

Ісаєв Т.С.

Хмельницький національний університет

Кисіль Т.М.

Хмельницький національний університет

МЕТОД РОЗПІЗНАВАННЯ ЕМОЦІЙ ЛЮДИНИ У РЕАЛЬНОМУ ЧАСІ ЗА ДОПОМОГОЮ МАШИННОГО ЗОРУ

У наш час існують різні методи і різні ознаки, які можуть бути використані для розпізнавання обличчя і емоцій, таких як обличчя, текст, мова і так далі. Вираз обличчя серед них є найкращою. Виявлення емоцій також широко використовується у багатьох сферах, таких як розуміння поведінки людини, виявлення психічних розладів, виявлення емоцій людей у натовпі тощо. У запропонованій концепції розпізнавання емоцій обличчя можна розділити на три етапи. На першому етапі відбувається виявлення людського обличчя на зображенні з камери, потім на другому етапі вхідні дані аналізуються на основі ознак за допомогою моделі згорткових нейронних мереж (CNN). На останньому етапі відбувається класифікація людських емоцій на 7 основних категорій: страх, гнів, відраза, сум, здивування, нейтральний та радість. Таким чином, запропонована модель розділена на три основні цілі: виявлення обличчя, розпізнавання обличчя та класифікація емоцій. Розпізнавання обличчя має багато важливих застосувань у різних сферах, таких як верифікація особи, безпека, біометрія, смарт-картки, системи спостереження тощо. Міміка передає невербальні сигнали, які відіграють важливу роль у міжособистісних відносинах. Система Cognitive Emotion AI – це процес ідентифікації емоційного стану людини. Основною метою нашого дослідження є розробка надійної системи, яка може виявляти та розпізнавати людські емоції з прямих трансляцій. Методологія цієї системи базується на двох етапах – розпізнавання обличчя здійснюється шляхом вилучення каскаду Хаара та розпізнавання рис обличчя за допомогою алгоритму Віолі-Джонса, а потім відбувається перевірка та розпізнавання емоції за допомогою методів штучного інтелекту. Система приймає зображення або кадр в якості вхідних даних, і шляхом аналізу зображення моделі, виконується попередня обробка та виділення рис обличчя, після чого система прогнозує емоційний стан. У даній роботі було зроблено згорткову нейронну мережу для класифікації людських емоцій за динамічною мімікою обличчя в реальному часі. Також, було проведено навчання нашої системи з використанням моделі Tensorflow, яка була навчена класифікувати людські емоції. Було використано відкриту базу даних із зображеннями для навчання моделі, а також власні фото для покращення результату. Для відображення результатів було підключено живий відеопотік. Отримані результати демонструють можливість застосування нейронних мереж в реальному часі для виявлення людських емоцій.

Ключові слова: розпізнавання емоцій, аналітика, тренування моделі, реальний час, машинне навчання, штучний інтелект.

Постановка проблеми. Емоції часто опосередковують і полегшують взаємодію між людьми. Таким чином, розуміння емоцій часто вносить контекст у, здавалося б, химерну та/або складну соціальну комунікацію. Емоції можна розпізнати за допомогою цілої низки засобів, як-от: голос, інтонація, мова тіла та більш складними методами, такими як електроенцефалографія (ЕЕГ) [11]. Однак, більш простим і практичним методом є дослідження міміки. Існує сім типів людських емоцій, які, як було доведено, є універсальними для всіх людей [2]: гнів, відраза, страх, щастя, смуток, здивування та нейтральний вираз. Таким чином, утиліта, яка розпізнає емоції за виразом

