчя. Сумісність: підтримка веб-камер, IP-камер, або відеопотоків з RTSP. Стійкість до шуму: робота при зміні освітлення, неідеальному ракурсі, мікровиразах.
Архітектурні вимоги	Розділення фронтенду (візуалізація емоцій) та бекенду (обробка зображення). Наявність механізму журналювання емоційних траєкторій з часовими мітками.
Інтерфейсні вимоги	Вивід поточної емоції на екрані або у вигляді графіка. API для отримання емоційного стану у форматі CSV. Конфігурація порогів та тривалості вікна через параметри користувача.

Проектування архітектури програмного модуля ґрунтується на модульному підході, який дозволяє розмежувати функціональність системи за окремими компонентами [14, 16, 17, 19, 25–27]. Обрана архітектура реалізує послідовну обробку відеопотоку, в якій кожен етап виконує чітко визначену функцію: від захоплення зображення до інтерпретації динаміки міміки з подальшою візуалізацією або експортом даних.

Функціональна структура системи передбачає п'ять основних етапів обробки. На першому етапі здійснюється захоплення відео, яке може надходити як з вебкамери, так і з попередньо записаного відеофайлу. Цей компонент повинен забезпечувати стабільне надходження кадрів з мінімальною затримкою, уніфікувати формат відеопотоку та, за потреби, здійснювати попередню нормалізацію зображення (наприклад, масштабування, конвертацію в сірий колір).

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		34

Далі активується модуль детекції обличчя, який у кожному кадрі локалізує область обличчя, повертаючи координати його обмежувального прямокутника. Визначена область передається до модуля landmarks — центрального компонента, який на основі нейронної мережі встановлює координати ключових точок обличчя, що є базовим представленням мімічної топології.

Наступним етапом є модуль аналізу зміни стану, який порівнює координати ключових точок у часовій послідовності, обчислюючи відстані, кути, співвідношення та їхню зміну в часі. Основне завдання цього блоку полягає в тому, щоб на основі обчислених геометричних ознак визначити тенденції динаміки міміки: наприклад, відкриття рота, стискання очей, нахил брів або поява посмішки. Він може реалізовувати як прості порогові алгоритми, так і більш складні моделі — наприклад, класифікацію ознак за допомогою машинного навчання. Завершальний компонент відповідає за виведення результатів. Він може формувати інтерактивні графіки, зберігати оброблені дані у форматах CSV, генерувати відео з накладеними анотаціями або забезпечувати взаємодію з користувачем через графічний або веб-інтерфейс.

Для реалізації даної архітектури обрано Python як основну мову програмування. Такий вибір зумовлений широкою підтримкою бібліотек для комп'ютерного зору, глибокого навчання, обробки масивів даних і візуалізації. Захоплення відео, попередня обробка зображення та генерація анотацій реалізуються за допомогою бібліотеки OpenCV, яка забезпечує як високу продуктивність, так і багатий набір інструментів для роботи з графічною інформацією. Для детекції обличчя і landmarks використовуються або MediaPipe, яка дозволяє працювати з високоточними моделями у реальному часі (включно з тривимірними точками), або Dlib, яка забезпечує стабільні результати на основі 68 стандартних ключових точок. Чисельна обробка координат точок, обчислення евклідових відстаней, нормалізація та побудова ознак виконуються за допомогою бібліотек NumPy і Pandas, що забезпечують ефективну роботу з великими обсягами даних і дозволяють проводити аналітику на рівні кожного кадру або агрегованих інтервалів часу.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						35
Змн.	Арк.	№ докум.	Підпис	Дата		

Для графічного відображення результатів обчислень використовуються бібліотеки Matplotlib та Seaborn. Вони дозволяють будувати графіки змін ключових ознак у часі, наприклад — графіки відкриття рота, амплітуди рухів брів чи ступеня стиснення очей. У випадку необхідності реалізації користувацького інтерфейсу застосовуються або PyQt (для побудови десктопного GUI), або Streamlit — зручна платформа для створення легких вебінтерфейсів без складної верстки. Такий підхід дозволяє зробити модуль доступним як для фахівців, так і для кінцевих користувачів, не знайомих із програмуванням [15, 17, 28–30].

Розробка ведеться у середовищах, зручних для модульної інтеграції та тестування. У фазі прототипування перевага надається JupyterLab. Для повноцінної розробки програмного продукту використовуються PyCharm або VSCode, що дозволяють організувати проєктну структуру з використанням модулів, тестових сценаріїв та інтеграції з системами контролю версій.

Після завершення реалізації перевіряється коректність детекції ключових точок на різних типах облич, стабільність трекінгу при зміні ракурсу та освітлення, а також точність обчислення метрик. Тестування також охоплює ситуації з відсутністю обличчя в кадрі, декількома обличчями або частковими перекриттями.

