

Вінницький національний технічний університет
Факультет інформаційних технологій та комп'ютерної інженерії
Кафедра комп'ютерних наук

Пояснювальна записка
до магістерської кваліфікаційної роботи
**на тему «Інформаційна технологія розпізнавання
емоцій людини»**

Виконав: студент 2 курсу,
групи 1КН-19м
спеціальності 122 «Комп'ютерні науки»
Зінов'єв Є. В.

Керівник: к. т. н., доц. Арсенюк І. Р.

Рецензент: к. т. н., доцент каф. ПЗ
Коваленко О. О.

Вінниця
2020

ЗАТВЕРДЖУЮ
Виконавчий директор
ТОВ «Аксиома ЮА»
Юрін О. Ю.
“ _____ ” _____ 2020 року

ЗАТВЕРДЖУЮ
Завідувач кафедри _____ КН _____
д. т. н., проф.. Яровий А. А.

(підпис)
“ _____ ” _____ 2020 року

ЗАВДАННЯ

на магістерську кваліфікаційну роботу на здобуття кваліфікації магістра зі спеціальності: 122 – «Комп'ютерні науки»

08-22.МКР.006.19.000.ПЗ

Магістранта групи 1КН-19м Зінов'єва Євгенія Вікторовича

Тема магістерської кваліфікаційної роботи: «Інформаційна технологія розпізнавання емоцій людини»

Вхідні дані: відеопотік із швидкістю зміни кадрів не менше 24 кадрів/с, кадри із роздільною здатністю від 40 × 40 пікселів до 1920 × 1080 пікселів, об'єктно-орієнтована мова програмування

Короткий зміст частин магістерської кваліфікаційної роботи:

1. Графічна: Загальний алгоритм роботи інформаційної системи розпізнавання емоцій, схема удосконаленої архітектури згорткової нейронної мережі структура інформаційної технології розпізнавання емоцій людини, UML-діаграма класів програмного засобу розпізнавання емоцій людини, головне вікно програмного засобу розпізнавання емоцій людини, приклад роботи програмного засобу розпізнавання емоцій людини.

2. Текстова (пояснювальна записка): вступ, аналіз сучасного стану розвитку інформаційних технологій розпізнавання емоцій, розробка інформаційної технології розпізнавання емоцій людини, програмна реалізація інформаційної технології розпізнавання емоцій людини, економічна частина, висновки, перелік використаних джерел, додатки.

КАЛЕНДАРНИЙ ПЛАН ВИКОНАННЯ МКР

№ етапу	Назва етапу	Термін виконання		Очікувані результати
		початок	кінець	
1	Аналіз сучасного стану розвитку інформаційних технологій розпізнавання емоцій Постановка задач дослідження			Аналітичний огляд літературних джерел, задачі досліджень, розділ 1 ПЗ
2	Розробка методу та інформаційної технології розпізнавання емоцій			Метод, інформаційна технологія, розділ 2
3	Програмна реалізація розробленої інформаційної технології, тестування та оцінка параметрів			Програмне забезпечення, розділ 3
4	Підготовка економічної частини			розділ 4
5	Апробація та/або впровадження результатів дослідження			тези доповідей/акт впровадження
6	Оформлення пояснювальної записки, графічного матеріалу та презентації			Пояснювальна записка, графічний матеріал, презентація

Консультанти з окремих розділів магістерської кваліфікаційної роботи

1. Науковий керівник _____ канд. техн. наук, доц., доц. кафедри КН
(підпис) наук. ступінь, вчене звання (посада)
 “ _____ ” _____ 20__ р. І. Р. Арсенюк
ініціали та прізвище

2. Економічна частина _____ к. е. н., доцент кафедри ЕПВМ
(підпис) наук. ступінь, вчене звання (посада)
 “ _____ ” _____ 20__ р. М. В. Бальзан
ініціали та прізвище

Дата попереднього захисту роботи “ _____ ” _____ 20__ р.

Рецензент _____ канд. техн. наук, доц., доц. кафедри ПЗ
(підпис) наук. ступінь, вчене звання (посада)
О. О. Коваленко
ініціали та прізвище

Завдання видав науковий керівник _____ канд. техн. наук, доц., доц. кафедри КН
(підпис) наук. ступінь, вчене звання (посада)
І. Р. Арсенюк
ініціали та прізвище

“ _____ ” _____ 20__ р.

Завдання отримав магістрант _____ Є. В. Зінов'єв
(підпис) ініціали та прізвище

“ _____ ” _____ 20__ р.

РЕФЕРАТ

Магістерська кваліфікаційна робота присвячена розробці інформаційної технології розпізнавання емоцій людини на основі згорткової нейронної мережі. Дана технологія забезпечує автоматизоване розпізнавання емоцій людини, що дає змогу досить точно визначити певну емоцію на обличчі людини в неперервному відеопотоці.

В ході роботи проведено аналіз предметної області розпізнавання емоцій людини. Розглянуто аналоги для розпізнавання емоцій людини. Визначено тип нейронної мережі, яку буде застосовано для реалізації розпізнавання емоцій людини. Удосконалено структуру нейронної мережі таким чином, щоб покращити якість та швидкість розпізнавання. Розроблено інформаційну технологію розпізнавання емоцій людини на основі згорткової нейронної мережі та реалізовано на мові програмування Python програмний засіб для розпізнавання емоцій людини на основі згорткової нейронної мережі.

ABSTRACT

The master's qualification work is devoted to the development of information technology for recognizing human emotions on the basis of a convolutional neural network. This technology provides automated recognition of human emotions, which allows you to qualitatively identify a particular emotion on a person's face in a video stream.

In the course of the work the analysis of the subject area of recognition of human emotions is carried out. Analogues for recognition of human emotions are considered. The type of neural network that will be used to implement the recognition of human emotions is determined. The structure of the neural network has been improved in such a way as to improve the quality and speed of recognition. The information technology of recognition of human emotions on the basis of a convolutional neural network is developed and the software for recognition of human emotions on the basis of a convolutional neural network is realized in Python programming language.

ЗМІСТ

1	АНАЛІЗ СУЧАСНОГО СТАНУ РОЗВИТКУ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ ЕМОЦІЙ.....	11
1.1	Аналіз існуючих методів розв’язання задачі розпізнавання емоцій	11
1.2	Аналіз існуючих програмних засобів розпізнавання емоцій	18
1.3	Постановка задачі розпізнавання емоцій	23
1.4	Висновок	25
2	РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ.....	26
2.1	Розробка методу розв’язання задачі розпізнавання емоцій	26
2.1.1	Обґрунтування вибору методу попередньої обробки зображення... ..	27
2.1.2	Обґрунтування вибору алгоритму отримання ознак емоції.....	30
2.1.3	Обґрунтування вибору алгоритму розпізнання емоції	33
2.2	Обґрунтування вибору архітектури згорткової нейронної мережі для розпізнавання емоцій та наведення її математичної моделі.....	35
2.3	Удосконалення архітектури згорткової нейронної мережі	44
2.4	Вибір набору даних для навчання удосконаленої нейронної мережі.....	47
2.5	Розробка алгоритму функціонування інформаційної системи для розпізнання емоцій людини	48
2.6	Розробка структури інформаційної технології розпізнавання емоцій людини.	50
2.7	Висновок	51
3	ПРОГРАМНА РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ	53
3.1	Розробка алгоритму роботи програмних засобів розпізнавання емоцій людини	53

3.2 Побудова UML-діаграми класів розробленого програмного засобу.....	56
3.3 Обґрунтування вибору мови програмування.....	59
3.4 Обґрунтування вибору нейромережевої бібліотеки.....	61
3.5 Обґрунтування вибору середовища розробки	63
3.6 Тестування розробленого програмного засобу розпізнання емоцій людини та аналіз результатів його роботи.....	66
3.7 Висновок	72
4 ЕКОНОМІЧНА ЧАСТИНА	74
4.1 Оцінювання комерційного потенціалу розробки	74
4.2 Прогнозування витрат на виконання науково-дослідної роботи та конструкторсько-технологічної роботи.....	75
4.3 Прогнозування комерційних ефектів від реалізації результатів розробки інформаційної технології	79
4.4 Розрахунок ефективності вкладених інвестицій та визначення періоду їх окупності.....	81
ВИСНОВКИ.....	84
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	86
ДОДАТКИ.....	92
Додаток А.....	92
Додаток Б	96
Додаток В.....	102

ВСТУП

Актуальність теми досліджень. Задача розпізнавання емоцій людини є досить актуальною на сучасному етапі розвитку людства. Її розв'язання є важливим для багатьох сфер життєдіяльності людей.

Так, наприклад, спостереження за емоціями може допомогти помітити людину, яка потенційно може створити проблеми на публічному заході. Аналіз емоцій людей на платформах станцій метро, у аеропортах, автовокзалах, залізничних вокзалах і т. п. допоможе визначати підозрілу поведінку та сповіщати про це відповідні органи (наприклад, про потенційну терористичну загрозу). Такі емоції, як сумнів і злість, можуть ховатися під маскою і контрастувати з тим, що говорить людина. Супермаркети та великі точки продажу також можуть отримати з цього вигоду, аналізуючи емоції людей, та використовуючи це у цільовому маркетингу та під час розв'язання задачі раціонального розміщення товарів. Розпізнавання емоцій обіцяє бути корисним і під час різноманітних опитувань, оскільки це дає можливість побачити, які саме питання (чи яка частина рекламного ролику) добре працюють і викликає емоційний відгук людей. Також, розпізнавання емоцій можна ефективно використати у медичній сфері та у сфері освіти, в тому числі і під час дистанційного навчання, яке отримало небувалої актуальності в умовах карантину внаслідок поширення коронавірусної інфекції Covid-19.

Саме тому дуже доцільною є розробка інформаційної технології, який вирішує задачу розпізнавання емоцій людини.

Зв'язок роботи з науковими програмами, планами, темами. Магістерська робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерних наук Вінницького національного технічного університету 22 К1 «Моделі, методи, технології та пристрої інтелектуальних інформаційних систем управління, економіки, навчання та комунікацій» та плану наукової та навчально-методичної роботи кафедри.

Мета та завдання дослідження. Метою дослідження магістерської кваліфікаційної роботи є підвищення точності розпізнавання емоцій людини у неперервному відеопотоці без збільшення навантаження на комп'ютер. Для досягнення поставленої мети слід розв'язати такі завдання:

- розглянути та проаналізувати існуючі програмні реалізації розв'язання задачі розпізнавання емоцій людини;
- запропонувати математичну модель для інформаційної технології розпізнавання емоцій людини;
- запропонувати етапи інформаційної технології та на їх основі розробити структуру та алгоритм роботи програмного засобу;
- виконати програмну реалізацію запропонованої інформаційної технології розпізнавання емоцій людини;
- провести тестування програмного продукту та виконати аналіз отриманих результатів.

Об'єкт дослідження – процес розпізнавання емоцій людини.

Предмет дослідження – інформаційна технологія розпізнавання емоцій людини.

Метою дослідження магістерської кваліфікаційної роботи є підвищення точності розпізнавання емоцій людини у неперервному відеопотоці без збільшення навантаження на систему комп'ютера.

Область застосування – психологія, відеонагляд, освіта.

Методи дослідження. У роботі використані такі методи наукових досліджень: методи оброблення цифрової інформації, теорія штучних нейронних мереж для реалізації інформаційної технології розпізнавання емоцій людини, методи математичної статистики для обрахунків результатів отриманих за допомогою програмного засобу, програмування на мовах високого рівня.

Наукова новизна одержаних результатів полягає в наступному: Удосконалено архітектуру згорткової нейронної мережі, що відрізняється від відомих тим, що додано проміжний згортковий шар, який дозволяє знизити

вимоги до потужності обчислювальних ресурсів без вагомих втрат точності розпізнавання емоцій.

Практичне значення одержаних результатів:

1. Розроблено алгоритм розпізнавання емоцій людини з використанням нейронних мереж.
2. Розроблено архітектуру нейронної мережі для підвищення точності розпізнавання емоцій людини на основі згорткової нейронної мережі в відеопотоці.
3. Розроблено програмний продукт розпізнавання емоцій людини з використанням нейронних мереж.

Достовірність теоретичних положень магістерської кваліфікаційної роботи підтверджується строгістю постановки задач, коректним застосуванням математичних методів під час доведення наукових положень, строгим виведенням аналітичних співвідношень, порівнянням результатів з відомими, та збіжністю результатів математичного моделювання з результатами, що отримані під час впровадження розроблених програмних засобів.

Особистий внесок здобувача. Усі результати, наведені у магістерській кваліфікаційній роботі, отримані самостійно.

Апробація результатів роботи. Результати досліджень апробовані на XLIX науково-технічній конференції підрозділів ВНТУ, а також на XII міжнародній науково-практичній конференції ІОН-2020 «ІНТЕРНЕТ-ОСВІТА-НАУКА-2020».

Публікації. За результатами досліджень опубліковано тезу доповіді науково-практичної конференції [1], тези доповіді міжнародної науково-практичної конференції [2]. Отримано свідоцтво про реєстрацію авторського права на твір – комп'ютерна програма «Нейромережевий модуль розпізнавання емоцій людини [3]»

1 АНАЛІЗ СУЧАСНОГО СТАНУ РОЗВИТКУ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ ЕМОЦІЙ

1.1 Аналіз існуючих методів розв'язання задачі розпізнавання емоцій

Виявлення розташування облич на зображенні зі складною фоновою структурою є першим і основним кроком у вирішенні задачі розпізнавання стану обличчя або розпізнавання емоційного стану людини по зображенню його обличчя.

На сьогоднішній день досягнуто значних успіхів при розв'язанні задач розпізнавання осіб на зображеннях, проте існує ряд складнощів, які істотно знижують надійність застосування сучасних методів [4]:

- Низьке розширення зображень. В результаті чого, ознаки об'єктів можуть бути погано помітні.
- Наявність складної фонової структури на зображеннях. Мається на увазі наявність на зображеннях сторонніх об'єктів, які можуть мати візуальні ознаки, схожі з шуканим об'єктом.
- Різні спотворення, отримані в процесі отримання зображень (Погані погодні умови, різні умови освітлення, шумові перешкоди).
- Кількість шуканих об'єктів на зображенні заздалегідь не відомо.

Все це вимагає застосування різних алгоритмів попередньої обробки, що в свою чергу ускладнює процес розпізнавання, робить його більш громіздким, збільшує обсяг і час обчислювальних процесів.

Тому на сьогоднішній день існує потреба в розробці методів і алгоритмів, які вирішують перераховані вище проблеми.

Існує велика кількість методів і алгоритмів розпізнавання об'єктів і їх класифікації на зображеннях зі складним фоном. Серед них можна відзначити наступні методи і алгоритми: метод порівняння з шаблоном, гістограмний аналіз характеристик яскравості зображення, метод головних компонент, метод опорних векторів, метод Віоли-Джонса, використання штучних

нейронних мереж, а також різні дескриптори і детектори локальних особливостей зображення.

Розглянемо основні методи ідентифікації емоцій людини. Одним з основних методів визначення емоцій людини є метод порівняння з шаблоном [4]. Принцип роботи даного методу полягає в порівнянні кожної області зображення з заданим шаблоном, в результаті чого визначається взаємна кореляція. Шаблон задається вручну або визначаються функцією і може відповідати цілому об'єкту або окремим компонентам об'єкту. Якщо коефіцієнт кореляції перевищує заданий поріг, то досліджувана область зображення відзначається такою, яка містить образ шуканого об'єкту.

Для роботи методу необхідні два невід'ємних компоненти – це вихідне зображення (наприклад, знімок з камери) і зображення-шаблон. Щоб виявити схожу ділянку, потрібно накласти шаблон на знімок і рухати шаблон по піксельно, зліва направо, зверху вниз. У кожному положенні зображення-шаблону на оригінальному документі обчислюється метрика, яка показує, чи є збіг чи ні.

Для кожного положення зображення-шаблону метрика заноситься в результуючу матрицю. Кожне положення шаблону в матриці містить метрику збігів. Після завершення операції порівняння, кращі співпадиння будуть знаходитися в глобальних мінімумах (при використанні методу найменших квадратів) або максимумам (при використанні методу кореляції або методу крос-кореляції).

Метод порівняння з шаблоном використовувався в основному в ранніх роботах по класифікації та розпізнавання об'єктів на зображеннях. Це були перші спроби сформулювати ознаки зображення об'єктів.

Недолік даного методу полягає в низькій швидкості роботи і високій чутливості до фотометричних та геометричних спотворень об'єктів на зображеннях.

Метод гістограмного аналізу [5] характеристик яскравості зображення ґрунтується на припущенні, що частотна характеристика області

розташування шуканого об'єкта відрізняється від інших областей зображення і має більш високу інтенсивність пікселів.

Для посилення просторових частот і придушення шумів на початковому етапі даного підходу виконується поліпшення контрасту, бінаризація або виділення кордонів на зображенні. В результаті цього, кордони об'єктів на зображенні стають більш контрастними, а фон затемнюється.

Ідея даного методу полягає в наступному: виконується сканування зображення і при цьому обчислюється середнє значення яскравості пікселів в кожному рядку матриці пікселів зображення (іноді обчислюють за стовпцями). В тому місці, де розташований шуканий об'єкт середня інтенсивність пікселів буде значно відрізнятися від інших областей зображення.

Перевага даного методу полягає в простій реалізації та високій швидкості роботи.

Недолік даних методів полягає в чутливості до будь-якої області зображення, метод не може коректно відрізнити саме емоцію людини, від інших об'єктів, а тому може спрацьовувати і на зображеннях, які навіть не містять зображення обличчя людини.

Метод головних компонент (Principal Component Analysis, PCA) [6]. Даний метод застосовується в різних областях, таких як: обробка і аналіз зображень, розпізнавання образів, стиснення даних і т. д. Метод спрямований на зменшення розмірності даних, при найменших втратах інформації, що буває дуже актуально в області аналізу і обробки зображень. В даному методі виконується лінійне ортогональне перетворення вхідного вектору X , що містить корельовані компоненти, в вектор меншого розміру Y , який містить некорельовані змінні. Некорельовані змінні називаються головними компонентами. Вектор X складається з усіх прикладів навчального набору зображень. Власні вектори отримуються при вирішенні наступного рівняння:

$$D = \Phi^T K \Phi, \quad (1.1)$$

де D – діагональна матриця власних чисел; K – матриця коваріації для вектору X – матриця власних векторів.

З матриці власних векторів Φ вибираються M найбільших власних чисел, з яких формується підматриця. Таким чином, отримаємо вираз: $y = \Phi_M^T \tilde{x}$, де $\tilde{x} = x - \bar{x}$, – нормалізований вектор з нульовим математичним очікуванням.

При виборі перших M компонент векторний простір розбивається на власний простір, що містить головні компоненти.

Спочатку обчислюються головні компоненти для навчального набору зображень об'єктів. Виконується фільтрація параметрів об'єктів, де відкидаються менш істотні з них і простір параметрів зменшується. Таким чином, об'єкти на зображенні можуть бути представлені у вигляді мінімального набору параметрів. Інші компоненти є другорядними і характеризують незначні відмінності між об'єктами і шум.

При класифікації об'єктів на зображенні обчислюється Евквідова відстань між головними компонентами досліджуваного зображення і зваженою комбінацією головних компонент навчального набору. Якщо головні компоненти зображення мають найбільш близьку відповідність з головними компонентами якого-небудь об'єкта з навчального набору, то вважається, що на зображенні міститься даний об'єкт.

Недолік даного методу полягає в низькій стійкості до різних шумових, і проєкційних спотворень об'єктів на зображеннях, що виникають при різних умовах реєстрації. Для ефективного застосування даного методу зображення об'єктів повинні бути отримані при однакових умовах.

Метод опорних векторів (SupportVectorMachine, SVM) є лінійним класифікатором і застосовується для вирішення завдань класифікації та регресійного аналізу [7].

Застосування методу SVM до вирішення задачі класифікації об'єктів на зображеннях можна описати таким чином. Дано безліч точок в просторі, які згруповані за двома класами. Для того щоб розділити безліч точок на класи

можна провести роздільну гіперплощину. З точки зору класифікації необхідно знайти оптимальну гіперплощину, відстань від якої до кожного класу буде максимальною. Чим більше відстань від роздільної гіперплощини до кожного класу, тим менше помилка класифікації. Вектори, розташовані в просторі ближче всього до гіперплощини, називаються опорними векторами.

Необхідно знайти функцію-класифікатор $f(x)$, яка приймає різні значення для векторів різних класів. Для визначення даної функції використовується навчальний набір даних. Функція, при якій очікувана помилка класифікації приймає мінімальне значення, є найбільш оптимальною функцією. Функція-класифікатор має вигляд:

$$f(x) = \text{sign}(\langle w, x \rangle + b), \quad (1.2)$$

де x – об'єкт, що належить простору; w – вектор нормалі до гіперплощини; $\langle w, x \rangle$, – скалярний добуток; b – допоміжний параметр.

Об'єкти, для яких функція $f(x)$ приймає значення рівні 1 або -1, відносяться до різних класів. Необхідно вибрати параметри w таким чином, щоб відстань до кожного класу була максимальною.

Перевагою методу SVM є те, що класифікація об'єктів виробляється на невеликому наборі даних. Завдяки цьому методу можна побудувати класифікатор, здатний мінімізувати верхню оцінку очікуваної помилки класифікації, навіть для тих об'єктів, які невідомі і їх не було в початковому наборі даних. Суть методу опорних векторів, безпосередньо в завданні виявлення емоцій, зводиться до пошуку гіперплощини в просторі ознак, яка буде кордоном між простором зображень емоцій (їх ознак) і тих зображень де емоції відсутні.