Структура проєкту (рисунок 3.1) описує послідовну модульну архітектуру системи аналізу емоцій у відеопотоці. Вона реалізує п'ятиетапну обробку: захоплення кадрів, детекція облич, визначення ключових точок, аналіз зміни геометрії, класифікація емоцій, згладжування, інтерпретація та виведення результатів.

Розгортання функціональних модулів системи передбачає послідовну реалізацію основних етапів обробки відеопотоку — від захоплення кадрів до виведення результатів аналізу. Архітектурно кожен етап є незалежним компонентом, що взаємодіє з попереднім через чітко визначені інтерфейси (наприклад, масив зображень, координати точок, значення ознак тощо), що дозволяє гнучко адаптувати систему до різних джерел даних і типів обробки.

На першому етапі реалізується модуль захоплення відео. Він відповідає за ініціалізацію відеопотоку з вебкамери або відеофайлу за допомогою функціоналу

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						36
Змн.	Арк.	№ докум.	Підпис	Дата		

cv2.VideoCapture з бібліотеки OpenCV. Модуль забезпечує читання кадрів у реальному часі та підготовку їх до подальшої обробки.

```
face_emotion_analyzer/
+-- app/
|   +-- __init__.py
|   +-- main.py           # Точка входу
|   +-- video_input.py    # Захоплення відео
|   +-- face_detector.py  # Детектор облич
|   +-- landmark_detector.py # Визначення ключових точок
|   +-- analyzer.py       # Аналіз зміни координат
|   +-- classifier.py     # Модуль класифікації емоцій
|   +-- smoother.py       # Ковзне згладжування
|   +-- output.py         # Виведення / експорт результатів
|   L-- config.py        # Конфігурація порогів, інтервалів
L-- tests/
    +-- test_analyzer.py
    +-- test_pipeline.py
    L-- test_integration.py
```

Рисунок 3.1 - Структура каталогів проекту

Другий етап передбачає детекцію обличчя в кожному кадрі. Залежно від обраної бібліотеки, використовується або MediaPipe Face Detection, яка забезпечує високу швидкодію та стабільність в реальному часі, або Dlib.get_frontal_face_detector() — класичний метод з хорошою точністю для зображень з фронтальними ракурсами. Модуль повертає координати обличчя у вигляді прямокутника, що охоплює відповідну ділянку кадру.

Третій етап полягає у виявленні ключових точок обличчя (landmarks), які формують топологічну модель міміки. У випадку використання Dlib отримується 68 двовимірних точок, що відповідають таким елементам, як очі, брови, ніс, губи, щелепа. Альтернативно MediaPipe Face Mesh дозволяє отримати до 468 3D-точок, що забезпечує детальніший аналіз рухів у просторі. Результатом модуля є координатна матриця ключових точок для кожного кадру.

Четвертий етап — аналіз динаміки — є центральним у структурі системи. Тут обчислюються евклідові відстані між відповідними точками на різних кадрах, а також зміни координат у вигляді різниць $\Delta x(t)$, $\Delta y(t)$, які формують часові ряди. На їхній основі визначаються зміни положення окремих зон обличчя (наприклад, ступінь відкриття рота, підняття брів, моргання). Цей модуль може бути

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						37
Змн.	Арк.	№ докум.	Підпис	Дата		

доповнений згладжуванням сигналу або використанням похідних для виявлення різких змін.

П'ятий етап — виведення та візуалізація — реалізує інтерфейс користувача або інтерфейс експорту. Оброблені координати та обчислені ознаки зберігаються у вигляді таблиць CSV.

Після завершення реалізації функціональних модулів проводиться тестування системи. Воно включає перевірку роботи алгоритмів на різних типах відео — статичних зображеннях, емоційно забарвлених роликах, а також у режимі реального часу. Аналізується точність та стабільність детекції облич і точок при зміні положення голови та умов освітлення. Особлива увага приділяється вимірюванню продуктивності системи, зокрема середньої кількості кадрів на секунду (FPS), що може бути досягнута на цільовому обладнанні.

На завершальному етапі здійснюється розгортання програмного модуля. Діаграма реалізації функціональних модулів зображена на рисунку 3.2.

3.2 Структуру коду програмного модуля

Опишемо структуру коду програмного модуля. Файл `main.py` починається з директив імпорту:

```
from video_input import get_video_stream
from face_detector import detect_faces
from landmark_detector import get_landmarks
from analyzer import analyze_changes
from output import visualize_frame, export_data
```

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						38
Змн.	Арк.	№ докум.	Підпис	Дата		

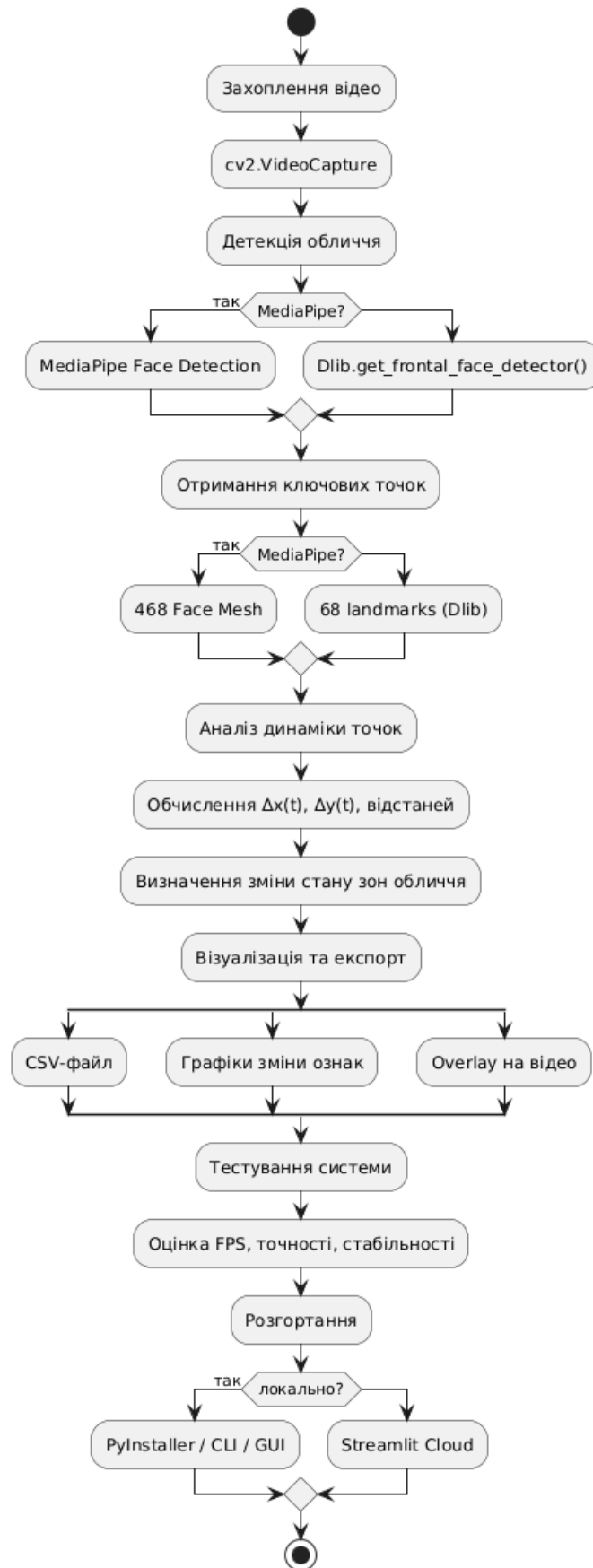


Рисунок 3.2 - Процес реалізації функціональних модулів

Далі main.py містить головну функцію запуску та управління всіма етапами:

```

for frame in get_video_stream("input.mp4"):
    faces = detect_faces(frame)
    for face in faces:
        landmarks = get_landmarks(frame, face)
        changes = analyze_changes(landmarks)
        visualize_frame(frame, landmarks, changes)
        export_data(changes)

```

Модуль video_input.py Обробляє відеофайл або потік з камери:

```

import cv2

def get_video_stream(source):
    cap = cv2.VideoCapture(source)
    while cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break
        yield frame
    cap.release()

```

Модуль face_detector.py детектує обличчя у кадрі:

```

import mediapipe as mp

face_detection = mp.solutions.face_detection.FaceDetection()

def detect_faces(frame):
    results = face_detection.process(frame)
    return results.detections if results.detections else []

```

Модуль landmark_detector.py визначає координати ключових точок:

```

import mediapipe as mp

face_mesh = mp.solutions.face_mesh.FaceMesh()

def get_landmarks(frame, face):
    results = face_mesh.process(frame)
    if results.multi_face_landmarks:
        return results.multi_face_landmarks[0]
    return None

```

Модуль analyzer.py Підраховує зміну координат landmark-ів між кадрами:

```

prev_landmarks = None

def analyze_changes(current_landmarks):
    global prev_landmarks
    deltas = []
    if prev_landmarks:
        for p1, p2 in zip(prev_landmarks, current_landmarks):

```

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						40
Змн.	Арк.	№ докум.	Підпис	Дата		

```

dx = p2.x - p1.x
dy = p2.y - p1.y
deltas.append((dx, dy))
prev_landmarks = current_landmarks
return deltas

```

Таблиця 3.3 - Формат вхідних та вихідних даних

Компонент	Вхідні дані	Вихідні дані
video_input	шлях до відеофайлу або камери	кадри зображення (BGR)
face_detector	кадр	список bounding boxes облич
landmark_detector	кадр, область обличчя	масив координат точок (x, y [, z])
analyzer	масив точок поточного кадру	список векторів зміщення (Δx , Δy)
output	кадр і зміни	візуалізація та/або збереження у CSV/JSON

Особливості інтеграції:

- усі модулі мають уніфіковані інтерфейси: передають/отримують типові структури даних (масиви, кортежі, словники).
- модуль analyzer зберігає попередні координати між кадрами, що дозволяє обчислювати зміну стану.
- Код може бути легко доповнений обробкою декількох облич або глибинною інформацією (z-координати).
- для визначення емоцій застосовано нейронну мережу — результат інтегрується в модуль analyzer.