До недоліків методу можна віднести те, що в процесі класифікації використовується тільки деяка частина зразків, що знаходиться на кордонах області класифікації. Перевагою ж є те, що класифікація методом опорних векторів виробляється на невеликому загальному наборі даних.

У 2001 році було запропоновано алгоритм Віоли-Джонса [8, 9]. Даний алгоритм дозволяє виконувати пошук об'єктів на зображеннях на основі ознак Хаара.

Ознаки Хаара є прямокутні області, що складаються з декількох суміжних частин. Різні класи об'єктів мають індивідуальними ознаками, які можна виразити у вигляді розподілу ознак Хаара (рис. 1.1).

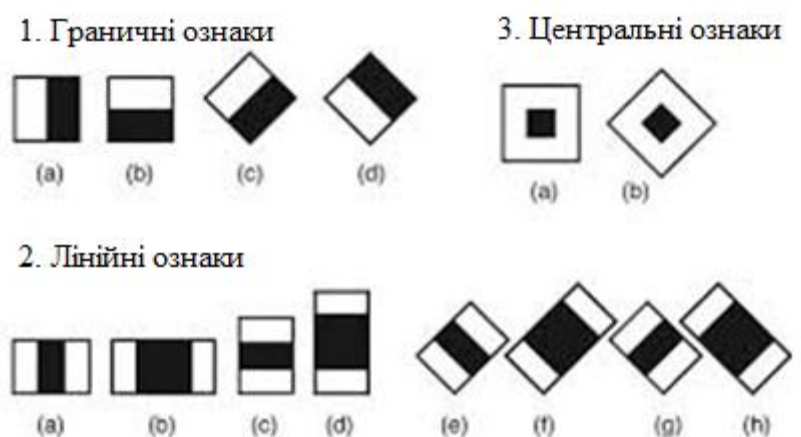


Рисунок 1.1 – Ознаки Хаара

У кожній області зображення обчислюється кілька тисяч варіантів розташування ознак, залежно від їхнього положення масштабу.

Відбувається підсумовування інтенсивностей пікселів в чорних і білих областях ознак Хаара і потім обчислюється різниця між сумами. Даний алгоритм показує досить високу точність детектування, а також високу швидкість роботи. Недоліком же є великий час навчання.

Штучна нейронна мережа (ІНС) [10, 11] являє собою математичну, програмну або апаратну модель, яка функціонує за принципом біологічних нервових клітин людського організму. ШНМ мають ієрархічну структуру і складаються з простих взаємодіючих між собою процесорів, які називаються нейронами, які згруповані в шари. Кожен нейрон має вхідні і вихідні зв'язки, що ведуть до нейронам сусідніх шарів.

Зв'язки між нейронами називаються синоптичними зв'язками і характеризуються певними ваговими коефіцієнтами, які в фізичному сенсі

можна порівняти з електричну провідність. Поточне значення нейрона визначається як зважена сума його вхідних значень:

$$S = \sum_{i=1}^N x_i w_{ij}, \quad (1.3)$$

де x_i – поточний вихід i -го нейрона; w_{ij} – ваговий коефіцієнт між i -м і j -м нейронами.

Кожен нейрон визначається функцією активації, яка призначена для визначення залежності сигналу на виході нейрона від зваженої суми сигналів на його входах. Приклад стандартної логістичної функції активації.

$$f(x) = \frac{1}{(1+e^{-x})}, \quad (1.4)$$

де S – зважена сума вхідних значень нейрона. Для всіх даних з навчальної вибірки, що надходять на вхід нейронної мережі, визначаються їх необхідні значення, які потім порівнюються з поточними значеннями і обчислюється помилка. Основний принцип навчання ШНМ полягає в зведенні до мінімуму емпіричної помилки класифікації:

$$E = \frac{1}{2} \sum_{i=0}^N (t_i - x_i)^2, \quad (1.5)$$

де x_i – дійсний вихід i -го нейрона; t_i – бажаний вихід i -го нейрона.

Після цього на основі методу градієнтного спуску обчислюється величина корекції синаптичних коефіцієнтів між нейронами:

$$\Delta w_{ij} = -\eta \delta_i x_i, \quad (1.6)$$

де δ_i – значення помилки нейрона i ; x_i – поточний вихід i -го нейрона; η – коефіцієнт швидкості навчання.

Для вирішення завдань детектування і розпізнавання об'єктів багатьма дослідниками застосовувалися ШНМ різної конфігурації.

Використання нейронної мережі дозволить значно прискорити та покращити реалізацію розпізнавання емоцій людини за рахунок можливості її навчання [12]. Потрібний продукт повинен мати змогу самоналаштуватися під потреби замовника, будучи при цьому повністю універсальним програмним забезпеченням (тобто не вимагати втручання програміста). Однак, для ефективного розв'язання поставленої задачі потрібно обрати деякий вид нейромережі, проаналізувавши існуючі варіанти реалізації останньої.

Отже, в ході аналізу існуючих методів розв'язання задачі розпізнавання емоцій було зазначено, що саме використання нейронної мережі забезпечить найбільш якісне вирішення поставленої задачі.

1.2 Аналіз існуючих програмних засобів розпізнавання емоцій

Розпізнавання емоцій людини є досить нетривіальним завданням. На сьогоднішній день для автоматизації таких дій розроблено не багато програмних аналогів [13].

Першим аналогом є (Noldus Information Technology) [14]. Цей додаток може вірно розпізнавати по виразу обличчя такі емоції, як «щастя», «смуток», «злість», «здивування», «страх», «відраза» і «спокій» (рис. 1.2). Крім того, FaceReader здатний визначити вік, стать, расу людей. Додаток не має потребу в навчанні і додаткового налаштування. У програмі реалізовані технології комп'ютерного зору. Зокрема, це метод Active Template, що полягає в накладенні деформованого шаблону на зображення обличчя:

Переваги даної реалізації:

- Середній відсоток розпізнавання емоцій дорівнює 60%, але залежить від нахилу і повороту особи площині, при русі обличчя в різні сторони відсоток точності падає до 58%.

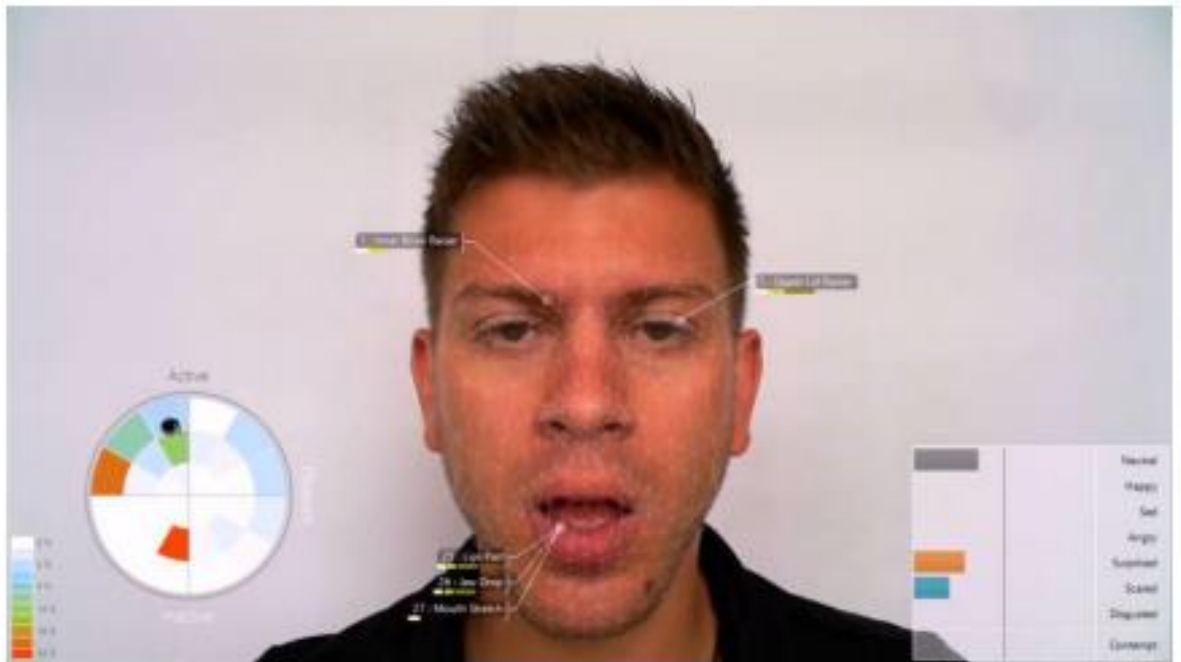


Рисунок 1.2 – Інтерфейс додатку FaceReader

– Програма працює з великою кількістю форматів відеофайлів з кодеками MPEG1, MPEG2, DivX4, DivX5, DivX6, DV-AVI [14]. При цьому розпізнавання емоцій може відбуватися покадрово, або повністю при перегляді всього відео.

- Можливості візуалізації: наявність гістограм, діаграм, відображення
- накладається сітки, відсотків висловлюються емоцій.

Недоліки:

- Відсоток розпізнавання емоцій і віку сильно падає при роботі з особами дітей віком менше 5 років.
- Визначення може відбуватися неточно, якщо людина носить окуляри.
- Програма не визначає обличчя, повернені в профіль.

EmoDetect (Нейробіотікс) [15]. Програмне забезпечення EmoDetect дозволяє визначити психоемоційний стан людини по вибірці зображень (відео або набір файлів). Розроблений класифікатор емоцій дозволяє визначити 6 базових емоцій: радість, здивування, смуток, злість, страх, відраза. Визначення емоцій відбувається за допомогою нейронних мереж: виконується

знаходження опорних точок на обличчі людини, і класифікатор будує загальну картину на основі їх становища щодо одне одного.

Компанія Нейроботікс розробляє і поставляє системи для дослідження фізіології людини і тварин і, використовуючи ці знання, розробляє нові технології в області біомеханіки, біоуправління і сенсорики у робототехніці.

Переваги:

- Класифікація емоцій трьома незалежними класифікаторами: нейронна мережа, система вирішальних правил, класифікатор по зваженій сумі ознак.
- Побудова графіка динамічного зміни інтенсивності емоцій випробуваного в часі і формування звіту про результати обробки відео.

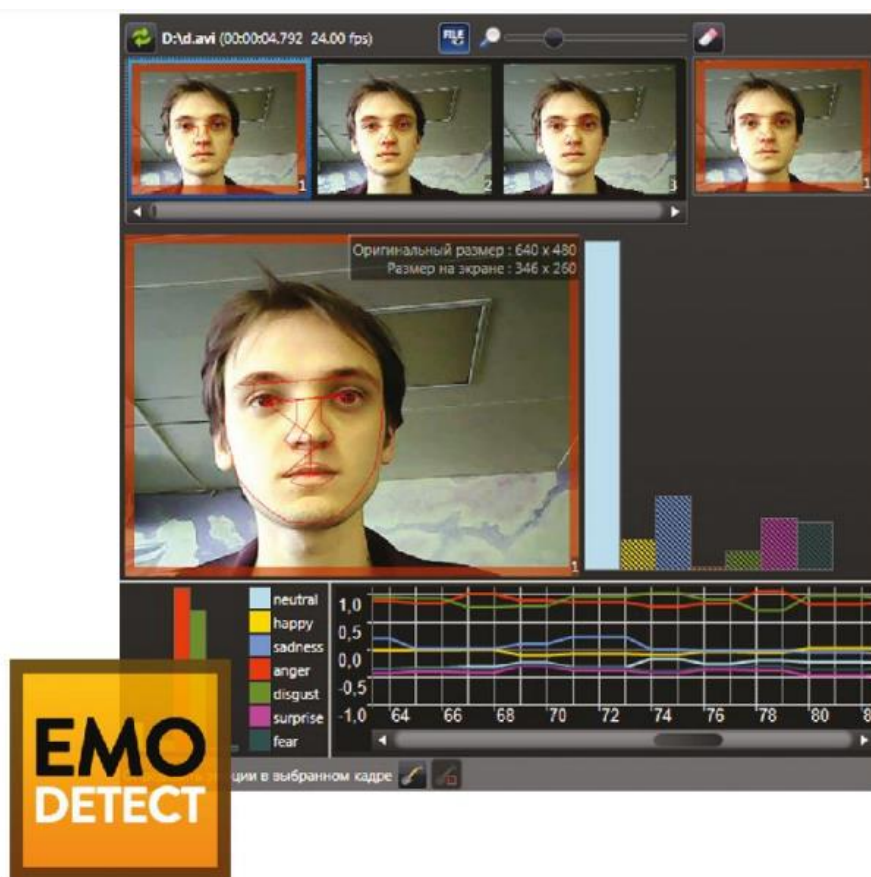


Рисунок 1.3 – Інтерфейс додатку EmoDetect

Недоліки:

- Програма не визначає обличчя, повернені в профіль.

Ще одним прикладом реалізації механізму розпізнавання емоцій є FaceSecurity (Cognitec, Німеччина)

Дана реалізація сканує потік вхідних даних і, при виявленні в кадрі осіб, порівнює їх з заведеною базою даних на предмет наявності збігів. У разі знаходження збігів система відправляє повідомлення операторам в реальному часі.

Цей додаток широко використовується в системі безпеки європейських країн для виявлення підозрілих осіб в громадських місцях.

Великі компанії ж використовують його для ідентифікації важливих клієнтів.

Рядові фірми використовують дану розробку для розпізнавання осіб своїх співробітників (рис. 1.4) і виявлення незареєстрованих відвідувачів, використовуючи бази даних з фотографіями своїх підлеглих.

Також дана розробка допомагає слідчим ідентифікувати особи в місцях злочину по фотографії і відеоспостереження шляхом зіставлення зображень облич в сховище агентств.

Переваги:

- Робота в реальному часі, можливість стеження за виявленим особою на декількох потоках даних одночасно.
- У режимі реального часу особа порівнюється з еталонами, занесеними в базу даних.
- Застосування C ++ API і Web Services API.
- Можливість масштабованості в межах «контрольного списку»: розмір, кількість відеопотоків і число видимих осіб на зображенні.

Недоліки:

- Неможливість реалізації обчислень для осіб, що мають відхилення від фронтального положення більш ніж 15 градусів.



Рисунок 1.4 – Інтерфейс додатку FaceSecurity

Ще одним прикладом є Microsoft Oxford Project Emotion Recognition (Microsoft) [16]. Принцип роботи Emotion Recognition аналогічний попередникам: машинні алгоритми аналізують наявність облич на завантаженій фотографії, після чого визначаються по міміці приблизні емоції. З огляду на тестовий режим роботи цього інструменту, результати часто виходять неточними і несподіваними. Але надання API в доступній формі дає розробникам всі можливості алгоритму машинного зору і розпізнавання емоцій для використання в своїх додатках.

Переваги:

- Обчислення відсотків відношення зображеної емоції.
- Доступно Emotion API, яке може бути використане в різних додатках.

Недоліки:

- Визначення емоцій тільки на статичних зображеннях.

- Програма не розпізнає обличчя, розташовані під великим кутом (більше 40 градусів).
- На даний момент доступна тільки демонстраційна версія.



Рисунок 1.5 – інтерфейс демонстраційної версії програми Microsoft Oxford Project Emotion Recognition

Отже в ході аналізу було виявлена потреба у створенні програмного продукту, який би надавав можливість якісно розпізнавати емоції людини у відеопотоці.

1.3 Постановка задачі розпізнавання емоцій

Протягом останніх десятиліть збільшення загальнодоступної обчислювальної потужності допомагало розвивати машини швидкого навчання, тоді як Інтернет надав величезну кількість даних для навчання. Ці дві розробки сприяли дослідженню інтелектуальних систем самонавчання, нейронні мережі – один з найбільш перспективних напрямів.

Розпізнавання облич може використовуватися для спостереження з боку правоохоронців, а також у натовпі. Інші сучасні програми передбачають автоматичне розмивання облич на кадрах Google Streetview та автоматичне розпізнавання друзів Facebook за фотографіями. Крім того, такі методи використовуються в цифрових камерах для автоматичного фотографування, коли користувач посміхається [17]. Однак найбільш перспективні розробки стосуються гуманізації штучних інтелектуальних систем. Якщо комп'ютери здатні відслідковувати психічний стан користувача, роботи можуть реагувати на це і вести себе належним чином. Отже, безперечно, розпізнавання емоцій відіграє ключову роль у поліпшенні взаємодії людини з машиною.

В ході дослідження ми зосередимось на системі, яка зможе розпізнавати емоції у неперервному відеопотоці. Крім того система матиме можливість розпізнавати емоції не лише на одному обличчі зображеному в кадрі, а і у декількох облич.

Така система є досить універсальною, і може бути доналаштована для кожного конкретного клієнта.

Модуль, що буде розроблений в ході дослідження, має бути представлений у вигляді додатку для операційної системи Windows, і мати графічний інтерфейс.

Так як для якісного тренування нейронної мережі повинен бути використаний великий набір даних, користувачу буде не раціонально створювати його самому. У процесі роботи програми користувач повинен мати змогу провести навчання нейронної мережі за допомогою набору даних, який представлений у комплекті із програмним продуктом.

Очевидно, що для роботи з програмою користувачу необхідно буде мати веб-камеру, підключену до комп'ютера. Після запуску і навчання нейронної мережі користувач натисне відповідну кнопку, буде ввімкнена веб-камера, зображення з якої буде виведене в програмі. На зображенні буде окреслено контур обличчя та написана відповідна емоція, яку програма розпізнає у відеопотоці.

Для доведення наукової актуальності необхідним буде модуль, який дозволить автоматизувати процес тестування роботи прогами на великому, відформатованому наборі даних.

Також необхідно розробити можливість валідації – тестування на невідформатованому наборі даних, що містить зображення різної роздільної здатності із можливими наявними вадами (шум) на них.

Отже, в кінцевому підсумку, розроблена інформаційна технологія забезпечить можливість на практиці показати результат проведеного в ході магістерської дипломної роботи дослідження.

1.4 Висновок

Проведено аналіз предметної області та досліджені основні методи розпізнавання емоцій. У ході аналізу було визначено, що використання нейронної мережі при розпізнанні емоцій може забезпечити високу точність і не потребує достатньо великої кількості ресурсів комп'ютера, що актуально під час розпізнавання емоцій у непевному відеопотоці.

Виконано огляд та аналіз існуючих програмних засобів, що дозволяють розпізнавати емоції людини. Більшість розглянутих програмних рішень не можуть забезпечити достатньо точне розпізнання емоції саме у неперервному відеопотоці. Деякі програмні засоби забезпечують достатньо високу точність розпізнавання, але ставлять достатньо жорсткі вимоги до формату кадру і ресурсів комп'ютерів.

Доведено доцільність розробки інформаційної системи, представленої у вигляді програмного додатку на основі математичної моделі, яка краще відтворює біологічний прототип і працює у неперервному відеопотоці не потребуючи великих ресурсів комп'ютера.

Наведено і обґрунтовано вимоги до розроблюваного програмного продукту.

2 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ

2.1 Розробка методу розв'язання задачі розпізнавання емоцій

На сьогоднішній день існує велика кількість алгоритмів [18] здатних автоматично розпізнавати емоції людини по міміці особи, що було описано раніше. Однак якість роботи систем розпізнавання емоцій за виразами обличчя погіршується через низку таких проблем:

- невелика кількість даних для навчання;
- етнічна приналежність, стать, вік;
- награність емоцій;
- оклюзія - ситуація, в якій два об'єкти розташовані приблизно на одній лінії і один об'єкт, розташований ближче до віртуальної камери або вікна перегляду частково або повністю закриває видимість іншого об'єкта;
- різний кут повороту голови;
- освітленість;
- відмінності в пропорціях особи.

Сучасні методи розпізнавання емоцій обличчя включають такі основні етапи [18]:

- попередня обробка зображення, яка складається з: знаходження області обличчя, обрізки і масштабування знайденої області, вирівнювання обличчя і регулювання контрастності;
- витяг візуальних ознак;
- класифікація емоцій.

2.1.1 Обґрунтування вибору методу попередньої обробки зображення

Попередня обробка дозволяє впоратися з описаними раніше проблемами, на даному етапі виконується локалізація області особи, обрізка і масштабування знайденої області, зміна контрастності зображення. витяг ознак, заснована на геометрії і зовнішньому вигляді обличчя. Під геометрією маються на увазі такі компоненти особи: (їх форма і розташування на обличчі), як очі, рот, ніс і т. д., а під зовнішнім виглядом особи – текстура шкіри.

Класифікація ознак спрямована на розробку відповідного алгоритму класифікації для ідентифікації виразів обличчя. Основною метою даного аналітичного огляду є порівняння методів попередньої обробки зображень обличчя, вилучення візуальних ознак і машинної класифікації емоцій, що дозволило визначити подальший напрямок дослідження для створення нової автоматичної системи розпізнавання емоцій людини по міміці особи.

Попередня обробка зображення дозволяє впоратися з такими проблемами, як недолік даних виразів обличчя, подібності, невеликі зміни зовнішнього вигляду особи, зміни пози голови, освітленості, а також підвищити точність системи розпізнавання емоцій. Попередня обробка зображень може включати в себе наступні етапи: локалізація області особи; обрізка та масштабування знайденої області; вирівнювання особи і регулювання контрастності.

1. Локалізація області обличчя дозволяє визначити розмір і місцезнаходження обличчя на зображенні. Найбільш часто використовувані методи локалізації:

- метод Віоли-Джонса (Viola-Jones object detection, VJ) [19];
- Single shot multibox detector (SSD) [20];
- гістограма спрямованих градієнтів (Histogram of oriented gradients, HOG) [21];
- Max margin object detection (MMOD) [22].

Під час аналізу методів локалізації облич більш детально необхідно зупинитись методі Віоли Джонса, оскільки він на сьогоднішній день є найпопулярнішим методом локалізації [19]. В оригінальній версії алгоритму Віоли-Джонса, наведений в пункті 1.1 використовувалися тільки примітиви без поворотів, а для обчислення значення ознаки сума яскравостей пікселів одній підобласті віднімалася з суми яскравостей іншої підобласті. У розвитку методу були запропоновані примітиви з нахилом на 45 градусів і несиметричних конфігурацій. Також замість обчислення звичайної різниці, було запропоновано приписувати кожній підобласті певне політичне значення і значення ознаки обчислювати як зважену суму пікселів різнотипних областей [23]. В основу методу лягли примітиви Хаара. Основною причиною була спроба піти від піксельного уявлення зі збереженням швидкості обчислення ознаки. Зі значень пари пікселів складно винести якусь осмислену інформацію для класифікації, в той час як з двох ознак Хаара будується наприклад, перший каскад системи з розпізнавання осіб, який має цілком осмислену інтерпретацію, наприклад як це зображено на рисунку 2.1

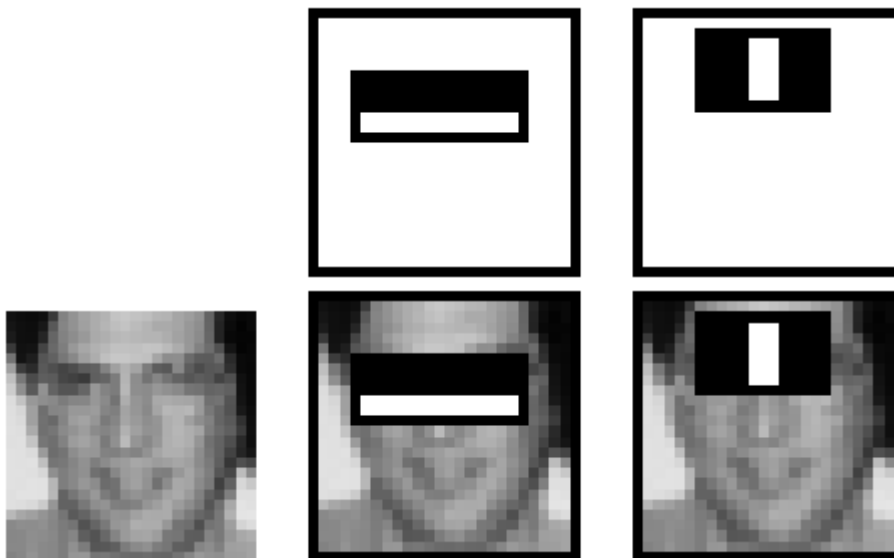


Рисунок 2.1 – Приклад використання першого каскаду Хаара системи розпізнавання осіб

Це дозволило створити швидкий алгоритм пошуку об'єктів, який користується успіхом вже більше десятиліття. Але доцільно повернутися до наведених раніше ознак. Для визначення приналежності до класу в кожному каскаді, перебувати сума значень слабких класифікаторів цього каскаду. Кожен слабкий класифікатор видає два значення в залежності від того більше або менше заданого порогу значення ознаки, що належить цьому класифікатором. В кінці сума значень слабких класифікаторів порівнюється з порогом каскаду і виносяться рішення знайдений об'єкт чи ні даними каскадом.

2. Обрізка і масштабування знайденої області особи здійснюється відповідно до координат, отриманих методами локалізації області особи. Через те, що знайдені області особи мають різний розмір, то необхідно виконувати масштабування зображення, тобто приведення всіх зображень до одного розширення. Для даних завдань застосовні [24]:

- вибірка Бесселя (Bessels correction);
- розподіл Гаусса (Gaussian distribution).

3. Вирівнювання обличчя дозволяє зменшити внутрікласові відмінності. Так, наприклад, для кожного виразу обличчя вибирається опорне зображення, яке розділяється по кольоровим компонентам або найбільш інформативним областям обличчя (наприклад лоб, очі), інші зображення вирівнюються щодо опорних зображень. для даної завдання застосовуються методи [25, 26]:

- масштабне інваріантне перетворення об'єктів (Scale-invariant feature transform, SIFT);
- області інтересу (Region of interest, ROI).

4. Регулювання контрастності дозволяє згладжувати зображення, зменшувати шум, підвищувати контрастність зображення обличчя і покращувати насиченість, що дозволяє впоратися, наприклад, з проблемою освітленості. Методами регулювання контрастності є [27]:

- вирівнювання гістограми (Histogram equalization, HE);
- лінійне контрастне розтягнення (Linear contrast stretching).

Різні маніпуляції із зображеннями, такі як обертання або зміщення, дозволяють збільшити різноплановість зображень для покращення розпізнання. Розширення бази даних варіативними зображеннями корисно для методів на основі глибокого навчання.

Тоді як для традиційних методів корисніше використовувати моделі на основі вирівнювання особи, які, навпаки, зменшують варіативність, пов'язану зі змінами пози голови, що дозволяє зменшувати внутрікласові відмінності і збільшити міжкласові подібності. У цьому випадку для кожного класу підбирається свій еталон, за яким відбувається вирівнювання зображень кожного класу. Підбір правильних методів попередньої обробки займає багато часу, так як від цього залежить швидкість і точність розпізнавання осіб.

Отже, основною задачею попередньої обробки зображення є локалізація області обличчя. Вхідне зображення буде відформатовано до необхідної роздільної здатності методом обрізки і масштабування. Для локалізації області обличчя передбачається використовувати метод каскаду Хаара, яка дозволить виділити краю об'єктів за допомогою лицьових орієнтирів, даний метод показує високу швидкість і точність виявлення області обличчя.

2.1.2 Обґрунтування вибору алгоритму отримання ознак емоції

Наступним етапом розпізнання емоцій є отримання ознак емоції. На даному етапі здійснюється знаходження елементів, які є найбільш інформативними для подальшої обробки. Залежно від виконуваних функцій методи вилучення інформативних візуальних ознак поділяються на кілька основних типів. Детально кожен метод вилучення ознак описаний далі.

1. Методи на основі геометричних об'єктів дозволяють витягати інформацію про геометричні об'єктах, таких як рот, ніс, брови та інші об'єкти, і визначати їх місце розташування. Маючи інформацію про геометричні об'єкти і їх розташування, можна розрахувати відстань між об'єктами; отримані відстані і координати положень об'єктів є ознаками для подальшої

класифікації емоцій. До методів на основі геометричних об'єктів відносяться [28]:

- дескриптор line edge map (LEM) визначає подібності виразів обличчя, ґрунтуючись на гранях об'єктів;
- активна модель форми (Active shape model, ASM), виявляє край об'єкту за допомогою лицьових орієнтирів, які представляють собою ланцюжок послідовностей точок ознак;
- активна модель зовнішнього вигляду (Active appearance model, AAM) – розширений варіант ASM, який також формує текстурні особливості виразів обличчя.
- нечіткі функції приналежності (Fuzzy membership functions), запропоновані в, витягають об'єкти особи, представляючи їх у вигляді чотирикутних форм і порівнюючи з чотирикутними геометричними фігурами, ступінь неточності представляється у вигляді дванадцяти нечітких функцій;
- дескриптор SIFT визначає потенційні точки інтересу на зображенні за допомогою визначення максимумів і мінімумів для функції гауссових фільтрів, фільтри застосовуються в різних масштабах і поворотах.
- курвлет-перетворення (Curvelet Transform) передає інформацію про місцезнаходження об'єкта і про просторової частоті.

2. Методи на основі моделей зовнішнього вигляду дозволяють витягати інформацію про текстурні особливості обличчя. Наприклад, різну кількість зморшок в області очей, говорить про різні вирази обличчя, тому застосування методів на основі моделей зовнішнього вигляду є найбільш інформативним для задач класифікації емоцій. До методів на основі моделей зовнішнього вигляду відносяться [29]:

- фільтр Габора (Gabor filter) є класичним методом виділення рис виразів осіб, який дозволяє виділити різні моделі деформації для кожної емоції.

- локальний бінарний шаблон (Local binary patterns, LBP) дозволяє уявити околиці пікселів зображення в двійковому коді, містять відмінні риси як локальних, так і глобальних (текстурних) областей виразів обличчя;

- локальне фазовий квантування (Local phase quantisation, LPQ) стійкий до розмиття зображень, заснований на короткочасному перетворенні Фур'є, яке дозволяє виявити періодичні компоненти в зображеннях виразів обличчя та оцінити їх внесок у формування вихідних даних.

- локальний дескриптор Вебера (Weber local descriptor) витягує ознаки у два етапи, перший етап розділяє зображення на локальні ділянки (рот, ніс і т. д.) і нормалізує зображення, другий етап витягує відмінні текстурні ознаки, використовуючи орієнтацію градієнта, що описує виразу обличчя.

3. Методами на основі глобальних і локальних об'єктів є:

- метод головних компонент (Principal component analysis, PCA) витягує відмінні ознаки виразів обличчя з коваріаційної матриці, зменшуючи розмірність векторів. Коваріаційна матриця - матриця, елементами якої є попарні лінійні залежності двох випадкових величин;

- лінійний дискримінантний аналіз [30] (Linear discriminant analysis, LDA) шукає вектори, які мають кращі відмінності між класами і групує ознаки одного і того ж класу.

- оптичний потік (Optical flow, OF) присвоює кожному пікселю вектор швидкості, так витягується інформація про рух м'язів обличчя з послідовності зображень, що дозволяє враховувати деформацію особи в динаміці.

З представлених методів вилучення ознак, методи на основі зовнішнього вигляду є більш корисними методами, так як вони дозволяють витягати текстурні особливості зовнішнього вигляду, які є важливими параметрами для РВЛ, проте вони менш адаптовані до оклюзії. Методи на основі геометричних

об'єктів краще адаптовані до оклюзії, відстань між лицьовими орієнтирами більше характеризують міжкласові відмінності. Використання гібридних методів для вилучення ознак виразів обличчя збільшує точність класифікації.

Отже, ході дослідження для отримання ознак буде використано алгоритм масштабування та фільтр Габора. Масштабування зображень буде здійснюватися з дозволом 48×48 , 64×64 , 128×128 , 224×224 пікселів, що дозволить зробити висновок про вплив дозволу на точність розпізнавання виразу обличчя. Фільтр Габора дозволить якісно визначити ознаки емоції людини.

2.1.3 Обґрунтування вибору алгоритму розпізнавання емоції

Класифікація є останньою стадією розроблюваному методі розпізнавання емоцій людини. На цій стадії здійснюється класифікація отриманих у попередньому етапі ознак на вирази обличчя: щастя, здивування, гнів, страх, відраза, сум і нейтральність.

Методи машинної класифікації емоцій діляться на традиційні методи і штучні нейронні мережі.

Детально кожен метод машинної класифікації емоцій розглянутий нижче.

1. Традиційні методи [31]:

- відстань Хаусдорфа (Hausdorff Distance) дозволяє вимірювати відстань між областями інтересу і порівнювати отримані відстані між класами;
- алгоритм k -найближчих сусідів (k -nearest neighbors algorithm, KNN) класифікує емоції особи за першими k -елементами, відстань між якими мінімальна;
- LDA представляє матричні дані у вигляді діаграми розсіювання, що дозволяє візуально розділяти ознаки класів;
- метод опорних векторів (Support vector machine, SVM) [32] будує гіперплощину, що розділяє об'єкти вибірки, чим більше відстань між

розділяючою гіперплощиною і об'єктами розділяючих класів, тим менше буде середня помилка класифікатора.

2. Штучні нейронні мережі:

- багатошаровий перцептрон (МП, Multilayer perceptron, MLP) має три шари, а саме: вхідний шар, прихований шар і шар обробки, кожен шар містить нейрони, які мають своє унікальне вагове значення. Даний вид перцептрону називається багатошаровим, тому що його прихований (той, якого навчають) шар може складатися з декількох шарів.

- глибокі нейронні мережі (Deep neural network, DNN) [33]:

- багатошарова пряма нейронна мережа (Multilayer feed forward neural network, MLFFNN) представляє собою взаємозв'язок перцептронів, де кількість шарів в нейронній мережі є кількістю перцептронів.

- згорткова нейронна мережа [34] (Convolutional neural network, CNN) – метод класифікації, який має три шари, а саме: шар згортки, шар підвибірки і повнозв'язний шар. CNN вловлює текстури на невеликих ділянках зображень. CNN навчається на нерухомих зображеннях, але досить успішно використовується і при виявленні емоції у відеопотоці.

- рекурентна нейронна мережа (Recurrent neural network, RNN) використовує контекстну інформацію про попередні зображення. Так, один навчальний набір містить послідовність зображень, класифікація всього набору буде відповідати класифікації останнього зображення.

Існують і інші глибокі нейронні мережі, які є модифікаціями представлених нейронних мереж.

Отже, для класифікації емоцій за отриманими масивами пікселів з різним розширенням планується використання згорткової нейронної мережі. Архітектура і принцип роботи загорткових нейронних мереж найбільш вдало

відтворює роботу людського мозку при розпізнанні різних об'єктів, зокрема емоцій.

2.2 Обґрунтування вибору архітектури згорткової нейронної мережі для розпізнавання емоцій та наведення її математичної моделі

З появою великих обсягів даних і великих обчислювальних можливостей стали активно використовуватися нейронні мережі. Особливу популярність отримали згорткові нейронні мережі, архітектура яких була запропонована Яном Лекуном [33] і націлена на ефективне розпізнавання зображень. Своєю назву архітектура мережі отримала через наявність операції згортки, суть якої в тому, що кожен фрагмент зображення множиться на матрицю (ядро) згортки поелементно, а результат підсумовується і записується в аналогічну позицію вихідного зображення. В архітектуру мережі закладені апріорні знання з предметної області комп'ютерного зору: піксель зображення сильніше пов'язаний з сусіднім (локальна кореляція) і об'єкт на зображенні може зустрітися в будь-якій частині зображення.

Згорткові мережі – нейронні мережі, в яких замість спільної операції множення на матрицю, хоча б в одному шарі, використовується згортка. У загальному вигляді згортка – операція над двома функціями дійсного аргументу. Функцію згортки можна подати у вигляді [34]:

$$S(t) = (x \cdot w)(t). \quad (2.1)$$

У термінології згорткових мереж перший аргумент (у нашому прикладі функція x) називається входом, а другий (функція w) – ядром. Вихід $S(t)$ називається картою ознак.

Оскільки під час розпізнавання емоцій на вхід буде подано зображення, представлене у вигляді двовимірної матриці, формула (2.1) набуде такого вигляду:

$$S(i, j) = (I \cdot K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n), \quad (2.2)$$

де I – вхідне зображення, K – ядро, i, j – координати елементів вхідного зображення, m, n – координати елементів ядра.

Успіх застосування згорткових нейронних мереж до класифікації зображень привів до безлічі спроб використовувати даний метод до інших місій. Останнім часом їх стали активно використовуватися для завдання класифікації емоцій.

Згорткова нейронна мережа, окрім вхідного і вихідного шару, зазвичай являє собою чергування згорткових шарів (convolution layers), субдескриптуючих шарів (subsampling layers) і при наявності повнозв'язних шарів (fully-connected layer) на виході. Всі три види шарів можуть чергуватися в довільному порядку [34]. У згортковому шарі нейрони, які використовують одні і ті ж ваги, об'єднуються в карти ознак (feature maps), а кожен нейрон карти ознак пов'язаний з частиною нейронів попереднього шару. При обчисленні мережі виходить, що кожен нейрон виконує згортку деякої області попереднього шару (яка визначається безліччю нейронів, пов'язаних з даними нейроном).

Приклад архітектури згорткової нейронної мережі наведений на рисунку.
2.2.

Нехай буде розглянуто основні шари, які використовують при побудові архітектур згорткових нейронних мереж для розпізнавання емоцій людини, та описано їх математичні моделі [34].

Вхідні дані являють собою кольорові зображення типу JPEG, розміру 48x48 пікселів. Якщо розмір буде занадто великий, то обчислювальна

складність підвищиться, визначення розміру в даній задачі вирішується методом підбору. Якщо вибрати розмір занадто маленький, то мережа не зможе виявити ключові ознаки осіб. Кожне зображення розбивається на 3 канали: червоний, синій, зелений. Таким чином виходить 3 зображення розміру 48x48 пікселів.

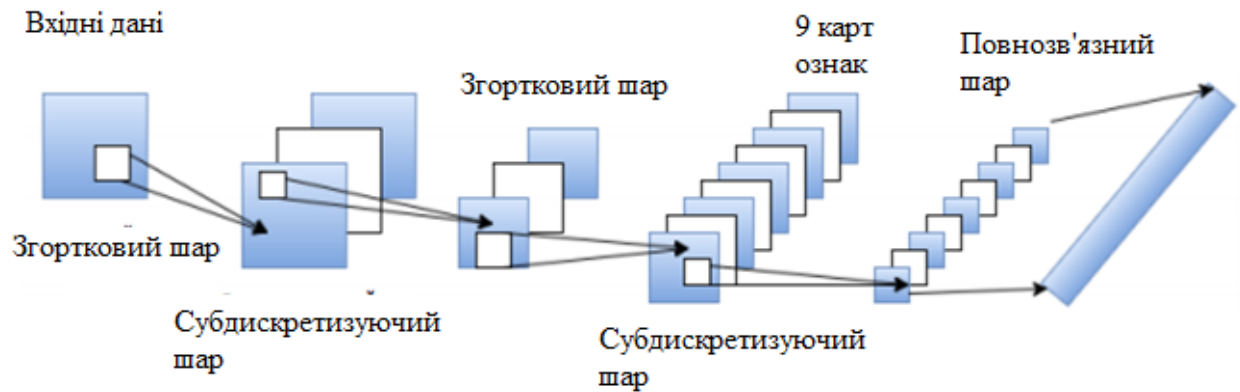


Рисунок 2.2 – Архітектура згорткової нейронної мережі

Вхідний шар враховує двовимірну топологію зображень і складається з декількох карт (матриць), карта може бути одна, в тому випадку, якщо зображення представлено в відтінках сірого, інакше їх 3, де кожна карта відповідає зображенню з конкретним каналом (червоним, синім і зеленим).

Вхідні дані кожного конкретного значення пікселя нормалізуються в діапазон від 0 до 1, за формулою 2.3:

$$f(p, min, max) = \frac{p - min}{max - min}, \quad (2.3)$$

де f – функція нормалізації, p – значення конкретного пікселя від 1 до 255, min – мінімальне значення пікселя 0, max – максимальне значення пікселя 255.

Згортковий шар – являє собою набір карт, (інша назва – карти ознак) у кожній карті є синоптичне ядро. Розмір у всіх карт згорткового шару – однакові і визначаються за формулою 2.4:

$$(w, h) = (mW - kW + 1, mH - kH + 1), \quad (2.4)$$

де (w, h) – обчислювальний розмір згорткової карти, mW – ширина попередньої карти; mH – висота попередньої карти, kW – ширина ядра, kH – висота ядра.

Ядро являє собою фільтр або вікно, яке ковзає по всій області попередньої карти і знаходить певні ознаки об'єктів. Наприклад, якщо мережу навчали на великій кількості зображень обличчя, то одне з ядер могло б в процесі навчання видавати найбільший сигнал в області очей, рота, брів або носа, інше ядро могло б виявляти інші ознаки емоції людини. Розмір ядра зазвичай беруть в межах від 3×3 до 7×7 . Якщо розмір ядра маленький, то воно не зможе виділити будь-які ознаки, якщо занадто велике, то збільшується кількість зв'язків між нейронами. Також розмір ядра вибирається таким, щоб розмір карт згорткового шару був парним, це дозволяє не втрачати інформацію при зменшенні розмірності в субдискретизуючому шарі, описаному нижче.

Ядро являє собою систему поділюваних ваг або синапсів, це одна з головних особливостей згорткової нейронної мережі. У звичайній багатосаровій мережі дуже багато зв'язків між нейронами, тобто синапсів, що вельми уповільнює процес детектування. У згортковій мережі – навпаки, загальні ваги дозволяють скоротити число зв'язків і дозволити знаходити одну і ту саму ознаку по всій області зображення.

Спочатку значення кожної карти згорткового шару рівні 0. Значення ваг ядер задаються випадковим чином в області від -0.5 до 0.5. Ядро ковзає по попередній карті і виробляє операцію згортки, яка часто використовується для обробки зображень, це описує формула 2.5:

$$(f * g)[m, n] = \sum_{k, l} f[m - k, n - l] * g[k, l], \quad (2.5)$$

де f – вихідна матриця зображення, g – ядро згортки.

На відміну від повнозв'язного, в згортковому шарі нейрон з'єднаний лише з обмеженою кількістю нейронів попереднього рівня, тобто, згортковий шар аналогічний застосуванню операції згортки, де використовується лише матриця ваг невеликого розміру (ядро згортки), яку ковзають по всьому оброблюваному шару.

Неформально цю операцію можна описати таким чином - вікном розміру ядра g проходимо з заданим кроком (зазвичай 1) все зображення f , на кожному кроці поелементно множимо вміст вікна на ядро g , результат підсумовується і записується в матрицю результату, як на рисунку 2.4.

При цьому в залежності від методу обробки країв вихідної матриці результат може бути менше вихідного зображення (*valid*), такого ж розміру (*same*) або більшого розміру (*full*). У спрощеному вигляді згортковий шар можна описати формулою 2.6:

$$x^l = f(x^{l-1} * k^l + b^l), \quad (2.6)$$

де x^l – вихід шару l , $f()$ – функція активації, b^l – коефіцієнт зсуву шару l , $*$ – операція згортки входу x з ядром k .