Діаграма класів наведена на кресленні КР.КІ.10660768.00.00.002.С1 Основні класи подано в таблиці. 3.4.

Таблиця 3.4 - Основні класи програмного модуля

Клас	Призначення та методи
VideoInput	забезпечує читання відеопотоку з файлу або камери. Основні методи: <code>init__(source: str)</code> — ініціалізація джерела відео. – <code>read_frame()</code> -> <code>ndarray</code> — повертає поточний кадр. – <code>release()</code> — звільняє ресурси.
Face Detector	виявляє обличчя на кадри. Основні методи: <code>__init__()</code> — ініціалізує детектор (Dlib). <code>detect(frame: ndarray) -> List[BoundingBox]</code> — повертає список координат знайдених облич.

Landmark Detector	визначає ключові точки (landmarks) на обличчі. Основні методи: __init__() — ініціалізація моделі детекції точок. get_landmarks(frame: ndarray, face_bbox: BoundingBox) -> List[Point] — повертає координати точок обличчя.
Feature Extractor	обчислює геометричні ознаки з набору ключових точок. Основні методи: extract(landmarks: List[Point]) -> List[float] — повертає вектор ознак, таких як відстані, кути, співвідношення
Emotion Classifier	визначає ймовірнісний розподіл емоцій за геометричними ознаками. Основні методи: __init__(model_path: str) — завантаження класифікаційної моделі. predict(features: List[float]) -> List[float] — повертає розподіл ймовірностей по класах емоцій.
Temporal Smoother	згладжує розподіл ймовірностей емоцій у часі. update(p: List[float]) -> List[float] — додає новий вектор у вікно та повертає згладжений розподіл. reset() — очищає буфер вікна
Emotion Analyzer	аналізує згладжений розподіл і визначає поточну емоцію, нейтральність або перехід. analyze(p_avg: List[float], prev_p: List[float]) -> EmotionState — повертає результат класифікації та динамічну зміну.
Output Manager	Призначення: зберігає або виводить результати обробки (візуалізація, логування). Основні методи: save_to_csv(data: List[Tuple]) — експортує дані у CSV. draw_overlay(frame: ndarray, landmarks: List[Point], emotion: str) — накладає анотації на зображення. plot_graphs(timeseries: Dict[str, List[float]]) — будує графіки змін.
Main Controller	координує весь процес: обробку відео, виклики до інших класів, логіку збереження та виводу. Основні методи: run() — головна функція запуску обробки. process_frame(frame: ndarray) — обробка одного кадру.

3.2 Тестування програмного модуля

Проведемо тестування програмних засобів [31]. Для аналізу міміки та емоцій на основі відео або зображень Застосовано бібліотеку Py-Feat. Вона використовує попередньо навчені моделі (зокрема на основі OpenFace та FER+), щоб оцінити ймовірності різних емоцій. Основні кроки виглядають так:

1. Виявлення обличчя з допомогою детектора обличчя Dlib.

2. Визначення точок-орієнтирів (рисунок 3.3).

3. Обчислення інтенсивностей скорочення окремих м'язів обличчя (за системою FACS — Facial Action Coding System). Py-Feat прогнозує значення кожної AU як неперервну величину, зазвичай від 0 до 5 або від 0 до 1.

4. Прогнозування емоцій в конкретному кадрі. На основі інтенсивностей AU та вхідного зображення Py-Feat обчислює ймовірність кожної з базових емоцій (наприклад, радість, сум, страх, гнів, відраза, здивування, нейтральний стан).

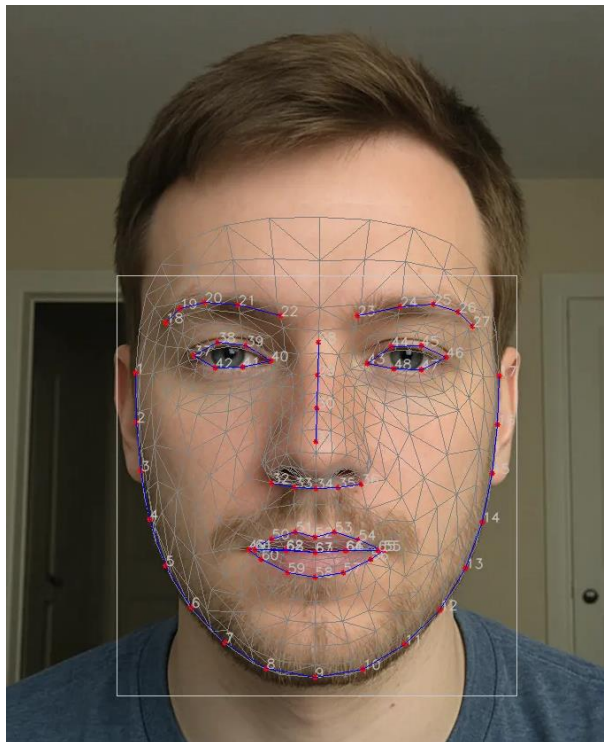
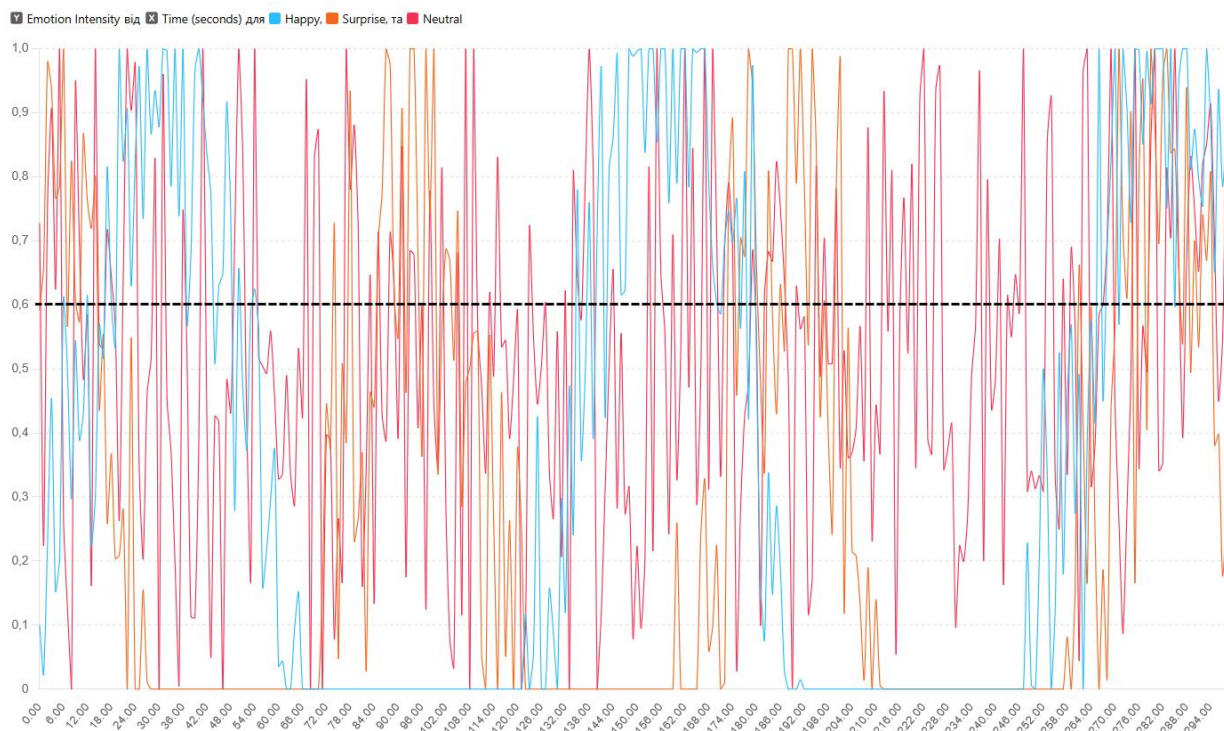


Рисунок 3.3 – Детекція обличчя та точок-орієнтирів

На рисунку 3.4 представлено визначення зміни емоційних станів у часі при зашумлених даних. Візуалізовані емоції: "Радість" (Happy), "Подив" (Surprise) та "Нейтральний стан" (Neutral). Сірим пунктиром відмічено поріг визначеності, вище якого емоція вважається стабільно домінуючою.

Стабільні зони характеризуються чіткою перевагою однієї емоції (значення перевищують поріг), тоді як перехідні зони демонструють моменти невизначеності, коли система не визначає чіткої домінуючої емоції через близькість або нижчі значення інтенсивностей різних емоцій.



Рисунку 3.4- Визначення зміни емоційних станів при зашумлених даних

На рисунку 3.5 представлено зміни емоційних станів у часі без додаткового шуму. Відображено чіткі та тривалі періоди стабільності емоційних станів: Стабільні періоди кожної емоції тривають довше та чітко виділяються значеннями, що значно перевищують пороговий рівень (сірий пунктир).