При цьому за рахунок крайових ефектів розмір вихідних матриць зменшується, що описує формула 2.7:

$$x_j^l = f(\sum_i x_i^{l-1} * k_j^l + b_j^l), \quad (2.7)$$

де x_j^l – карта ознак j (вихід з шару l), $f()$ – функція активації, b_j^l – коефіцієнт зсуву шару l для карти ознак j , k_j^l – ядро згортки карти ознак j , шару l , $*$ – операція згортки входу x з ядром k .

На рисунку 2.3 наведена схема роботи згорткового шару з ядром згортки розміру 3×3 .

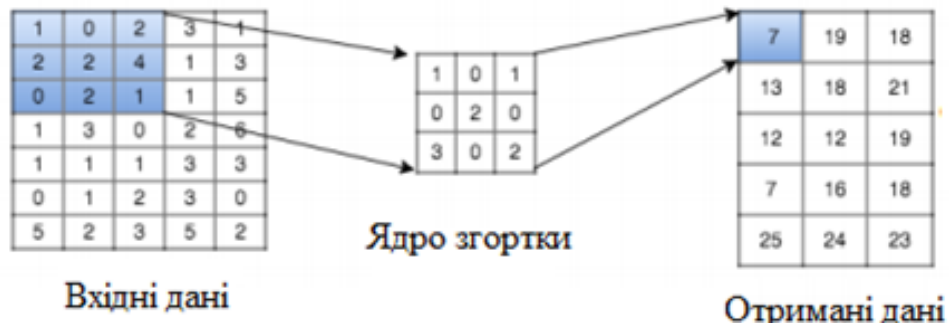


Рисунок 2.3 – Схема роботи згорткового шару

Субдискретизуючий шар. Субдискретизуючий шар також, як і згортковий має карти, але їх кількість співпадає з попереднім (згортковим) шаром. Мета шару - зменшення розмірності карт попереднього шару. Якщо на попередній операції згортки вже були виявлені деякі ознаки, то для подальшої обробки настільки докладне зображення вже не потрібно, і воно ущільнюється до менш докладного. До того ж фільтрація вже непотрібних деталей допомагає не перенавчатися.

Формально шар може бути описаний формулою :

$$x^l = f(a^l * \text{subsample}(x^{l-1}) + b^l), \quad (2.8)$$

де x^l – вихід шару l , $f()$ – функція активації, a^l, b^l – коефіцієнти зсуву шару l , $\text{subsample}()$ – операція вибірки локальних максимальних значень.

У процесі сканування ядром субдискретизуючого шару (фільтром) карти попереднього шару, субдискретизуюче ядро не перетинається на відміну від згорткового шару. Тобто, шари цього типу виконують зменшення розмірності (зазвичай в кілька разів). Це можна робити різними способами, але найчастіше використовується метод вибору максимального елемента - вся карта ознак поділяється на осередки, з яких вибираються максимальні за значенням. На рисунку 2.4 зображена схема роботи субдискретизуючого шару з методом вибору максимального елемента.

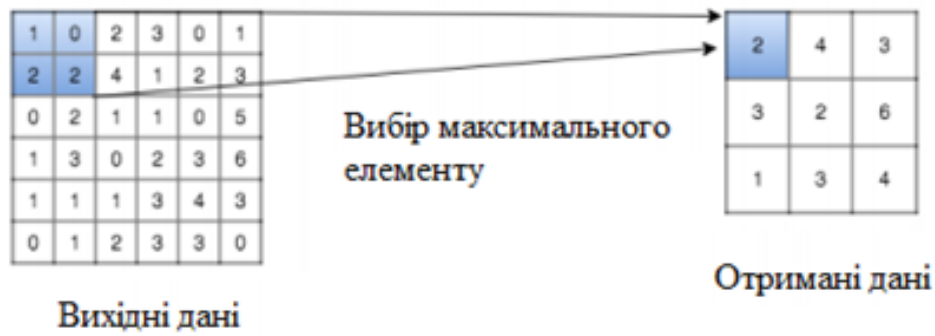


Рисунок 2.4 – схема роботи субдискретизуючого шару

Повнозв'язний шар - шар в якому кожен нейрон з'єднаний з усіма нейронами на попередньому рівні, причому кожен зв'язок має свій ваговий коефіцієнт. Мета шару - класифікація, моделює складну нелінійну функцію, оптимізуючи яку, поліпшується якість розпізнавання. Нейрони кожної карти попереднього субдискретизуючого шару пов'язані з одним нейроном прихованого шару. Таким чином число нейронів прихованого шару дорівнює числу карт субдискретизуючого шару, але зв'язки можуть бути не обов'язково такими, наприклад, тільки частина нейронів будь-якої з карт субдискретизуючого шару зв'язана з першим нейроном прихованого шару, а частина залишилася з другим, або всі нейрони першої карти пов'язані з нейронами 1-го і 2-го прихованого шару. Обчислення значень нейрону можна описати формулою:

$$x_j^l = f(\sum_i x_i^{l-1} * w_{i,j}^{l-1} + b_j^{l-1}), \quad (2.9)$$

де x_j^l – карта ознак j (вихід з шару l), $f()$ – функція активації, b_j^l – коефіцієнт зсуву шару l для карти ознак j , $w_{i,j}^l$ – матриця вагових коефіцієнтів шару l .

На рисунку 2.5 показана схема роботи повнозв'язного шару.

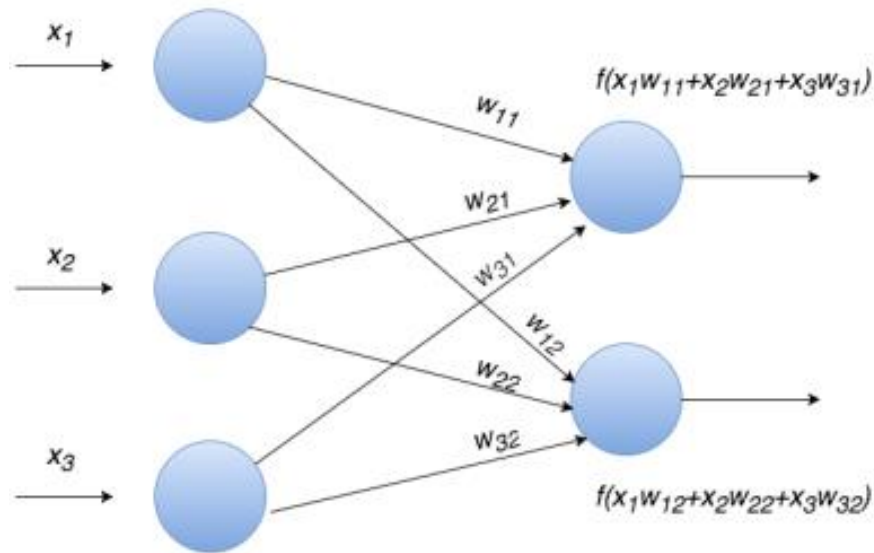


Рисунок 2.5 – Схема роботи повнозв'язного шару

Вихідний шар пов'язаний з усіма нейронами попереднього шару. Кількість нейронів відповідає кількості розпізнаваних класів, тобто кількості основних емоцій.

На основі описаних шарів та їх математичних моделей побудовані основні архітектури згорткових нейронних мереж, зокрема, для розпізнавання емоцій. Розглянемо три основні архітектури згорткових нейронних мереж, які найчастіше використовують для розпізнання емоцій людини.

(А) Перша розглянута мережа для вибору і подальшої роботи базується на наведених дослідженнях Крижевського та Хінтона [35]. Це найпростіша мережа (з трьох розглянутих), і вона потребує найменших обчислювальних вимог. Варто зауважити, що оскільки розроблювальний додаток може бути представлений у вигляді розпізнавання емоцій в реальному часі у вбудованих системах, алгоритми швидкої роботи є дуже бажаними. Дана мережа складається із трьох згорткових шарів та двох повнозв'язних шарів, для зменшення розміру зображення, та шаром проріджування для зменшення ймовірності перенавчання. Гіперпараметри вибираються такими, що кількість обчислень у кожному згортковому шарі залишається приблизно однаковою (забезпечує збереження інформації по всій мережі). Навчання

здійснюється за допомогою різного числа згорткових фільтрів, щоб оцінити їх вплив на результативність.

(B) У 2012 році була розроблена згорткова мережа AlexNet для класифікації зображень у більш ніж 1000 різних класах, використовуючи 1,2 мільйона зразкових зображень із набору даних ImageNet [36]. Внаслідок того, що у цьому дослідженні модель має розрізняти сім основних емоцій, а завдяки обмеженим обчислювальним ресурсам розмір оригінальної мережі вважається занадто великим, що може негативно відобразитись на продуктивності обробки зображення, особливо в умовах відеопотоку, тому автори спробували зменшити кількість згорткових шарів з п'яти до трьох. В отриманих трьох повнозв'язних шарах кількість вузлів кожного повнозв'язного шару також було зменшено з 4096 до 1024. Хоча початкова мережа була оптимізована для паралельних тренувань, було відзначено, що це не було необхідно для простішої її версії. Мережа також використовує локальну нормалізацію для прискорення шарів тренувань та відсіву, щоб зменшити ступінь перенавчання.

(C) Ще ряд експериментів було проведено на основі мережі, наведеної у роботі Гуді [37]. Оскільки це дослідження також мало на меті розпізнати сім емоцій за допомогою набору даних FEREC-2013. Початкова мережа починається з шару введення розмірністю 48×48 , що відповідає розміру вхідних даних. За цим шаром роташований один згортковий шар. Взагалі дана мережа містить два згорткових шари та одним повнозв'язний шар, з'єднаний з вихідним шаром. Метод проріджування було застосовано до повністю пов'язаного шару, і весь шар містить блоки ReLu.

Архітектура згорткової нейронної мережі (C) виходячи з дослідженні [37] показала досить високу точність розпізнання емоцій .при цьому зменшивши навантаження на ресурси комп'ютера. Отже, архітектура повинна стати гарною відправною точкою для побудови інформаційної системи для розпізнання емоцій.

2.3 Удосконалення архітектури згорткової нейронної мережі

В якості бази для побудови інформаційної системи для розпізнання емоцій людини було використано архітектуру (С), описану вище. Було вирішено застосувати другий проміжний згортковий шар для зменшення кількості параметрів. Це дозволить знизити вимоги до потужності обчислювальних ресурсів, здатних реалізувати дану мережу [36]. Крім того, покращується швидкість навчання без вагомих втрат точності розпізнавання емоцій. Таким чином під (С) мається на увазі удосконалена архітектура згорткової нейронної мережі. Схема удосконаленої архітектури нейронної мережі (С) зображена на рисунку 2.5.

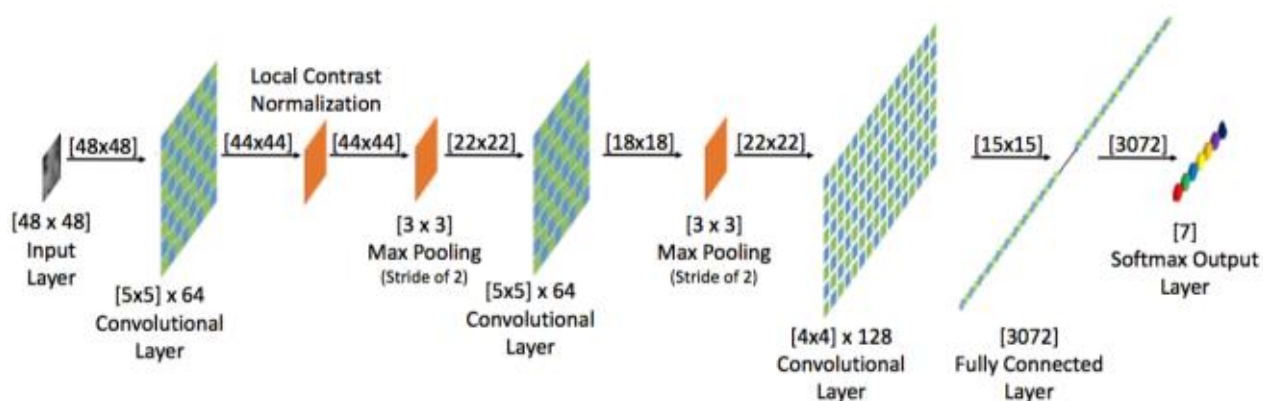


Рисунок 2.6 – Схема удосконаленої архітектури нейронної мережі

Усі мережі (А) – (С) пройшли навчання протягом 60 епох. На рисунку 1 наведено різні деталі процесу навчання [36]. Для мережі (А), точність даних під час валідації становить близько 63%. Вже через 10 епох точність піднялася вище 60%, що свідчить про можливість швидкого навчання. Крім того, варто зазначити, що коригування розміру фільтра не мало великого впливу на точність, хоча це і не вплинуло на час обробки. Це означає, що швидкі моделі можна реалізувати з досить високою продуктивністю. Важливо зазначити, що значно складніша мережа (В), також швидко вчиться, але з точністю до 54%. Таким чином видно, що зменшення розміру мережі суттєво знизили показники

точності оригінальної мережі (B). Разом із значно вищими вимогами до потужності обчислювальних ресурсів, а отже, і повільнішими показниками роботи, ця модель не є достатнім викликом двох інших архітектур. Мережа (C) показує дещо повільнішу криву навчання, але показники точності розпізнавання емоцій, під час тестування на наборі валідацій, аналогічні відповідним показникам мережі (A). Вимоги до обробки вхідних зображень не такі жорсткі як у мережі (B), але жорсткіші за мережу (A). Тому виходячи з цього факту, мережа (A) здається найбільш перспективною для нашого завдання розпізнавання емоцій. Однак продуктивність мережі (C) на додатковому тестовому наборі RaFD суттєво краща (60%), ніж у мережі (A) (50%). Це свідчить про кращі узагальнюючі можливості, що дуже важливо для розробки на її основі нашого програмного додатку.

На рисунку 2.6 зображене порівняння

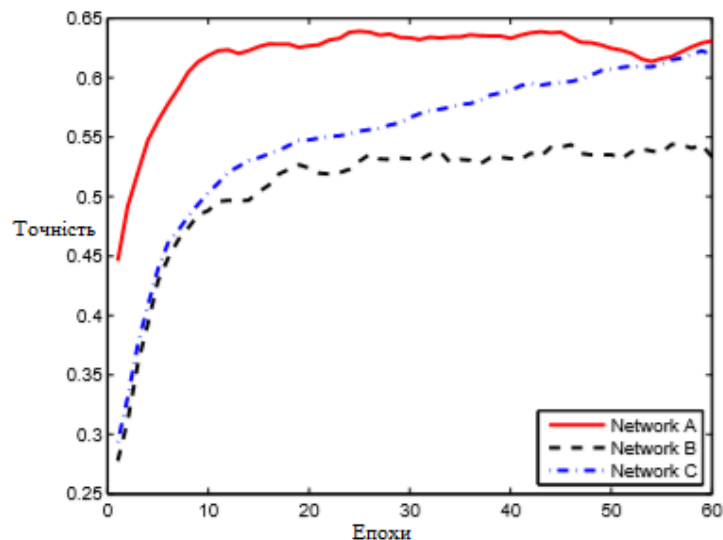


Рисунок 2.7 – Порівняння точності нейронних мереж (A) – (C)

Метод навчання та цільову функцію буде використано такі ж як і в основоположній мережі (C). Навчання мережі проводиться з використанням стохастичного градієнтного спуску з імпульсом в пакетному режимі з партіями по 100 зразків даних.

Стохастичний градієнтний спуск (англ. stochastic gradient descent, incremental gradient descent) — ітеративний метод оптимізації градієнтного

спуску за допомогою стохастичного наближення [38]. Використовується для прискорення пошуку цільової функції шляхом використання обмеженого за розміром тренувального набору, який вибирається випадкового при кожній ітерації.

Існує 2 підходи в реалізації градієнтного спуску [38]:

- Пакетний (batch), коли на кожній ітерації навчальна вибірка переглядається цілком, тільки після чого змінюється w . Це потребує великих обчислювальних затрат.
- Стохастичний (stochastic/online), коли на кожній ітерації алгоритму з навчальної вибірки випадковим чином обирається лише один об'єкт. Таким чином вектор w налаштовується кожен раз на новобраний об'єкт.

Вибрана функція активації = функція ReLU. Rectified Linear Unit – це найчастіше використовувана функція активації при глибокому навчанні. Данна функція повертає 0, якщо приймає заперечний аргумент, у випадку, якщо позитивний аргумент, функція повертає само число. Тож вона може бути записана як $f(z) = \max(0, z)$. При першому огляді може бути показано, що вона лінійна і має те проблеми, що і лінійна функція, але це не так і її можна використовувати в нейронних сетях з багатьма шарами. Функція ReLU надає кілька переваг перед сигмоїдовою та гіперболічним тангенсом:

- Дуже швидко і просто вважається похідна. Для від'ємних значень буде 0, для позитивних – буде 1.
- Розрідженість активації. У мережах з дуже великою кількістю нейронів використання сигмоїдної функції або гіперболічного тангенса як активаційної функції тягне активацію багатьох нейронів, що може позначитися на продуктивності навчання моделі. Якщо ж використовувати ReLU, то кількість включаються нейронів стане менше, в силу характеристик функції, і сама мережа стане легше.

Швидкість навчання для тренінгу ініціалізується як 0,0025, і лінійно зменшується до 0,001 протягом 50 епох навчання.

Отже, враховуючи вище проаналізоване, остання, модифікована мережа (С) виявилася найперспективнішою для подальших практичних застосувань при побудові інформаційної системи для розпізнання емоцій людини.

2.4 Вибір набору даних для навчання удосконаленої нейронної мережі

Нейронні мережі та, зокрема, глибокі мережі відомі своєю потребою у великих обсягах навчання даних. Більше того, вибір зображень, використаних для тренувань, відповідає за більшу частину вистави можливої моделі. Це означає потребу в якісному, і кількісному наборі даних. Для розпізнавання емоцій у відкритому доступі наявно кілька наборів даних, що варіюється від кількох сотень фотографій із високою роздільною здатністю до десятків тисяч менших зображень.

Три, які ми обговоримо – це Facial Expression Recognition Challenge (FERC-2013) [39], Розширений CohnKanade (СК) [40] та Radboud Faces Database (RaFD) [41], приклади зображень з наведених наборів даних показано на рисунку 2.8.

Крім того, можна помітити, що на обличчі висловлюються вирази в СК та RaFD (тобто 'чистий', або форматований для роботи), тоді як набір FERC-2013 демонструє емоції при довільній поведінці людей.



Рисунок 2.8 – Приклади зображень з наборів даних FERC-2013 (ліворуч), СК(по-центру) та RaFD (праворуч)

Це робить знімки з FERCC2013 складнішими для інтерпретації, але з огляду на великий розмір набору даних різноманітність може бути корисною для точності моделі.

В ході дослідження було вирішено, що після навчання на FERCC-2013 набір зображень із "чистих" наборів даних можна легко класифікувати, але не навпаки. Отже, для удосконаленої мережі навчання проводитиметься з використанням 29000 зразків за даними FERCC-2013 та ще 9000 невідформатованих зразків для валідації. Графік розподілу зображень відповідно до семи основних [41] емоцій наведено на рисунку 2.9.

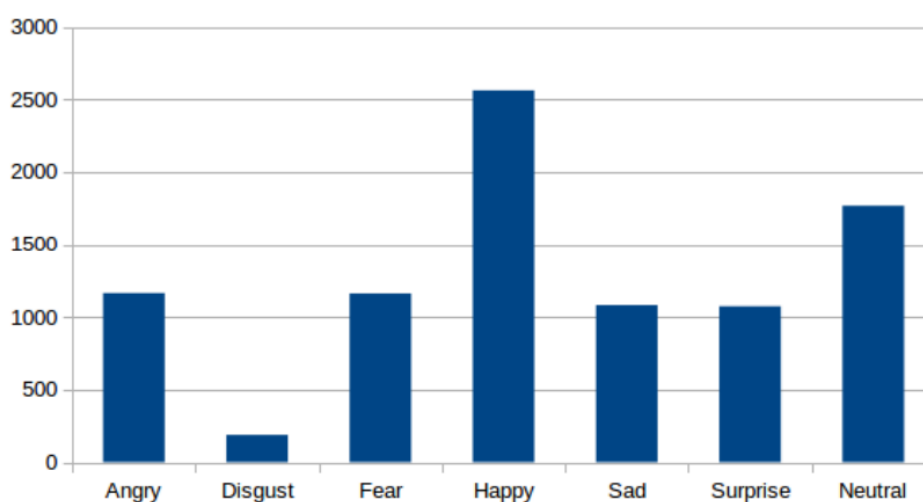


Рисунок 2.9 – Графік розподілу зображень відповідно до семи основних емоцій

2.5 Розробка алгоритму функціонування інформаційної системи для розпізнання емоцій людини

Дану інформаційну систему можна розділити на 4 модулі, які взаємодіють між собою. Структурна схема функціонування інформаційної системи для розпізнання емоцій людини зображена на рисунку 2.10.

Модуль навчання відповідає за налаштування вагів взаємозв'язків у нейронній мережі. Це надає можливість програмі запам'ятовувати особливості емоцій людини, і використовувати цю інформацію для розпізнання іншої емоції. Модуль містить в собі вбудований відформатований для навчання

тестовий набір зображень певного формату, за допомогою якого можна провести навчання нейронної мережі.