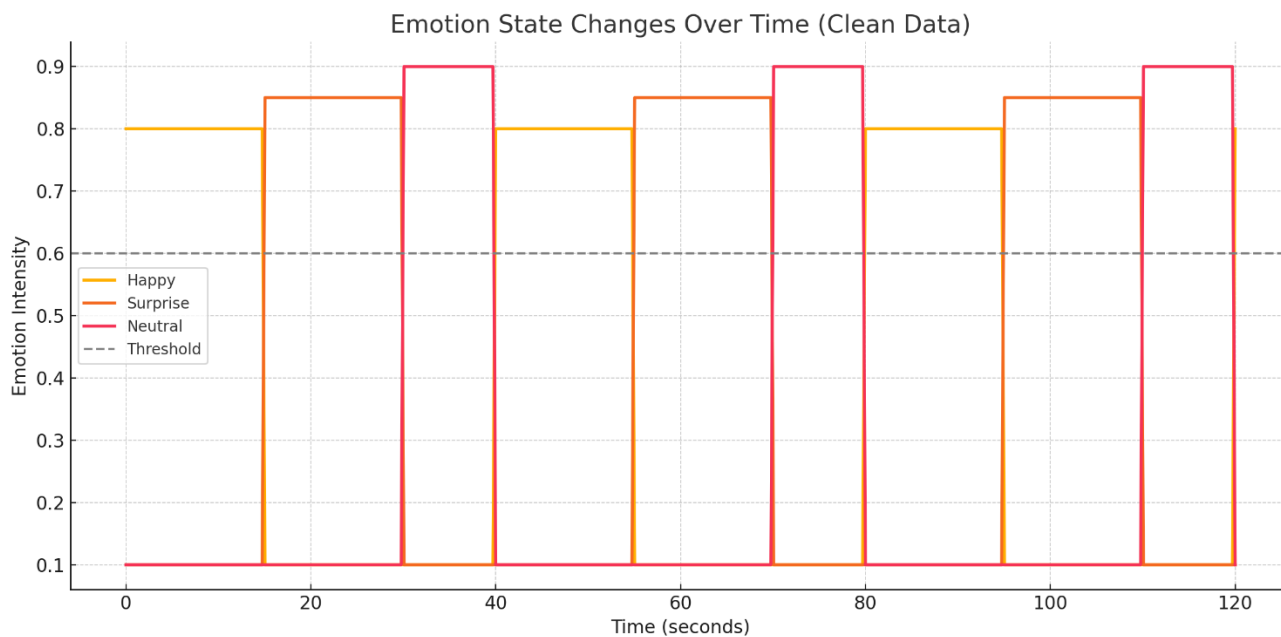


Рисунок 3.5 - Зміни емоційних станів у часі без додаткового шуму

Оранжевими вертикальними лініями позначено моменти переходу між стабільними зонами. На рисунку 3.6 тестовий сценарій зі значною випадковістю в змінах емоційних станів ("Радість", "Подив" і "Нейтральний стан").

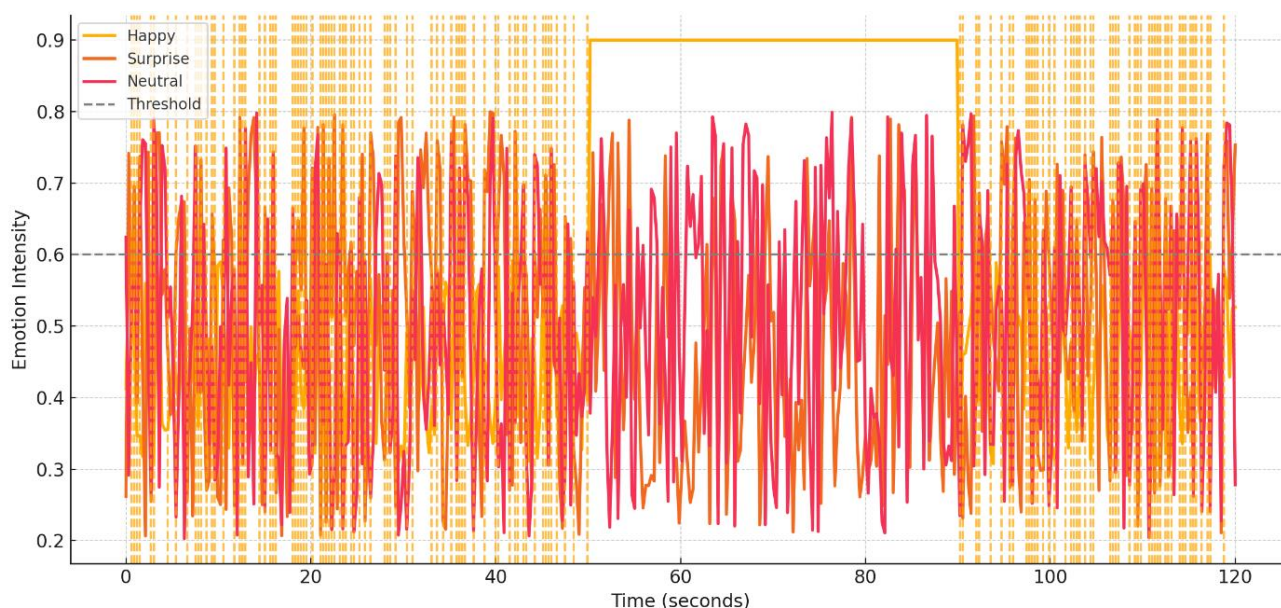


Рисунок 3.6 – Визначення стабільного стану

В період від 50 до 90 секунд чітко визначено стабільний стан емоції "Радість". Інші проміжки характеризуються частими випадковими змінами інтенсивності емоцій, що створюють численні перехідні зони, які позначені оранжевими вертикальними пунктирними лініями.

У ході проведеного тестування розробленого алгоритму підрахунку змін стану елементів людського обличчя було підтверджено його здатність ефективно розпізнавати стабільні емоційні періоди, а також чітко визначати перехідні стани з високим ступенем невизначеності. Алгоритм виявив стійкість до випадкових змін і показав можливість стабільного розпізнавання емоцій, навіть за умов значних коливань інтенсивності виразів обличчя.

Практична цінність розробленого алгоритму полягає у його широкій сфері застосування, яка включає освітні технології, інтерактивні інтерфейси, системи відеоспостереження, маркетингові дослідження.

ВИСНОВКИ

1. Виконано системний аналіз сучасних алгоритмів розпізнавання та аналізу обличчя, що дозволило виявити переваги та обмеження різних підходів — від класичних евристичних методів до моделей глибокого навчання — і сформулювати обґрунтовану базу для вибору оптимальних технологій, які ляжуть в основу реалізації програмного модуля динамічного аналізу міміки.

2. Розроблено моделі та алгоритм для визначення змін стану елементів людського обличчя у відеопотоці, що охоплюють геометричний аналіз ключових точок, оцінку емоцій за часовими ознаками та виявлення перехідних або змішаних емоційних станів на основі ймовірнісного класифікаційного підходу.

3. Спроектовано архітектуру програмного модуля, яка реалізує поетапну обробку відеопотоку з виявленням облич, розпізнаванням ключових точок і аналізом динаміки змін елементів обличчя, що забезпечує модульність, масштабованість та інтегровану візуалізацію.

4. Запропонована модель та алгоритм аналізу емоцій у часі забезпечують динамічне виявлення як стабільних емоційних станів, так і перехідних або змішаних виразів обличчя. Модель використовує вектор геометричних ознак для формування ймовірнісного розподілу емоцій і враховує зміну цих розподілів у часі, що дозволяє точно фіксувати моменти зміни настрою.

5. Реалізований програмний модуль довів свою функціональну придатність: він дозволяє обробляти відеодані, коректно виявляти зміни міміки та генерувати відповідні графіки або дані для подальшого аналізу, що підтверджено успішними результатами тестування на різних відеоматеріалах.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						46
Змн.	Арк.	№ докум.	Підпис	Дата		

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Костецький В. В. Алгоритми та засоби для аналізу емоцій у відеопотоці. // *Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Інтелектуальні комп'ютерні системи та мережі» (ІКСМ весна 2025)*. (м. Тернопіль, . м. Тернопіль, С. 104–105.

2. Liu W., Anguelov D., Erhan D. та ін. SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016*. ред. Bastian Leibe, Jiri Matas, Nicu Sebe, Max Welling. Cham : Springer International Publishing, 2016. С. 21–37. DOI:10.1007/978-3-319-46448-0_2.

3. shape-predictor-68-face-landmarks · GitHub Topics · GitHub. URL: <https://github.com/topics/shape-predictor-68-face-landmarks> (accessed 12/05/2025).

4. William I., Ignatius Moses Setiadi D. R., Rachmawanto E. H. et al. Face Recognition using FaceNet (Survey, Performance Test, and Comparison). *2019 Fourth International Conference on Informatics and Computing (ICIC) 2019 Fourth International Conference on Informatics and Computing (ICIC)*. (Semarang, Indonesia, 10.2019). Semarang, Indonesia : IEEE, 2019. DOI:10.1109/ICIC47613.2019.8985786. P. 1–6.

5. GitHub - TadasBaltrusaitis/OpenFace: OpenFace – a state-of-the art tool intended for facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation. URL: <https://github.com/TadasBaltrusaitis/OpenFace> (accessed 12/05/2025).

6. Facial Emotion Recognition Dataset. URL: <https://www.kaggle.com/datasets/tapakah68/facial-emotion-recognition> (accessed 12/05/2025).