Рисунок 2.10 – Загальна структура інформаційної системи розпізнавання емоцій людини

Модуль графічного користувацького інтерфейсу PyQt відповідає за відображення графічного інтерфейсу користувача. Він виконує роль оболонки програми. Усі модулі мають можливість виклику із цього модуля. Також даний інструмент надає усі елементи інтерфейсу для взаємодії з користувачем. Іншими словами це набір інструментів для побудови графічного інтерфейсу.

Модуль розпізнавання відповідає за основну можливість програмного модуля: безпосередньо за розпізнавання людської емоції. Нейронна мережа, на основі знань (особливостей), яким вона навчена, здійснює класифікацію образу емоції та відносить його до одного з уже відомих їй.

Для дослідження отриманих результатів, було розроблено модуль візуалізації історії тестування нейронної мережі у процесі навчання. У процесі навчання і автоматичного тестування після кожної епохи зберігається результат тестування. Таким чином, аби відобразити залежність кількості правильних результатів розпізнавання емоцій (точності, Accuracy) відносно до кількості епох (Epoch), які зменшують швидкодію побудовано графік Model Accuracy. Тестування здійснювалося на двох наборах даних: відформатований

для навчання тестовий набір із приблизно 29000 зображень певного формату (train) і набір із більше 7000 фотографій довільного формату (val).

При запуску програми користувач матиме змогу вибрати, з яким із модулів йому працювати. Модуль навчання надасть можливість автоматично навчити нейронну мережу із вбудованого набору даних.

Модуль розпізнання запустить веб-камеру на комп'ютері користувача, після чого відеопотік буде виведено у вигляді окремого вікна. Якщо програма знайде у відопотоці образ обличчя людини, його буде виділено прямокутником, а поруч з ним з'явиться емоція, яку програма розпізнала. При роботі з модулем валідації, буде виконано процес валідації і тестування, в результаті якого користувач зможе впевнитись у правильності вибору архітектури нейронної мережі, та порівняти отримані результати з іншими дослідженнями.

Таким чином досягається усі задання поставлені на початку дослідження.

2.6 Розробка структури інформаційної технології розпізнавання емоцій людини.

На вхід неперервний відеопотік з веб-камери комп'ютера зі швидкістю зміни кадрів не менше 24 кадрів на секунду. Роздільна здатність вхідного кадру не повинна бути меншою за 40×40 пікселів і більшою за 1920×1080 пікселів. Після цього відбувається зміна масштабу зображення до заданого у налаштуваннях нейронної мережі. Значення кожного пікселя перетворюється так, щоб їх діапазон був від 0 до 1. Після цього зображення потрапляє на вхід нейронної мережі і відбувається процес локалізації образу обличчя методом каскаду Хаара. На отримані області відбувається пошук ознак емоції за допомогою фільтра Габора. На основі отриманих ознак виконується розпізнавання – класифікація емоцій людини на зображеному кадрі. Наступним є процес обробки вихідних даних та виведення отриманого результату.

Структура інформаційної технології колоризації чорно-білих зображень з використанням нейронних мереж зображено на рисунку 2.11.

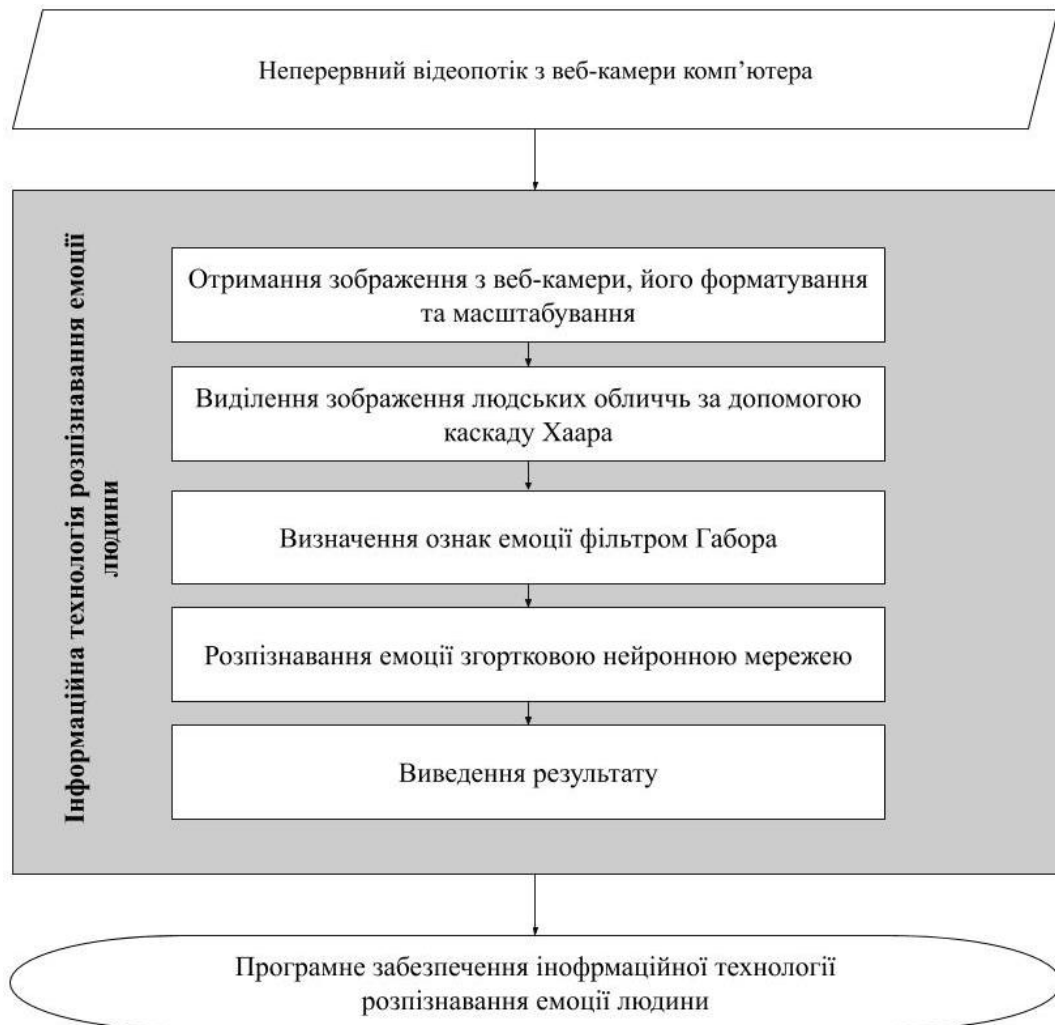


Рисунок 2.11 – Структура інформаційної технології розпізнавання емоції людини

2.7 Висновок

У результаті дослідження було виділено 3 основних етапи роботи методу розпізнавання емоцій людини:

- попередня обробка зображення;
- отримання візуальних ознак;
- класифікація емоцій.

Після проведеного аналітичного огляду існуючих реалізацій цих етапів було доведено: для попередньої обробки зображення слід використовувати метод каскаду Хаара; для отримання візуальних ознак доцільно використовувати фільтр Габора, для класифікації емоцій варто застосовувати згорткову нейронну мережу.

Запропоновано математичну модель згорткової нейронної мережі та шарів, які входять до складу її архітектури, на основі якої буде побудована така архітектура нейронної мережі, яка дозволить з достатньою точністю розпізнавати емоції людини, не потребуючи додаткових навантажень на апаратне забезпечення комп'ютера.

В результаті аналізу існуючих архітектур згорткових нейронних мереж, було обрано згорткову нейронну мережу, запропоновану у роботі Гуді, як найбільш перспективну для подальшого дослідження.

Було удосконалено архітектуру вище обраної нейронної мережі, та доведено доцільність такого удосконалення.

Обґрунтовано вибір FERС-2013, як набору даних для навчання вдосконаленої згорткової нейронної, який відрізняється великим обсягом набору даних і різноманітністю його елементів.

Розроблено алгоритм роботи інформаційної системи, який базується на даній вдосконаленій згортковій нейронній мережі.

Запропоновано структуру інформаційної технології для розпізнавання емоцій людини, побудованої на основі вдосконаленої архітектури згорткової нейронної мережі.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ

3.1 Розробка алгоритму роботи програмних засобів розпізнавання емоцій людини

Загальний алгоритм роботи інформаційної системи розпізнавання емоцій наведено на рис. 3.1.

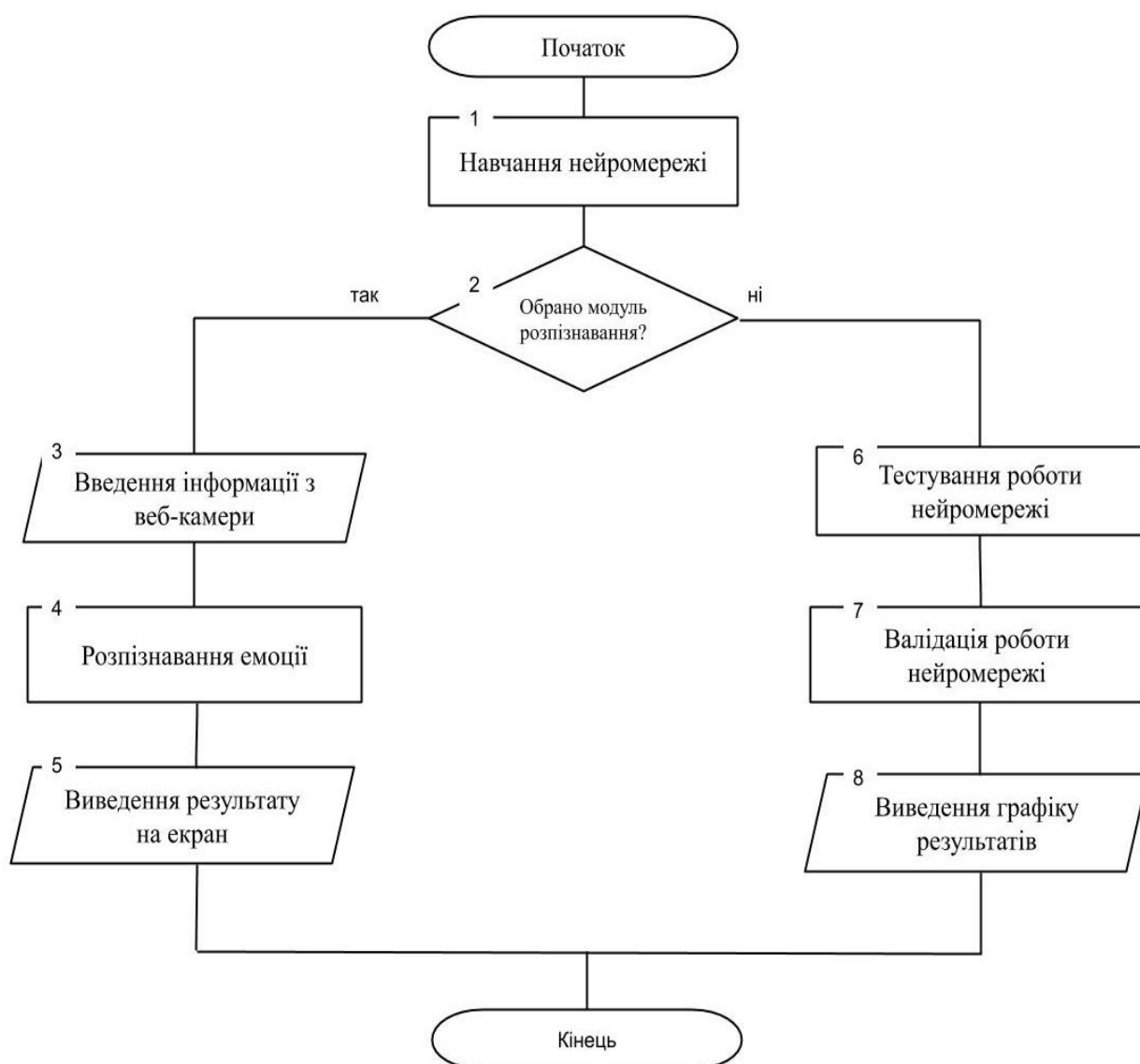


Рисунок 3.1 – Загальний алгоритм роботи інформаційної системи розпізнавання емоцій

Інформаційна технологія складається з трьох основних модулів – модуля навчання нейронної мережі, модуля розпізнавання емоцій людини та модуля тестування та валідації.

Модуль навчання нейронної мережі виконує функцію тренування нейронної мережі на заданому наборі даних – фотографій із зображеннями обличчя людини. Так як для навчання використано метод навчання з учителем, набір даних містить дані про правильний результат розпізнавання. Нейронна мережа, яка на початку тренування заповнена випадковими значеннями ваг, корелює значення ваг в залежності від значення, які необхідні для досягнення правильного результату. Алгоритм роботи модуля навчання нейронної мережі зображено на рисунку 3.2.

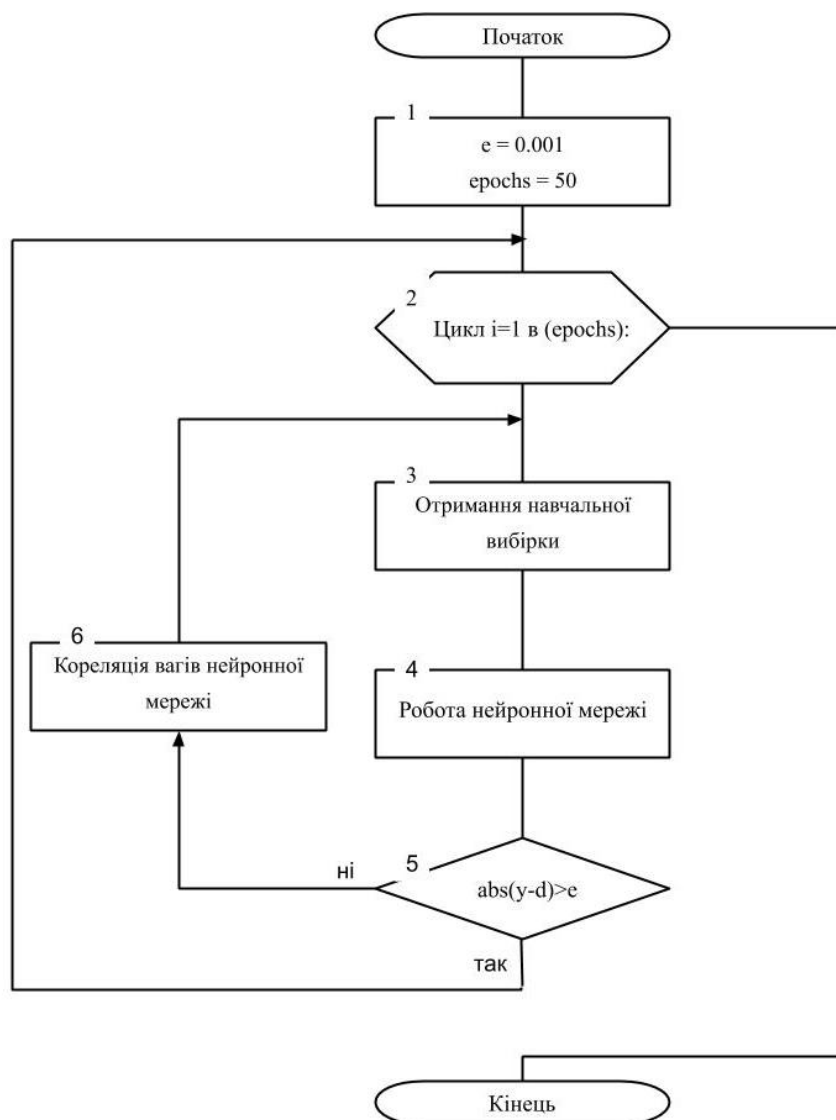


Рисунок 3.2 – Алгоритм роботи модуля навчання нейронної мережі

Другим модулем, там модулем, який виконує основну функцію інформаційної системи є модуль розпізнавання емоцій людини. В модулі розпізнавання система підключає веб камеру, у отриманому на вхід зображенні розпізнає обличчя людини та визначає їх кількість методом каскаду Хаара. Після цього виділяються ознаки емоції, на основі яких класифікується емоція. Зображення з веб камери виводиться на екран із окресленим контуром обличчя та назвою розпізнаної на ньому емоції. Розпізнавання виконується у неперервному відеопотоці. Результат роботи даного модуля підтверджує практичну доцільність розробки інформаційної технології для розпізнавання емоцій людини. Схема роботи модуля розпізнавання емоцій наведена на рисунку 3.3.



Рисунок 3.3 – Схема роботи модуля розпізнавання нейронної мережі

Модуль тестування і валідації використовується для підтвердження наукової доцільності розробки інформаційної системи для розпізнавання

емоцій людини. Для тестування використовується набір даних для навчання із відформатованими зображеннями. Програма розпізнає емоції, після чого результат порівнюється із набором даних, який містить дані про правильний результат розпізнавання. Результат порівняння зберігається як правильний, якщо емоцію розпізнано правильно і помилковий, якщо емоція розпізнана невірно.

Для валідації використовується набір даних із невідформатованими зображеннями. Це можуть бути зображення із довільним розширенням, із присутніми на них шумами та іншими завадами. Програма розпізнає емоції на невідформатованому зображенні, після чого результат порівнюється із набором даних, який містить дані про правильний результат розпізнавання. Результат порівняння зберігається як правильний, якщо емоцію розпізнано правильно і помилковий, якщо емоція розпізнана помилково.

Після проведення тестування і валідації на основі збережених даних будується графік із порівнянням результатів даних операцій. Графік виводиться на екран і зберігається у відповідній папці із програмою.

3.2 Побудова UML-діаграми класів розробленого програмного засобу

Для правильного та швидкого проектування інформаційної системи було використано об'єктно-орієнтовану парадигму програмування. Розроблена архітектура класів в повній мірі відображає загальну архітектуру програми.

UML - діаграма класів програми зображена на рисунку 3.4.

Наведемо опис основних класів, використаних в програмі.

Conv2D – клас що використовується для роботи двовимірного згорткового шару. Цей шар описує ядро згортки, яке спільно з входом шару створює тензор вихідних даних. Використовуючи цей шар як перший шар у моделі, вказується аргумент ключового слова `input_shape` (кортеж цілих чисел,

не включає вісь зразка), наприклад `input_shape = (128, 128, 3)` для 128×128 RGB-зображень.

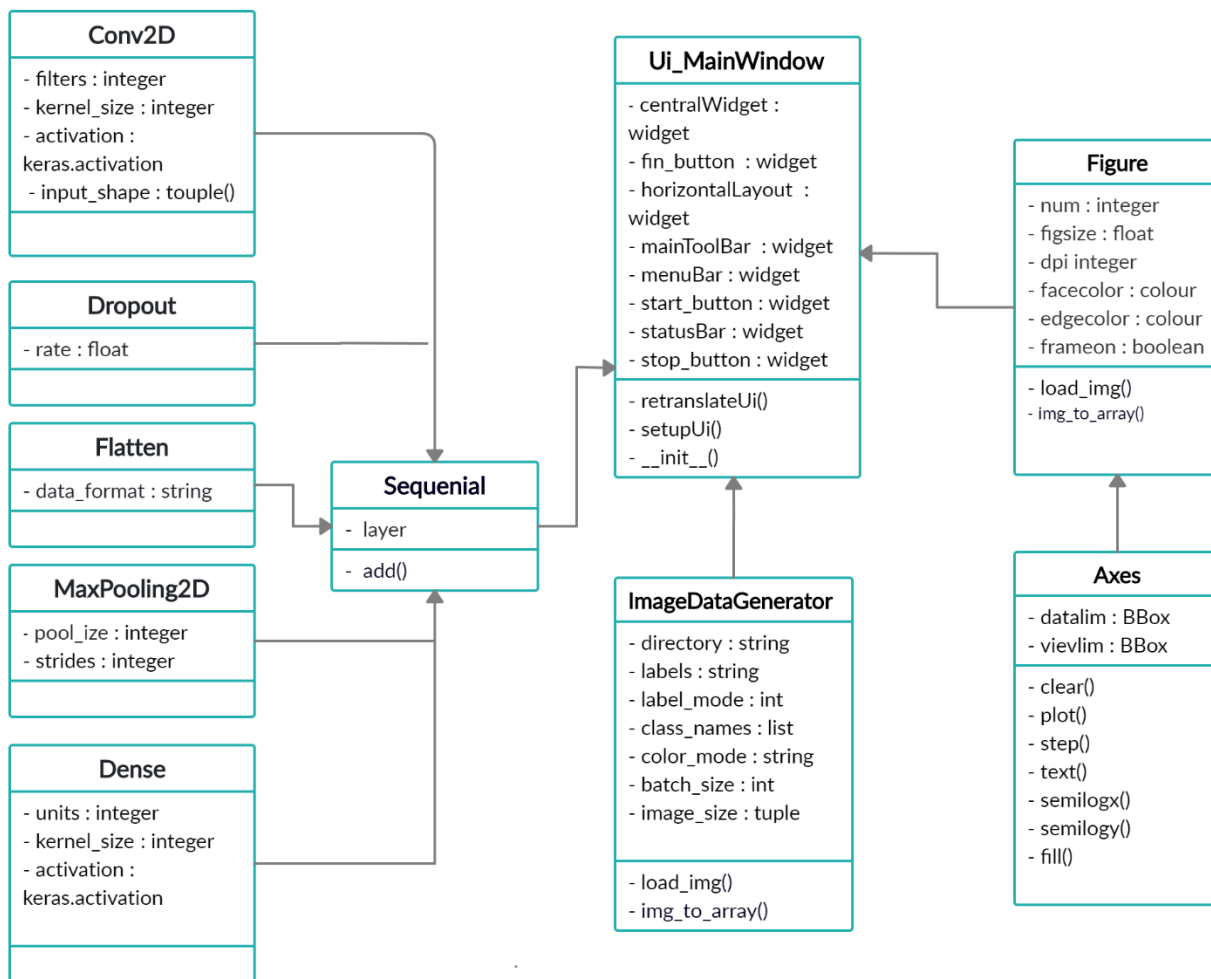


Рисунок 3.4 – UML діаграма класів програми

DropOut – клас, що застосовується для описання шару випадання. Шар випадіння вирішує проблему перенавчання нейронної мережі. Рівень випадіння випадково встановлює вхідні значення вагів на 0 з частотою `rate` на кожному кроці під час навчання, що допомагає запобігти перенавчанню нейронної мережі. Входи, не встановлені на 0, масштабуються на $1 / (1 - \text{rate})$ таким чином, що сума по всіх входах залишається незмінною.

Dense – клас що описує повнозв'язний шар нейронної мережі. Dense реалізує операцію: `activation(dot(input, kernel))`, де `activation` – це функція активації, передана як аргумент активації, `kernel` – це матриця ваг, створена шаром, а `dot` – функція добутку вхідної `input` матриці на `kernel`.

Flatten – клас, що описує шар для перетворення шару у плоский шар. Плоский шар – шар, де видалені всі виміри окрім одного.

MaxPooling2 – клас, що описує шар дискретизації на основі вибірки. Мета полягає в тому, щоб зменшити вибірку вхідного представлення (зображення, вихідну матрицю прихованого шару і т. д.). Зменшивши його розмірність і зробивши припущення щодо властивостей про властивості, що містяться в обраних субрегіонах зображення. Працює на основі вибору максимального значення з підвибірки розміром `pool_size`. За замовчуванням підвибірка зсувається без перекриття.

Sequential – клас, що описує модель використовуваної згорткової нейронної мережі. Модель нейронної мережі описує архітектуру та конфігурацію, а також використовує алгоритми навчання. Метод `add()` додає до моделі один із шарів, описаних класом, що наведені вище.

Axes – клас осей для побудови графіку. Містить опис більшості елементів фігури: вісь, крапку, лінію, текст, багатокутник тощо та встановлює систему координат. А екземпляри Axes підтримують зворотні виклики через атрибут зворотних викликів. Метод `plot()` у модулі осей бібліотеки `matplotlib` використовується для побудови графіків у по x як рядків та/або маркерів. Метод `clean()` очищує графік, `step()` продовжує лінію у точку з визначеними координатами, `semilogx` і `semilogy()` задають властивості осей координат відповідно.

Figure – клас, що описує це найважливіший зовнішній контейнер для графіки `matplotlib`, який може включати в себе кілька об'єктів Axes. Можна розглядати об'єкт Figure як такий, що схожий на ящик контейнера, що містить один або кілька об'єктів Axes (справжні графіки). Під об'єктами Axes, в порядку ієрархії розташовані менші об'єкти, такі як індивідуальні лінії, позначки, легенди і текстові бокси. Практично кожен «елемент» діаграми – це власний маніпульований об'єкт Python, аж до ярликів і позначок.

ImageDataGenerator – клас, може бути використаний, наприклад, для розширення навчального або тестового наборів даних, або для створення

набору даних на основі наявних зображень. У випадку інформаційної системи розпізнавання емоцій – для генерування зображень ознак емоції, для подальшого їх перетворення у ядро згортки. Властивість `Directory`: каталог, де знаходяться дані зображення `label` – назва зображення. `Color_mod` – один із «відтінків сірого», «`rgb`», «`rgba`». За замовчуванням: «`rgb`». `batch_size`: Розмір пакетів даних. За замовчуванням: 32. `image_size` – параметр для зменшення розміру зображень після їх зчитування з диска. За замовчуванням 256×256 . Оскільки конвеєр обробляє партії зображень, які повинні мати однаковий розмір, це повинно бути передбачено. Метод `load()` завантажує зображення у об'єкт, метод `img_to_array()` перетворює його в масив даних для подальшого використання у програмі, зокрема в якості вхідного шару нейронної мережі.

`GUIDesign` – клас, що відповідає за відображення головного меню програми і взаємодію з користувачем. Містить інформацію про графічний інтерфейс, зокрема про його віджети: кнопки, панелі.

Розроблені схеми роботи модулів та UML-діаграма класів забезпечують необхідні знання та алгоритми для розробки програмного модуля.

3.3 Обґрунтування вибору мови програмування

При виборі засобів для розробки програмного проекту необхідно врахувати надзвичайно велику кількість різноманітних аспектів, найбільш важливим із яких є мова програмування, тому що вона в значній мірі визначає інші доступні засоби. При виборі мови програмування основними критеріями були: продуктивність програмування, продуктивність роботи додатку та ефективність використання пам'яті також можливість роботи з процесами та потоками

Важливим чинником для вибору мови, в даному випадку є її універсальність, оскільки використання багатьох мов в одному проекті може негативно вплинути на його якість та ускладнити підтримку. Для розробки

програмних систем такого рівня існує три мови, які відповідають вимозі універсальності: Java, C# та Python [42].

Java – об'єктно-орієнтована мова програмування, випущена компанією Sun Microsystems у 1995 році як основний компонент платформи Java. Зараз мовою займається компанія Oracle, яка придбала Sun Microsystems у 2009 році. Синтаксис мови багато в чому походить від C та C++. У офіційній реалізації, Java програми компілюються у байткод, який при виконанні інтерпретується віртуальною машиною для конкретної платформи. Oracle надає компілятор Java та віртуальну машину Java, які задовольняють специфікації Java Community Process, під ліцензією GNU General Public License.

C# (вимовляється Сі-шарп) – об'єктно-орієнтована мова програмування з безпечною системою типізації для платформи .NET. Розроблена Андерсом Гейлсбергом, Скотом Вілтанутом та Пітером Гольде під егідою Microsoft Research (при фірмі Microsoft). Синтаксис C# близький до C++ і Java. Мова має строгу статичну типізацію, підтримує поліморфізм, перевантаження операторів, вказівники на функціїчлени класів, атрибути, події, властивості, винятки, коментарі у форматі XML. Перейнявши багато що від своїх попередників – мов C++, Delphi, Модула і Smalltalk – C#, спираючись на практику їхнього використання, виключає деякі моделі, що зарекомендували себе як проблематичні при розробці програмних систем, наприклад множинне спадкування класів (на відміну від C++).

Python (рекомендоване прочитання – «Пайтон», запозичено назву з британського шоу Монті Пайтон) – інтерпретована об'єктно-орієнтована мова програмування високого рівня з динамічною семантикою. Розроблена в 1990 році Гвідо ван Россумом. Структури даних високого рівня разом із динамічною семантикою та динамічним зв'язуванням роблять її привабливою для швидкої розробки програм, а також як засіб поєднання існуючих компонентів [42]. Python підтримує модулі та пакети модулів, що сприяє модульності та повторному використанню коду. Інтерпретатор Python та стандартні бібліотеки доступні як у скомпільованій так і у вихідній формі на

всіх основних платформах. В мові програмування Python підтримується кілька парадигм програмування, зокрема: об'єктно-орієнтована, процедурна, функціональна та аспектноорієнтована.

І C#, і Java та Python підтримують можливість роботи з процесами та потоками. В силу підтримки операційною системою Windows, найбільшою швидкістю при роботі з потоками і процесами володіє мова програмування C#. Мова програмування Python компенсує нижчу швидкість простим, привабливим для роботи (в силу простоти написання коду) модулем для роботи з процесами – «multiprocessing», а також величезною довідкою для застосування можливостей, описаних в модулі [43].

Python – широко використовується програмістами через його чистої граматики і синтаксису, приємного дизайну. Різні структури даних, велика кількість фреймворків для тестування, співвідношення високого і низького рівня програмування роблять Python одним з найпопулярніших мов програмування для штучного інтелекту. А такі програмні пакети як pandas, scikit-learn і Tensorflow, роблять Python надійним варіантом для сучасних додатків в області машинного навчання [44].

Основним недоліком мови C# перед Java та Python є її пропрієтарна ліцензія та можливість використання усіх її можливостей в повному обсязі лише на ОС Windows. Тому реальним є вибір між Java та Python. Швидкість розробки програмного забезпечення мовою Python є на порядок вищою у порівнянні із Java завдяки його динамічній типізації. Крім того Python є інтерпретованою мовою, що полегшує внесення змін та налаштування програмної системи на етапі впровадження. Тому основною мовою програмування обрано мову Python

3.4 Обґрунтування вибору нейромережевої бібліотеки

Як зазначалося раніше, Python є найбагатшим у програмуванні бібліотек нейронної мережі.

Keras — відкрита нейромережна бібліотека, написана мовою Python [45]. Вона здатна працювати поверх TensorFlow. Спроектвану для уможливлення швидких експериментів з мережами глибинного навчання, її зосереджено на тому, щоби вона була зручною в користуванні, модульною та розширюваною. Її було створено як частину дослідницьких зусиль проєкту ONEIROS (англ. Open-ended Neuro-Electronic Intelligent Robot Operating System), а її основним автором та підтримувачем є Франсуа Шолле (фр. François Chollet), інженер Google.

2017 року команда TensorFlow Google вирішила підтримувати Keras в основній бібліотеці TensorFlow. Шолле пояснив, що Keras було замислено радше як інтерфейс, аніж як самостійну систему машинного навчання. Вона пропонує високорівневий, інтуїтивніший набір абстракцій, який робить розробку глибинно-нейромережних моделей простою незалежно від використовуваного обчислювального тилу.

Keras містить численні втілення широко вживаних нейромережних будівельних блоків, таких як шари, цільові та передавальні функції, оптимізувальники та безліч інструментів для спрощення роботи із зображеннями та текстом, щоби спрощувати кодування, потрібне для написання глибинно-нейромережного коду. Її код розміщено на GitHub, а до форумів спільнотної підтримки належать сторінка питань GitHub та канал Slack.

На додачу до стандартних нейронних мереж, Keras містить підтримку згорткових та рекурентних нейронних мереж, що є дуже актуальним в умовах розробки інформаційної системи розпізнання емоцій людини. Вона підтримує інші поширені службові шари, такі як виключення, пакетне унормовування та агрегування.

Keras дає своїм користувачам можливість виробляти продукти на основі глибинних моделей для смартфонів (iOS та Android), веб-сайтів та віртуальної машини Java. Вона також дозволяє використовувати розподілене тренування

моделей глибинного навчання на кластерах графічних (ГП) та тензорних (ТП) процесорів переважно у зв'язці з CUDA.

Саме завдяки описаним вище перевагам для розробки інформаційної системи розпізнавання емоцій людини було вибрано бібліотеку Keras.

3.5 Обґрунтування вибору середовища розробки

Інтегроване середовище розробки, ІСР (англ. IDE, Integrated development environment) – система програмних засобів, використовувана програмістами для розробки програмного забезпечення (ПО) [46]. Зазвичай, середовище розробки містить:

- текстовий редактор;
- компілятор і / або інтерпретатор;
- засоби автоматизації збирання;
- відладчик.

Іноді містить також засоби для інтеграції з системами управління версіями і різноманітні інструменти для спрощення конструювання графічного інтерфейсу користувача.

Багато сучасних середовищ розробки також включають браузер класів, інспектор об'єктів і діаграму ієрархії класів – для використання при об'єктно-орієнтованій розробці ПЗ. Хоча й існують ІСР, призначені для декількох мов програмування – такі як Eclipse, Embarcadero RAD Studio, Qt Creator, останні версії NetBeans, Xcode або Microsoft Visual Studio, але зазвичай ІСР призначається для одного певної мови програмування – як, наприклад, Visual Basic, Delphi, Dev-C++.

Інтегровані середовища розробки були створені для того, щоб максимізувати продуктивність програміста завдяки тісно пов'язаним компонентам з простими користувача інтерфейсами. Це дозволяє розробнику зробити менше дій для перемикання різних режимів, на відміну від дискретних програм розробки. ІСР, зазвичай, являє собою єдину програму, в якій

проводилася вся розробка. Вона, зазвичай, містить багато функцій для створення, зміни, компілювання, розгортання і налагодження програмного забезпечення.

Мета середовища розробки полягає в тому, щоб абстрагувати конфігурацію, необхідну, щоб об'єднати утиліти командного рядка в одному модулі, який дозволить зменшити час, щоб вивчити мову, і підвищити продуктивність розробника. Також вважається, що важка інтеграція завдань розробки може далі підвищити продуктивність. Наприклад, ICP дозволяє проаналізувати код і тим самим забезпечити миттєвий зворотний зв'язок і повідомити про синтаксичні помилки.

Сьогодні існує досить велика кількість ICP для Python: Boa Constructor, Eclipse, PyDev, Eric, IDLE, Komodo, PyCharm, PyScripter, SPE, Wing IDE, Visual Studio Code [43]. Проте їхні можливості поки що не можуть задовольнити усіх потреб розробників програмного забезпечення. Винятками є Visual Studio Code та PyCharm. PyCharm – інтегрована середовище розробки для мови програмування Python.

Надає засоби для аналізу коду, графічний відладчик, інструмент для запуску юніт-тестів і підтримує веб-розробку на Django та Flask. PyCharm розроблена чеською компанією JetBrains на основі IntelliJ IDEA [46].

Можливості:

- Статичний аналіз коду, підсвічування синтаксису і помилок.
- Навігація за проектом і вихідного коду: відображення файлової структури проекту, швидкий перехід між файлами, класами, методами і використанням методів.
- Рефакторинг: перейменування, витяг методу, введення змінної, введення константи, підйом і спуск методу і т. д.
- Інструменти для веб-розробки з використанням фреймворку Django та Flask.
- Вбудований відладчик для Python.
- Вбудовані інструменти для юніт-тестування.

- Підтримка систем контролю версій: загальний користувацький інтерфейс для Mercurial, Git, Subversion, Perforce і CVS з підтримкою списків змін та злиття.
- Велика кількість плагінів, що дозволяють значно розширити можливості ICP, зокрема плагіни інтеграції із MongoDB.

Основним середовищем для розробки паралельного алгоритму сортування злиттям було обрано Visual Studio Code

Visual Studio Code — засіб для створення, редагування та відладки сучасних застосунків і програм для хмарних систем [46]. Visual Studio Code розповсюджується безкоштовно і доступний у версіях для платформ Windows, Linux і OS X. Компанія Microsoft представила Visual Studio Code у квітні 2015 на конференції Build 2015.

Це середовище розробки стало першим кросплатформовим продуктом у лінійці Visual Studio. За основу для Visual Studio Code використовуються напрацювання вільного проекту Atom, що розвивається компанією GitHub. Зокрема, Visual Studio Code є надбудовою над Atom Shell, що використовують браузерний рушій Chromium і Node.js. Примітно, що про використання напрацювань вільного проекту Atom на сайті Visual Studio Code і в прес-релізі і в офіційному блозі не згадується.

Редактор містить вбудовані засоби налагодження, інструменти для роботи з Git і засоби рефакторингу, навігації по коду, автодоповнення типових конструкцій і контекстної підказки. Продукт позиціонується як легковаге рішення, що дозволяє обійтися без повного інтегрованого середовища розробки Visual Studio також дозволяє замінювати кодову сторінку при збереженні документа, символи перекладу рядка і мову програмування поточного документа.

З 2018 року з'явилися розширення Python для Visual Studio Code з відкритим вихідним кодом. Воно надає розробникам широкі можливості для редагування, налагодження і тестування коду.

3.6 Тестування розробленого програмного засобу розпізнання емоцій людини та аналіз результатів його роботи

Основною задачею даного тестування є перевірка того, чи було реалізовано можливість розпізнання емоцій людини. І хоча самій програмі процес тестування автоматизовано, та існує модуль, що відповідає за тестування і валідацію, протестуємо основні функції на конкретному прикладі.

Після запуску програми користувач побачить перед собою основний інтерфейс програми – головне меню. Головне меню програми зображено на рисунку 3.5

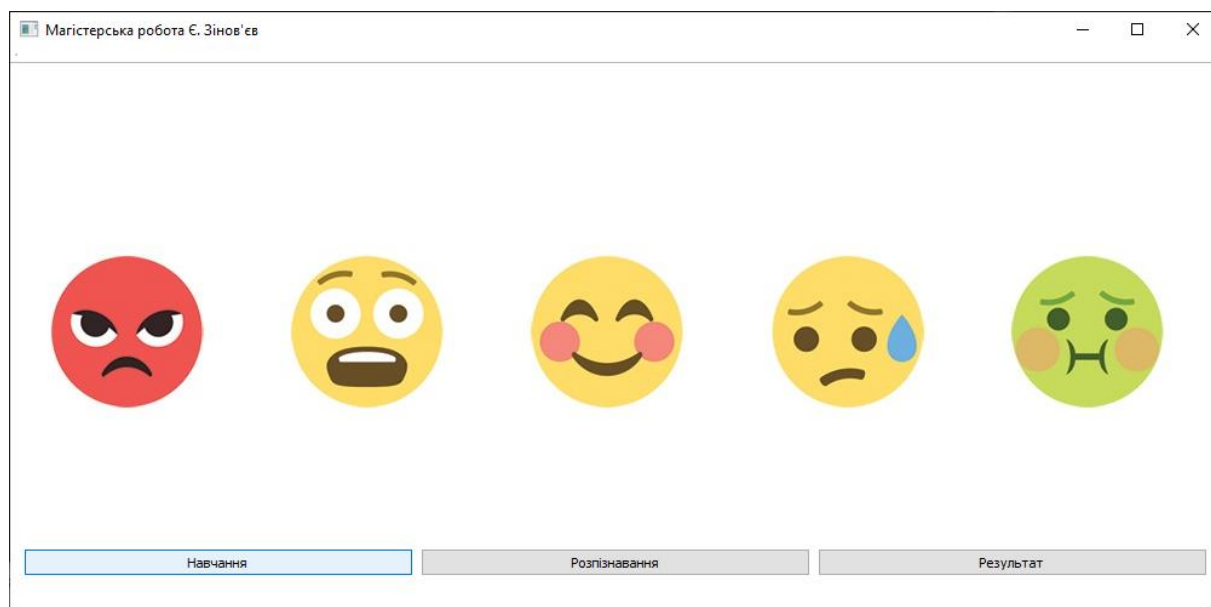


Рисунок 3.5 – Головне меню програми

У головному меню користувач може вибрати модуль, з яким йому необхідно працювати. Основна функція програми – саме демонстрація розпізнавання емоцій людини за допомогою згорткової нейронної мережі, отже інтерфейс виконано у мінімалістичному стилі.

Якщо програмний додаток запущено вперше – для коректного розпізнавання емоцій необхідно провести навчання нейронної мережі, натиснувши на кнопку «Навчання». У комплекті із програмою постачається 2

набори даних – для тренування і тестування та набір для валідації. Вони знаходяться у папці із програмою у папці «data», у вкладених папках «train» та «test» відповідно. Для тренування використані зображення, що розташовані у папці «train».

Якщо користувач одразу вирішить провести розпізнавання, натиснувши на відповідну кнопку, але при цьому нейронна мережа не буде навчена, процес навчання розпочнеться автоматично. Після початку тренування буде відкрите нове вікно. Форма навчання нейронної мережі наведено на рисунку 3.6.

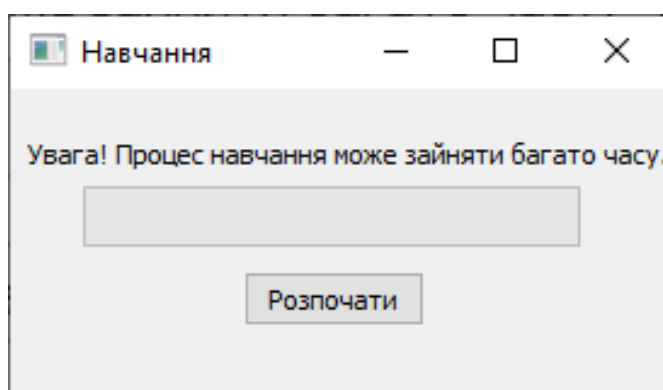


Рисунок 3.6 – Форма навчання нейронної мережі для розпізнавання емоцій людини

Швидкість процесу навчання залежить від ресурсів системи, де розгорнуто програму, але в середньому процес навчання може займати декілька годин. Тому форма навчання містить попередження про те, що процес навчання може зайняти тривалий час. Але процес навчання, як правило, проводять лише при першому запуску програми, а далі працюють із навченим набором даних. Форма, що показує прогрес процес навчання нейронної мережі зображений на рисунку 3.7.

Після закінчення навчання користувачу буде видано повідомлення, після закриття форми навчання програма повернеться до головного меню. Далі протестуємо основний модуль програми – розпізнавання.

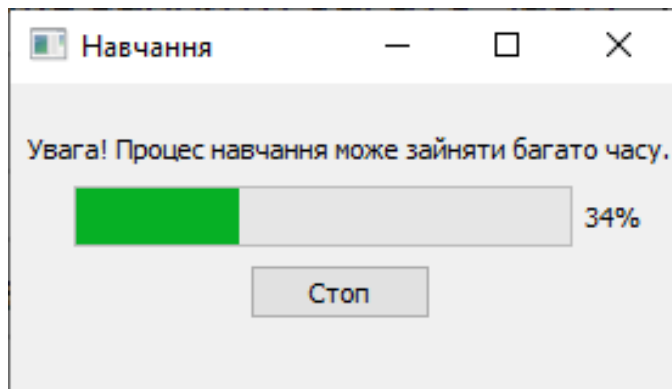


Рисунок 3.7 – Форма, що показує прогрес процес навчання нейронної мережі для розпізнавання емоцій людини.