7. AffectNet Dataset | Papers With Code. URL: <https://paperswithcode.com/dataset/affectnet> (accessed 24/05/2025).

8. Mollahosseini A., Hasani B., Mahoor M. H. AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. 2017. DOI:10.48550/ARXIV.1708.03985.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

9. Real-world Affective Faces (RAF) Database. URL: <http://www.whdeng.cn/RAF/model1.html> (accessed 24/05/2025).
10. Dhall A., Goecke R., Gedeon T. та ін. Emotion recognition in the wild. *Journal on Multimodal User Interfaces*. Вип. 10, № 2. С. 95–97. DOI:10.1007/s12193-016-0213-z.
11. CheyneyComputerScience/CREMA-D. Crowd Sourced Emotional Multimodal Actors Dataset (CREMA-D). Cheyney Computer Science, 2025. URL: <https://github.com/CheyneyComputerScience/CREMA-D> (accessed 25/05/2025).2025.
12. DAiSEE: Dataset for Affective States in E-Environments. URL: <https://people.iith.ac.in/vineethnb/resources/daisee/index.html> (accessed 25/05/2025).
13. Глибовець М. М., Олецький О. В. Штучний інтелект: Підручник. К. : Вид. дім „КМ Академія”, 2002. 366 с.
14. Luger G. F. Artificial intelligence : structures and strategies for complex problem solving. 6th. Pearson Education, 2009. 779 p. Also available online, URL: http://www.uoitc.edu.iq/images/documents/informatics-institute/exam_materials/artificial%20intelligence%20structures%20and%20strategies%20for%20%20complex%20problem%20solving.pdf
15. Garvey C. Artificial Intelligence and Japan’s Fifth Generation. *Pacific Historical Review*. Vol. 88, Issue 4. P. 619–658.
16. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 4th. Pearson Education, 2020. 1136 p.
17. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. URL: <http://www.deeplearningbook.org>
18. Amato G., Falchi F., Gennaro C. et al. A Comparison of Face Verification with Facial Landmarks and Deep Features. 2018.
19. Матеріали Восьмої Всеукраїнської міжнародної конференції з оброблення сигналів і зображень та розпізнавання образів УкрОбраз’2006. 2006. URL: <http://www.uasoiro.org.ua/files/Zbirnyk/proceedind.html>

20. Березький О. М., Батько Ю. М., Березька К. М. та ін. Методи, алгоритми і програмні засоби опрацювання біомедичних зображень. Тернопіль : Економічна думка, ТНЕУ, 2017. 330 с.

21. Ткаченко Р. О., Ткаченко П. Р., Ізонін І. В. Нейромережеві засоби штучного інтелекту: навч. посібник. Львів : Видавництво Львівської політехніки, 2017. 208 с.

22. Шаховська Н. Б., Камінський Р. М., Вовк О. Б. Системи штучного інтелекту: навч. посібник. Львів : Видавництво Львівської політехніки, 2018. 392 с.

23. Dupré D., Krumhuber E. G., Küster D. та ін. A performance comparison of eight commercially available automatic classifiers for facial affect recognition. *PLOS ONE*. ред. Sidney D'Mello. Вип. 15, № 4. С. e0231968. DOI:10.1371/journal.pone.0231968.

24. microsoft/FERPlus. The FER+ annotations provide a set of new labels for the standard Emotion FER dataset. This is the FER+ new label annotations for the Emotion FER dataset. Microsoft, 2025. URL: <https://github.com/microsoft/FERPlus> (accessed 13/05/2025).2025.

25. Haykin S. S. Neural Networks. A Comprehensive Foundation. 2th. Pearson Education India, 2005. 823 p. Also available online, URL: [https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural %20Networks%20-%20A%20Comprehensive%20Foundation%20-%20Simon%20Haykin.pdf](https://cdn.preterhuman.net/texts/science_and_technology/artificial_intelligence/Neural%20Networks%20-%20A%20Comprehensive%20Foundation%20-%20Simon%20Haykin.pdf)

26. Haykin S. S. Neural Networks and Learning Machines. Pearson Education India, 2010. 944 p.

27. Simons G. L. Towards fifth-generation computers. NCC Publications, 1983. 226 p.

28. Bratko I. Prolog Programming for Artificial Intelligence. 4th. Pearson Education Canada, 2011. 696 p.

29. Deng L., Yu D. Deep Learning: Methods and Applications. *Foundations and Trends in Signal Processing*. Vol. 7, Issue 3–4. P. 197–387.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						49
Змн.	Арк.	№ докум.	Підпис	Дата		

30. Gonzalez R. C., Woods R. E. Digital Image Processing. 4th. Pearson Education Limited, 2018. 1022 p. Also available online, URL: <https://dl.icdst.org/pdfs/files4/01c56e081202b62bd7d3b4f8545775fb.pdf>

31. Березький О. М., Мельник Г. М., Дубчак Л. О. та ін. Методичні вказівки до випускних кваліфікаційних робіт освітнього рівня “Бакалавр” спеціальності “Комп’ютерна інженерія”. ред. О. М. Березький. Тернопіль : ЗУНУ, 2024. 52 с.

					КР.КІ. 10660768.00.00.000.ПЗ	Арк.
						50
Змн.	Арк.	№ докум.	Підпис	Дата		