Після натискання кнопки «Розпізнавання» в головному меню відкриться вікно та буде запущено веб-камеру. На екрані буде виведено неперервний відеопотік, в разі наявності на ньому обличчя, програма омалює його контуром, та розпізнає емоцію. Робота модуля розпізнавання зображена на рисунку 3.8.

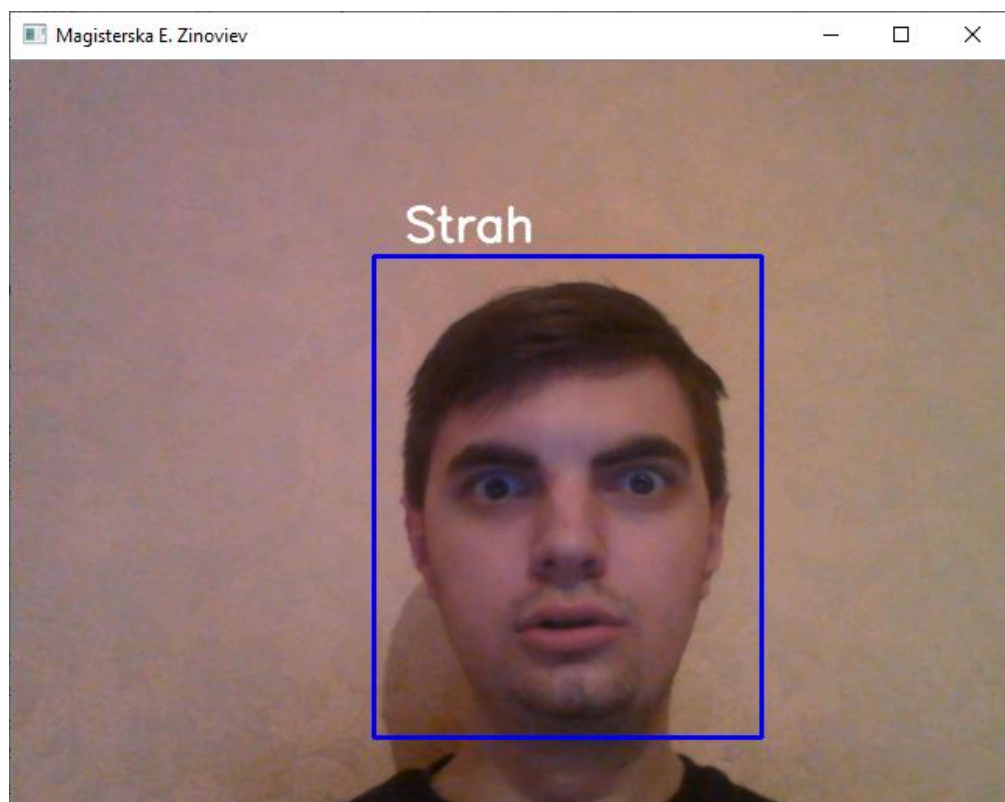


Рисунок 3.8 – Робота модуля розпізнавання емоцій людини

Роботу модуля розпізнавання емоцій людини також необхідно протестувати при зміщеному зображенні обличчя, це зображено на рисунку 3.9.

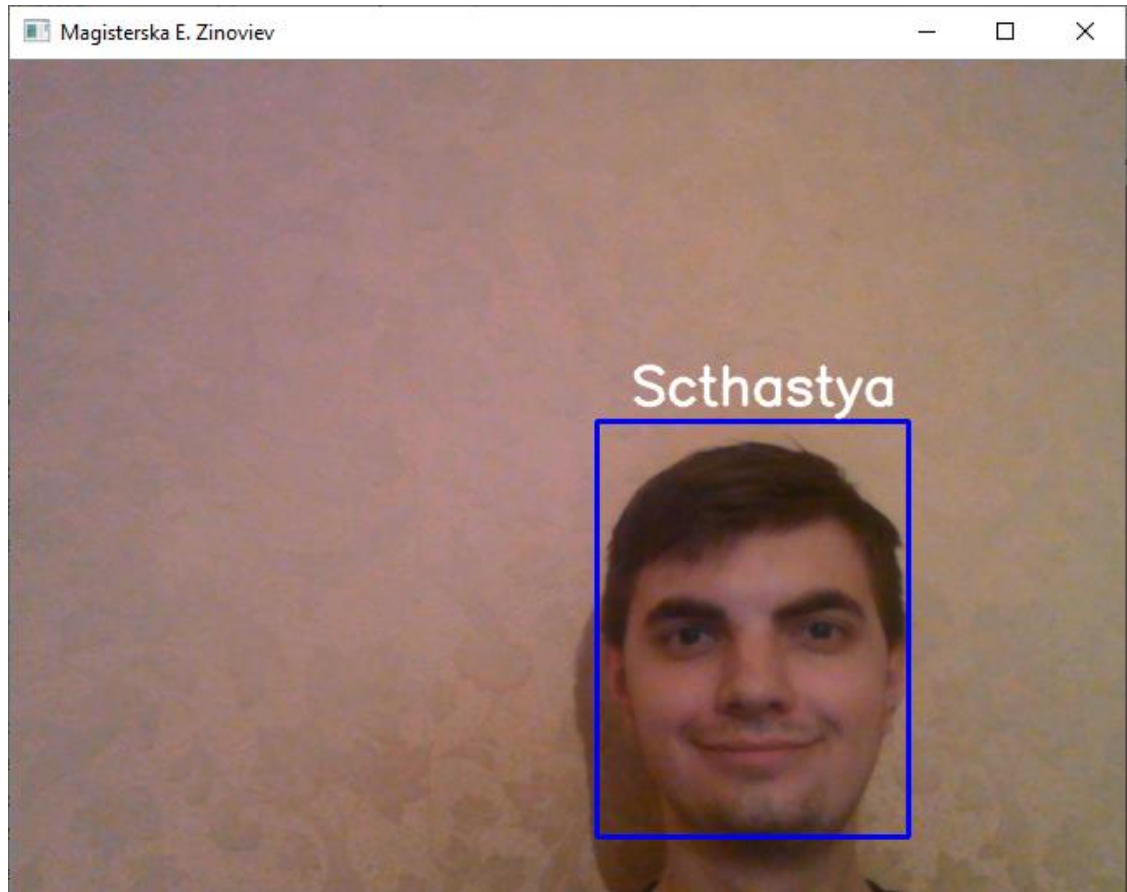


Рисунок 3.9 – Робота модуля розпізнавання емоцій людини при зміщеному положенні обличчя

Також необхідно перевірити можливість розпізнавання емоцій, коли в кадрі 2 або більше обличчя, та при наявності завад, наприклад шумів. Робота модуля розпізнавання при наявності завад та більше двох обличч зображена на рисунку 3.10.

Отже, тестування модулів тренування і розпізнавання емоції довели те, що програма задовольняє вимогам до неї, надаючи можливість точно і без великих затрат ресурсів системи розпізнавати емоцію людини у неперервному відеопотоці. Тим самим доведено практичну доцільність розробки.

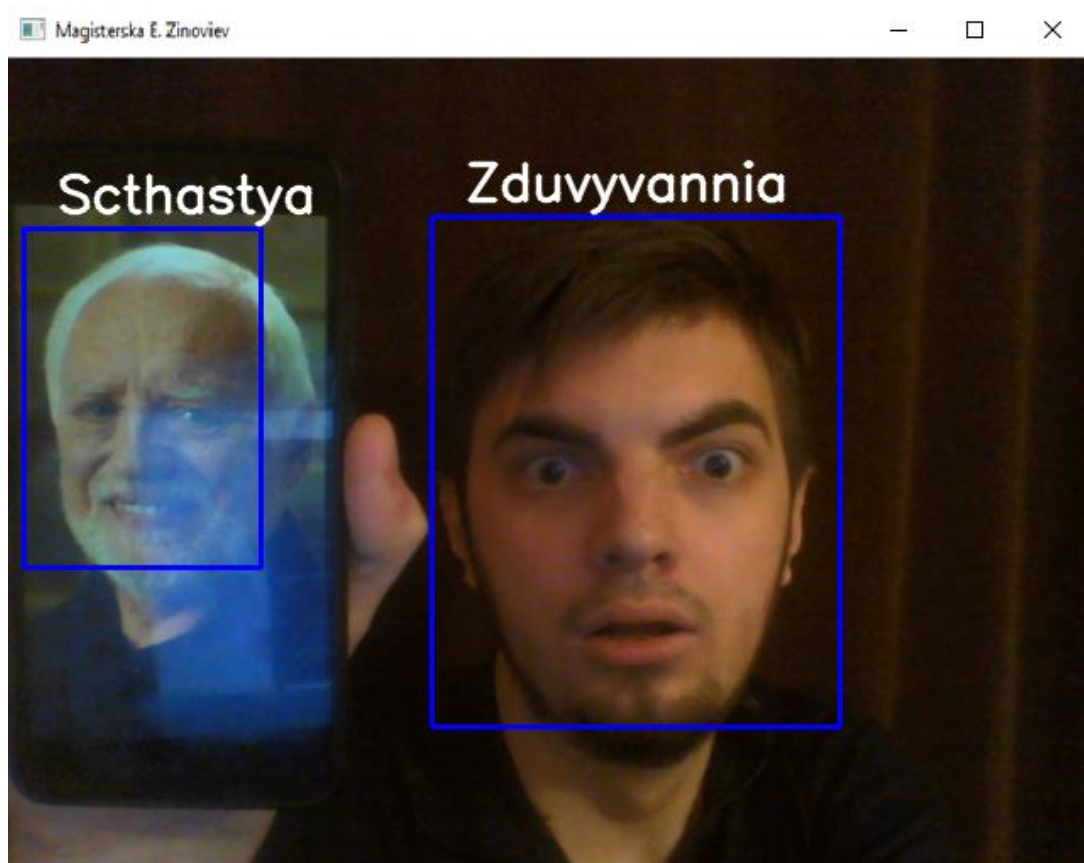


Рисунок 3.10 – Робота модуля розпізнавання при наявності завад та більше двох облич

Для проведення аналізу розробки у програмі передбачено модуль тестування і валідації. Щоб застосувати його, у головному меню програми необхідно натиснути кнопку «Результат». Програма почне проведення тестування та валідації, відкривши користувачу форму проведення тестування і валідації. Форма проведення тестування і валідації зображена на рисунку 3.11.

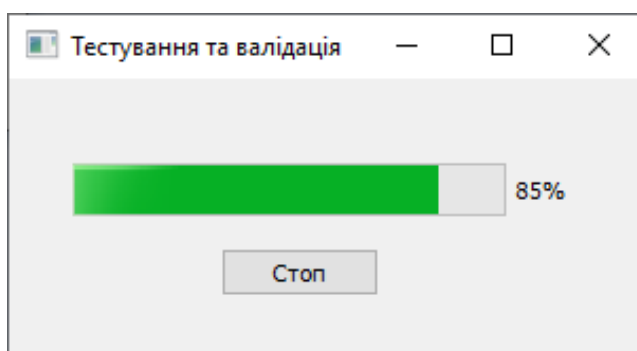


Рисунок 3.11. – Форма проведення тестування і валідації

Після закінчення тестування і валідації у формі буде відображено графік точності розпізнавання емоцій людини розробленим додатком. Графік точності розпізнавання емоцій людини розробленим додатком зображено на рисунку 3.12.

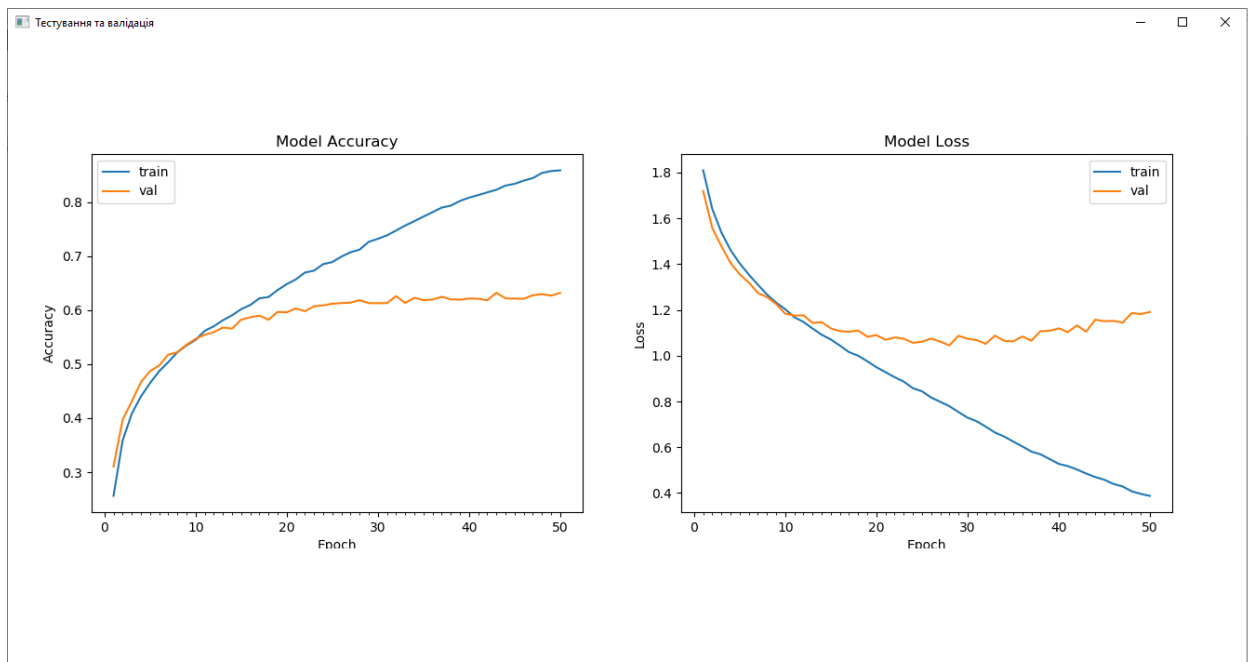


Рисунок 3.11. – Форма із графіками точності та функцією втрат розпізнавання емоцій людини

У лівій частині «Model Accuracy» наведена залежність точності від кількості епох навчання нейронної мережі. «train» – на відформатованому наборі даних, «val» – не відформатованому. Отже, під час тестування точність становила близько 90%, а під час валідації – 62% (а найближчий, за характеристиками аналог, наведений в пункті 1.2 – додаток FaceReader – 56%, тобто точність зросла в середньому на 6 – 7%). Отже, з урахуванням усіх можливостей програми, можна констатувати, що програма перевершує можливості своїх аналогів.

У правій частині «Model Loss» наведено залежність значення функції втрат від кількості епох навчання нейронної мережі. Функції втрат виражають

розбіжність між передбаченнями тренованої моделі та дійсними зразками задачі. «train» – на відформатованому наборі даних, «val» – не відформатованому. Тут необхідно відмітити, що на тестовому наборі даних функція втрат спадала зі збільшенням епох, але при валідації після проходження 40 епохи почала зростати. З одного боку це може свідчити про перенавчання нейронної мережі, з іншої – про збільшення кількості завад на використаних зображеннях набору.

Варто зазначити, що під час роботи програмного додатку втрат кадрів виявлено не було, що свідчить про плавність роботи розробленої програми. Навантаження на систему, в особливості на процесор – було стабільним, різких навантажень під час появи у кадрі обличчя людини не виявлено.

3.7 Висновок

Розроблено алгоритм роботи програмного засобу для розпізнавання емоцій людини, який складається з трьох модулів: навчання, розпізнавання та валідації.

Розроблено алгоритм навчання вдосконаленої архітектури згорткової нейронної мережі.

Розроблено схему роботи модуля розпізнавання емоцій людини, що базується на вдосконаленій архітектурі згорткової нейронної мережі.

Розроблено UML-діаграму класів програми для розпізнавання емоцій людини.

У результаті порівняльного аналізу об'єктно-орієнтованих мов на основі об'єктивних переваг, таких як наявність великої кількості нейромережевих бібліотек та простота написання коду, було обґрунтовано доцільність застосування мови програмування Python.

Обґрунтовано вибір нейромережевої бібліотеки Keras для подальшої програмної реалізації згорткової нейронної мережі для розпізнання емоцій людини. Наведено її можливості та основні переваги.

Виконано аналіз існуючих найпопулярніших середовищ програмування для обраної мови Python. Було обрано середовище програмування Visual Studio Code, яке задовольняє усім потребам при розробці програмного забезпечення, має зручний графічний інтерфейс а також засоби налагодження.

На основі здійснених досліджень розроблено програмне забезпечення, що дозволяє практично реалізувати технологію розпізнавання емоцій людини.

Здійснено тестування розробленого програмного додатку, яке підтвердило коректність його роботи.

Розроблено автоматизовану систему тестування додатку. Доведено залежність точності розпізнавання емоцій людини від формату вхідного зображення та кількості епох навчання нейронної мережі із вдосконаленою архітектурою. Тестування додатку продемонструвало підвищення точності розпізнавання емоції людини у відеопотоці до 62%, що перевершує показники розглянутих аналогів без суттєвого навантаження на апаратні ресурси комп'ютера (процесор).

4 ЕКОНОМІЧНА ЧАСТИНА

4.1 Оцінювання комерційного потенціалу розробки

Метою проведення технологічного аудиту є оцінювання комерційного потенціалу розробки. Для проведення технологічного аудиту було залучено двох незалежних експертів. Такими експертами будуть Арсенюк І.Р. та Сілагін О.В.

Оцінювання комерційного потенціалу розробки буде здійснено за 12-ма критеріями за 5-ти бальною шкалою. Результати оцінювання комерційного потенціалу розробки наведено в таблиці 4.1.

Таблиця 4.1 – Результати оцінювання комерційного потенціалу розробки

Критерії	Прізвище, ініціали, посада експерта	
	Арсенюк І.Р.	Сілагін О.В.
	Бали, виставлені експертами:	
1	4	4
2	3	3
3	3	4
4	4	3
5	3	4
6	4	
7	3	3
8	4	4
9	4	3
10	4	3
11	3	4
12	3	4
Сума балів	СБ ₁ = 43	СБ ₂ = 43
Середньоарифметична сума балів $\overline{СБ}$	$\overline{СБ} = \frac{\sum_{i=1}^3 СБ_i}{2} = 43$	

Отже, з отриманих даних таблиці 4.1 видно, що нова розробка має високий рівень комерційного потенціалу.

4.2 Прогнозування витрат на виконання науково-дослідної роботи та конструкторсько–технологічної роботи.

Для розробки нового програмного продукту необхідні такі витрати.

Основна заробітна плата для розробників визначається за формулою (4.1):

$$Z_o = \frac{M}{T_p} \cdot t, \quad (4.1)$$

де M – місячний посадовий оклад конкретного розробника;

T_p – кількість робочих днів у місяці, $T_p = 22$ дні;

t – число днів роботи розробника, $t = 50$ днів.

Розрахунки заробітних плат для керівника і програміста наведені в таблиці 4.2.

Таблиця 4.2 – Розрахунки основної заробітної плати

Працівник	Оклад M , грн.	Оплата за робочий день, грн.	Число днів роботи, t	Витрати на оплату праці, грн.
Науковий керівник	10000	454,55	5	2272,75
Інженер-програміст	6000	272,73	50	13636,36
Всього:				15909,11

Розрахуємо додаткову заробітну плату:

$$Z_{\text{дод}} = 0,1 \cdot 15909,11 = 1590,91 \text{ (грн.)}$$

Нарахування на заробітну плату операторів НЗП розраховується як 22% від суми їхньої основної та додаткової заробітної плати:

$$H_{\text{зп}} = (Z_o + Z_p) \cdot \frac{\beta}{100}, \quad (4.2)$$

$$H_{\text{зп}} = (15909,11 + 1590,91) \cdot \frac{22}{100} = 3850,00 \text{ (грн.)}$$

Розрахунок амортизаційних витрат для програмного забезпечення виконується за такою формулою:

$$A = \frac{Ц \cdot H_a}{100} \cdot \frac{T}{12}, \quad (4.3)$$

де Ц – балансова вартість обладнання, грн;

H_a – річна норма амортизаційних відрахувань % (для програмного забезпечення 25%);

T – Термін використання (T=3 міс.).

Таблиця 4.3 – Розрахунок амортизаційних відрахувань

Найменування програмного забезпечення	Балансова вартість, грн.	Норма амортизації, %	Термін використання, міс.	Величина амортизаційних відрахувань, грн
Персональний комп'ютер	10000	25	3	625
Всього:				625

Розрахуємо витрати на комплектуючі. Витрати на комплектуючі розрахуємо за формулою:

$$K = \sum_1^n H_i \cdot C_i \cdot K_i, \quad (4.4)$$

де n – кількість комплектуючих;

H_i - кількість комплектуючих i -го виду;

C_i – покупна ціна комплектуючих i -го виду, грн;

K_i – коефіцієнт транспортних витрат (прийmemo $K_i = 1,1$).

Таблиця 4.4 - Витрати на комплектуючі, що були використані для розробки ПЗ.

Найменування матеріалу	Одиниці виміру	Ціна, грн.	Витрачено	Вартість витрачених матеріалів, грн.
Флешка	шт.	180	1	180
Пачка паперу	уп.	130	1	130
Ручка	шт.	10	1	10
Всього з урахуванням транспортних витрат				352

Витрати на силову електроенергію розраховуються за формулою:

$$V_e = V \cdot P \cdot \Phi \cdot K_n, \quad (4.5)$$

де V – вартість 1кВт-години електроенергії ($V=1,68$ грн/кВт);

P – установлена потужність комп'ютера ($P=0,6$ кВт);

Φ – фактична кількість годин роботи комп'ютера ($\Phi=200$ год.);

K_n – коефіцієнт використання потужності ($K_n < 1$, $K_n = 0,7$).

$$V_e = 1,68 \cdot 0,6 \cdot 200 \cdot 0,7 = 141,12 \text{ (грн.)}$$

Розрахуємо інші витрати $V_{ін}$.

Інші витрати I_B можна прийняти як $(100...300)\%$ від суми основної заробітної плати розробників та робітників, які були виконували дану роботу, тобто:

$$V_{ін} = (1..3) \cdot (Z_o + Z_p). \quad (4.6)$$

Отже, розрахуємо інші витрати:

$$V_{ін} = 1 * (15909,11+1590,91) = 17500,02 \text{ (грн.)}$$

Сума всіх попередніх статей витрат дає витрати на виконання даної частини роботи:

$$B = Z_o + Z_d + H_{зп} + A + K + B_e + I_B$$

$$B = 15909,11+1590,91+3850,51+625+352+141,12+17500,02 = 39968,67 \text{ (грн.)}$$

Розрахуємо загальну вартість наукової роботи $B_{заг}$ за формулою:

$$B_{заг} = \frac{B_{ін}}{\alpha}, \quad (4.7)$$

де α – частка витрат, які безпосередньо здійснює виконавець даного етапу роботи, у відн. одиницях = 1.

$$B_{заг} = \frac{39968,67}{1} = 39968,67.$$

Прогнозування загальних витрат $ЗВ$ на виконання та впровадження результатів виконаної наукової роботи здійснюється за формулою:

$$ЗВ = \frac{В_{заг}}{\beta}, \quad (4.8)$$

де β – коефіцієнт, який характеризує етап (стадію) виконання даної роботи.

Отже, розрахуємо загальні витрати:

$$ЗВ = \frac{39968,67}{0,9} = 44409.63 \text{ (грн.)}$$

4.3 Прогнозування комерційних ефектів від реалізації результатів розробки інформаційної технології

Спрогнозуємо отримання прибутку від реалізації результатів нашої розробки. Зростання чистого прибутку можна оцінити у теперішній вартості грошей. Це забезпечить підприємству (організації) надходження додаткових коштів, які дозволять покращити фінансові результати діяльності .

Оцінка зростання чистого прибутку підприємства від впровадження результатів наукової розробки. У цьому випадку збільшення чистого прибутку підприємства $\Delta \Pi_i$ для кожного із років, протягом яких очікується отримання позитивних результатів від впровадження розробки, розраховується за формулою:

$$\Delta \Pi_i = \sum_1^n (\Delta \Pi_{я} \cdot N + \Pi_{я} \Delta N)_i \quad (4.9)$$

де $\Delta \Pi_{я}$ – покращення основного якісного показника від впровадження результатів розробки у даному році;

N – основний кількісний показник, який визначає діяльність підприємства у даному році до впровадження результатів наукової розробки;

ΔN – покращення основного кількісного показника діяльності підприємства від впровадження результатів розробки;

$P_{\text{я}}$ – основний якісний показник, який визначає діяльність підприємства у даному році після впровадження результатів наукової розробки;

n – кількість років, протягом яких очікується отримання позитивних результатів від впровадження розробки.

В результаті впровадження результатів наукової розробки витрати на виготовлення інформаційної технології зменшаться на 25 грн (що автоматично спричинить збільшення чистого прибутку підприємства на 25 грн), а кількість користувачів, які будуть користуватись збільшиться: протягом першого року – на 150 користувачів, протягом другого року – на 125 користувачів, протягом третього року – 100 користувачів. Реалізація інформаційної технології до впровадження результатів наукової розробки складала 500 користувачів, а прибуток, що отримував розробник до впровадження результатів наукової розробки – 400 грн.

Спрогнозуємо збільшення чистого прибутку від впровадження результатів наукової розробки у кожному році відносно базового.

Отже, збільшення чистого продукту ΔP_1 протягом першого року складатиме:

$$\Delta P_1 = 25 \cdot 500 + (400 + 25) \cdot 150 = 76250 \text{ грн.}$$

Протягом другого року:

$$\Delta P_2 = 25 \cdot 500 + (400 + 25) \cdot (150 + 125) = 129375 \text{ грн.}$$

Протягом третього року:

$$\Delta P_3 = 25 \cdot 500 + (400 + 25) \cdot (150 + 125 + 100) = 171875 \text{ грн.}$$

4.4 Розрахунок ефективності вкладених інвестицій та визначення періоду їх окупності

Визначимо абсолютну і відносну ефективність вкладених інвестором інвестицій та розрахуємо термін окупності.

Абсолютна ефективність E_{abc} вкладених інвестицій розраховується так:

$$E_{abc} = (ПП - PV), \quad (4.10)$$

де $\Delta П_i$ – збільшення чистого прибутку у кожному із років, протягом яких виявляються результати виконаної та впровадженої НДДКР, грн;

t – період часу, протягом якого виявляються результати впровадженої НДДКР, 3 роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні; для України цей показник знаходиться на рівні 0,1;

t – період часу (в роках) від моменту отримання чистого прибутку до t . 2, 3, 4.

Рисунок, що характеризує рух платежів (інвестицій та додаткових прибутків) буде мати вигляд, зображений на рисунку 4.1.

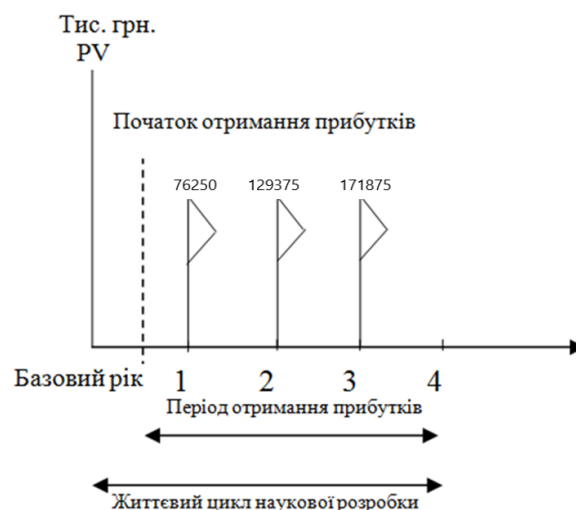


Рисунок 4.1 – Вісь часу з фіксацією платежів, що мають місце під час розробки та впровадження результатів НДДКР

Розрахуємо вартість чистих прибутків за формулою:

$$ПП = \sum_1^m \frac{\Delta\Pi_i}{(1+\tau)^t} \quad (4.11)$$

де $\Delta\Pi_i$ – збільшення чистого прибутку у кожному із років, протягом яких виявляються результати виконаної та впровадженої НДДКР, грн;

τ – період часу, протягом якого виявляються результати впровадженої НДДКР, роки;

τ – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні; для України цей показник знаходиться на рівні 0,1;

t – період часу (в роках) від моменту отримання чистого прибутку до точки.

Отже, розрахуємо вартість чистого прибутку:

$$ПП = \frac{44409,63}{(1+0,1)^0} + \frac{76250}{(1+0,1)^2} + \frac{129375}{(1+0,1)^3} + \frac{171875}{(1+0,1)^4} = 322020,45(\text{грн.})$$

Тоді розрахуємо $E_{абс}$:

$$E_{абс} = 322020,45 - 44409,63 = 277610,82 \text{ грн.}$$

Оскільки $E_{абс} > 0$, то вкладання коштів на виконання та впровадження результатів НДДКР буде доцільним.

Розрахуємо відносну (щорічну) ефективність вкладених в наукову розробку інвестицій E_B за формулою:

$$E_B = \sqrt[T]{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.12)$$

де $E_{абс}$ – абсолютна ефективність вкладених інвестицій, грн;

PV – теперішня вартість інвестицій $PV = 3B$, грн;

$T_{ж}$ – життєвий цикл наукової розробки, роки.

Тоді будемо мати:

$$E_B = \sqrt[3]{1 + \frac{277610,82}{44409,63}} - 1 = 1,69 \text{ або } 169 \%$$

Далі, розраховану величина E_B порівнюємо з мінімальною (бар'єрною) ставкою дисконтування τ_{\min} , яка визначає ту мінімальну дохідність, нижче за яку інвестиції вкладатися не будуть. У загальному вигляді мінімальна (бар'єрна) ставка дисконтування τ_{\min} визначається за формулою:

$$\tau = d + f,$$

де d – середньозважена ставка за депозитними операціями в комерційних банках; в 2019 році в Україні $d = 0,2$;

f – показник, що характеризує ризикованість вкладень, величина $f = 0,1$.

$$\tau = 0,2 + 0,1 = 0,3$$

Оскільки $E_B = 169\% > \tau_{\min} = 0,3 = 30\%$, то у інвестор буде зацікавлений вкладати гроші в дану наукову розробку.

Термін окупності вкладених у реалізацію наукового проекту інвестицій. Термін окупності вкладених у реалізацію наукового проекту інвестицій $T_{ок}$ розраховується за формулою:

$$T_{ок} = \frac{1}{E_B},$$
$$T_{ок} = \frac{1}{1,69} = 0,69 \text{ року.}$$

Обрахувавши термін окупності даної наукової розробки, можна зробити висновок, що фінансування даної наукової розробки буде доцільним.

ВИСНОВКИ

У ході виконання магістерської кваліфікаційної роботи розроблено інформаційну технологію розпізнання емоцій людини на основі згорткової нейронної мережі з вдосконаленою архітектурою.

Здійснено аналіз предметної області та досліджено основні методи розпізнавання емоцій. У ході аналізу було визначено, що використання нейронної мережі при розпізнанні емоцій може забезпечити високу точність і не потребує достатньо великої кількості ресурсів комп'ютера, що актуально під час розпізнавання емоцій у непевному відеопотоці.

Виконано огляд та аналіз існуючих програмних засобів, що дозволяють розпізнавати емоції людини. Визначено, що більшість розглянутих програмних рішень не можуть забезпечити достатньо точне розпізнання емоції саме у неперервному відеопотоці.

Доведено доцільність розробки інформаційної системи, представленої у вигляді програмного додатку. Наведено та обґрунтовано вимоги до розроблюваного програмного продукту.

Після проведеного аналітичного огляду існуючих реалізацій основних етапів технології розпізнання емоцій людини було доведено: для попередньої обробки зображення слід використовувати метод каскаду Хаара; для отримання візуальних ознак доцільно використовувати фільтр Габора, для класифікації емоцій варто застосовувати згорткову нейронну мережу.

Запропоновано математичну модель інформаційної технології розпізнавання емоцій людини, на основі якої базується згорткова нейронна мережа, та шари, які входять до складу її архітектури.

У результаті аналізу існуючих архітектур згорткових нейронних мереж, було обрано згорткову нейронну мережу, запропоновану у роботі Гуді, як найбільш перспективну для подальшого дослідження. Архітектуру обраної нейронної мережі удосконалено шляхом додавання згорткового проміжного шару, та доведено доцільність такого удосконалення.

Обґрунтовано вибір FEREC-2013, як набору даних для навчання вдосконаленої згорткової нейронної мережі, який відрізняється великим розміром набору даних і різноманітністю його елементів.

Запропоновано структуру інформаційної технології для розпізнавання емоцій людини, побудованої на основі вдосконаленої архітектури згорткової нейронної мережі. На основі структури інформаційної технології визначено алгоритм роботи програмного засобу для розпізнавання емоцій людини, який складається з трьох модулів: навчання, розпізнавання та валідації, а також розроблено алгоритм роботи кожного з модулів.

Обґрунтовано доцільність застосування мови програмування Python для розробки програмного засобу. Для програмної реалізації згорткової нейронної мережі обрано нейромережеву бібліотеку Keras. Програмна реалізація програмного додатку виконана у середовищі програмування Visual Studio Code.

На основі здійснених досліджень розроблено програмне забезпечення, що дозволяє практично реалізувати технологію розпізнавання емоцій людини.

Здійснено тестування розробленого програмного додатку, яке підтвердило коректність його роботи.

Розроблено автоматизовану систему тестування додатку. Під час тестування точність становила близько 90% для відформатованого статичного набору даних, а під час валідації у відеопотоці – 62% (найближчий, за характеристиками аналог, додаток FaceReader – 56%, тобто точність зросла в середньому на 6 – 7%). У результаті тестування суттєвого збільшення навантаження на апаратні ресурси комп'ютера та втрати кадрів не виявлено. Отже, можна констатувати, що програма перевершує можливості своїх аналогів

У результаті виконання даної кваліфікаційної роботи поставлені задачі було виконано у повному обсязі. Отже, мета магістерської роботи досягнута

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Зінов'єв Є., Арсенюк І. Дослідження методів розпізнавання емоцій за допомогою нейронних мереж // Матеріали XLIX науково-технічної конференції підрозділів ВНТУ. Вінниця, 2020.
2. Матеріали XI міжнародної науково-практичної конференції «ІОН-2020» [Електроний ресурс] – Режим доступу: <http://ies.vntu.edu.ua/reports/program/WORK-IES-2020.pdf>
3. Свідоцтво про реєстрацію авторського права на твір. «Нейромережевий модуль колоризації чорно-білих зображень» / Арсенюк І. Р., Зінов'єв Є. В. .
4. Брилюк, Д. Распознавание человека по изображению лица и нейросетевые методы // Д. Брилюк, В. Старовойтов. – Минск: Институт Технической Кибернетики Национальной Академии Наук Беларуси, 2001.
5. Глазунов, А. Компьютерное распознавание человеческих лиц // Открытые Системы. 2000. № 3.
6. 2 Gavrilu D. M., Giebel J., Munder S. Vision-based pedestrian detection: the protector system // Proceedings of the IEEE Intelligent Vehicles Symposium, Parma, Italy. – 2004. – pp. 13-18.
7. Хейдоров И. Э., Янь Цзинбинь, У Ши, Сорока А. М., Трус А. А. Классификация эмоционально окрашенной речи с использованием метода опорных векторов // Речевые технологии. – 2008. – №3. – С. 53-61.
8. Viola, P. Rapid object detection using a boosted cascade of simple features / P. Viola // IEEE Conf. on Computer Vision and Pattern Recognition. – Kauai, Hawaii, USA, 2001. – V. 1. – P. 511–518.
9. Viola, P. Robust realtime face detection / P. Viola// International Journal of Computer Vision. – 2004. – V. 57. – № 2. – P. 137–154.
10. Karungaru, S. Human Face Detection In Visual Scenes Using Neural Networks / S. Karungaru, M. Fukumi, N. Akamatsu // Proc. of ICONIP 2000. 2000. P. 1-6.

11. Lin, S.-H. Face Recognition Detection by Probabilistic Decision-Based Neural Network / S.-H. Lin, S.-Y. Kung, L.-J. Lin // IEEE Trans. Neural Networks. 1997. Vol. 8. P. 114-132.
12. Feraud, R. A fast and accurate face detector based on neural networks / R. Feraud, O. Bemier, J.-E. Viallet // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. № 23. P. 42-53.
13. Исследование рынка систем распознавания эмоций [Электронный ресурс] – Режим доступа: <https://habr.com/ru/post/133686/>
14. About Noldus [Electronic resource] – Mode of access: <https://www.noldus.com/about-noldus>
15. Распознавание эмоций по лицу - EmoDetect 3.0 [Электронный ресурс] – Режим доступа: <https://neurobotics.ru/catalog/psixofiziologiya/raspoznvanie-emoczij-po-liczu-emodetect/>
16. This Microsoft tool recognizes emotions in pictures [Electronic resource] – Mode of access: <https://blogs.microsoft.com/ai/happy-sad-angry-this-microsoft-tool-recognizes-emotions-in-pictures/>
17. Распознавание выражений лица [Электронный ресурс] – Режим доступа: <https://medium.com/@neurohiveru>
18. Varma S., Shinde M., Chavan S.S. Analysis of PCA and LDA features for facial expression recognition using SVM and HMM classifiers. Techno-Societal 2018. Proc. 2nd International Conference on Advanced Technologies for Societal Applications, vol. 1, 2020, pp. 109–119. doi: 10.1007/978-3-030-16848-3_11
19. Viola P., Jones M.J. Robust real-time face detection. International Journal of Computer Vision, 2004, vol. 57, no. 2, pp. 137–154.
20. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C.-Y., Berg A.C. SSD: single shot multibox detector. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2016, vol. 9905, pp. 21–37.

21. Déniz O., Bueno G., Salido J., De la Torre F. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters*, 2011, vol. 32, no. 12, pp. 1598–1603.
22. Gross R. Face databases. In: Li S.Z., Jain A.K., editors. *Handbook of Face Recognition*. Springer Verlag. 2005. pp. 301–327.
25. Phillips P.J., Moon H., Rizvi S.A., Rauss P.J. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000. 22(10):1090–1104.
23. Owusu E., Abdulai J.-D., Zhan Y. Face detection based on multilayer feed-forward neural network and Haar features. *Software: Practice and Experience*, 2019, vol. 49, no. 1, pp. 120–129.
24. Mohan P.G., Prakash C., Gangashetty S.V. Bessel transform for image resizing. *Proc. 18th International Conference on Systems, Signals and Image Processing (IWSSIP 2011)*, Sarajevo, Bosnia-Herzegovina, 2011, pp. 75–78.
25. Naz S., Ziauddin S., Shahid A.R. Driver fatigue detection using mean intensity, SVM, and SIFT. *International Journal of Interactive Multimedia and Artificial Intelligence*, 2019, vol. 5, no. 4, pp. 86–93.
26. Priya R.V. Emotion recognition from geometric fuzzy membership functions. *Multimedia Tools and Applications*, 2019
27. Oloyede M., Hancke G., Myburgh H., Onumanyi A. A new evaluation function for face image enhancement in unconstrained environments using metaheuristic algorithms // *EURASIP Journal on Image and Video Processing*. 2019.
28. Fawwad Hussain M., Wang H., Santosh K.C. Gray level face recognition using spatial features. *Communications in Computer and Information Science*, 2019
29. Ahsan T., Jabid T., Chong U.P. Facial expression recognition using local transitional pattern on Gabor filtered facial images // *IETE Technical Review*. 2013

30. Martínez A.M., Kak A.C. PCA versus LDA. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, vol. 23, no. 2, pp. 228–233
31. Babu D.R., Shankar R.S., Mahesh G., Murthy K.V.S.S. Facial expression recognition using bezier curves with hausdorff distance // Proc. IEEE International Conference on IoT and Application (ICIOT 2017). Nagapattinam, India. 2017
32. Степанова О.А., Ивановский Л.И., Хрящев В.В. Использование глубокого обучения и сверточных нейронных сетей для анализа выражения лица // DSPA: Вопросы применения цифровой обработки сигналов. 2018. Т. 8. № 4. С. 170–173.
33. 8. Talegaonkar I., Joshi K., Valunj S., Kohok R., Kulkarni A. Real time facial expression recognition using deep learning // Proc. of International Conference on Communication and Information Processing (ICCIPI). 2019 [Электронный ресурс]. Режим доступа: <https://ssrn.com/abstract=3421486>.
34. Гудфеллоу Я., Бенджио И., Курвилль А. Г93 Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.: цв. ил.
35. A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images, 2009.
36. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
37. A. Gudi. Recognizing semantic features in faces using deep learning. arXiv preprint arXiv:1512.00743, 2015.
38. Matt Taddy. Stochastic Gradient Descent // Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions. — New York: McGraw-Hill, 2019
39. Kaggle. Challenges in representation learning: Facial expression recognition challenge, 2013.

40. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 94–101. IEEE, 2010.
41. Jesorsky O., Kirchberg K., Frischholz R. Face detection using the Hausdorff distance. In Bigun J., Smeraldi F., editors. Audio and Video based Person Authentication. LNCS Springer. 2001. pp. 90 95.
42. Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение / Плас Дж. Вандер — СПб.: Питер, 2018. — 576 с.:
43. Рашка С. «Python и машинное обучение» / пер. с англ. А. В. Логунова. - М.: ДМК Пресс, 2017. - 418 с.:
44. Жерон, Орельен. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем./ Пер. с англ. - СПб.: ООО "Альфа-книга": 2018. - 688 с.:
45. This Is What Makes Keras Different, According To Its Author [Electronic resource] – Mode of access: <https://www.forbes.com/sites/quora/2016/08/25/this-is-what-makes-keras-different-according-to-its-author/?sh=341957d566cf>
46. VS Code [Electronic resource] – Mode of access: <https://code.visualstudio.com/docs/getstarted/locales>
47. O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. van Knippenberg. Presentation and validation of the radboud faces database. Cognition and emotion, 24(8):1377– 1388, 2010.
48. Jesorsky O., Kirchberg K., Frischholz R. Face detection using the Hausdorff distance. In Bigun J., Smeraldi F., editors. Audio and Video based Person Authentication. LNCS Springer. 2001. pp. 90 95.
49. Caltech 10000 Web Faces. Human Faces Collected from Google Image Search. CalTech University. Pasadena, California, USA [Internet]. 2005.

[Електронний ресурс]. – Режим доступу: www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces. / – Дата доступу : 21.09.2020.

50. Яровий А. А. Розпізнавання мімічних мікровиразів обличчя людини на основі Time Delay Neural Network / Яровий А. А., Кашубін С. Г., Кулик О. О., Липкань І. М. // Вісник Хмельницького національного університету. Технічні науки. – 2015. – № 1. – С. 122 – 126.

51. Lienhart R., Kuranov E., Pisarevsky V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: PRS 2003, pp. 297-304 (2003)