обличчя, могла б знайти широке застосування. Завдання розпізнавання емоцій є особливо складним з двох причин: 1) не існує великої бази даних навчальних зображень і 2) класифікація емоцій може бути складною в залежності від того, чи є вхідне зображення статичним, чи зображення рухається. Більшість програм розпізнавання емоцій досліджують статичні зображення виразу обличчя. Дослідження застосування згорткових нейронних мереж (ЗНМ) для розпізнавання емоцій у реальному часі за допомогою вхідного відеопотоку є дуже важливими на сьогодні. Враховуючи обчислювальні вимоги та складність штучної моделі, оптимізація мережі для ефективних

обчислень для покадрової покадрової класифікації є необхідною. Крім того, необхідно враховувати варіації освітлення та положення об'єктів. Було розроблено систему для виявлення людських емоцій у різних сценах, ракурсах та умовах освітлення в режимі реального часу.

Аналіз останніх досліджень і публікацій. За останні два десятиліття дослідники значно просунулися в розпізнаванні людських емоцій на обличчі за допомогою комп'ютерного зору. Історично склалося так, що існувало багато підходів до вирішення цієї проблеми, включаючи використання пірамідальних гістограм градієнтів (PHOG) [12], AU, що враховує риси обличчя, покращені LBP-дескриптори та RNN [13]. Однак нещодавні найкращі роботи, подані на конкурс «Емоції в дикій природі» 2020 року [18] використовували глибокі згорткові нейронні мережі (CNN), що дають до 62 % точності тесту.

Нещодавня розробка G. Levi et. al [14] показала значне покращення розпізнавання емоцій на обличчі за допомогою штучного інтелекту. Головною проблемою для дослідників є: 1) мала кількість даних, доступних для навчання глибоких нейронних мереж, та 2) варіації зовнішнього вигляду, які зазвичай спричинені варіаціями освітлення. Вони використовували локальні

бінарні патерни (LBP) для перетворення зображення в інваріантний до освітлення тривимірний простір, який може слугувати вхідними даними для CNN.

Ця спеціальна обробка даних була застосована до різних загальнодоступних модулів, таких як VGG S [15]. Фінальні результати показали точність тесту до 54,56 %. Існує багато підходів до навчання моделі, але вони не мають аналізу точності розпізнавання емоцій, що є великим недоліком тому, що аналіз точності моделі у реальному часі допоможе покращити саму модель та краще зрозуміти які фактори впливають на її роботу та що потрібно для отримання оптимального результату щоб використовувати модель більш досконало.

Постановка завдання. Метою роботи є створення моделі, що дозволить аналізувати емоційний стан людини у реальному часі.

Виклад основного матеріалу. Створення та дослідження моделі із застосуванням машинного та штучного інтелекту відрізняється від розробки додатків, програм та ігор у тому, що процес моделювання включає постійні дослідження та оновлення даних, здійснюючи експерименти, щоб навчити нейронну мережу розрізняти об'єкти та емоції людини. У машинного навчання власний життєвий цикл, що означає проходження даних через певні етапи розробки та впровадження системи прогнозування. В порівнянні з циклом розробки програмного забезпечення, створення моделей машинного навчання включає етап експериментування з наборами даних для досягнення поставлених цілей і завдань, при цьому використовуються нові дані після проходження етапу навчання.

Початок роботи із моделлю та її тренування

На рис. 1 показано структуру запропонованої в цьому дослідженні системи. Як показано на рис. 1, було використано два підходи для виявлення емоцій суб'єкта: виявлення емоцій за допомогою мімічних орієнтирів та виявлення емоцій за допомогою сигналів ЕЕГ [11].

Для тренування моделі необхідно було обрати відкриту базу даних, яка містить картинки людей з різними емоціями та надасть змогу натренувати систему досить гарно. Було обрано базу даних Fer2013 [16]. Ця база даних знаходиться у відкритому доступі, а отже кожен бажаючий без проблем зможе використати картинку без ризику. Вона містить досить багато картинок людей різного віку, які показують різні емоції. Саму систему баз даних із картинками зображено на рис. 2.

У цьому дослідженні було використано HD-камеру для зйомки обличчя суб'єктів та створення

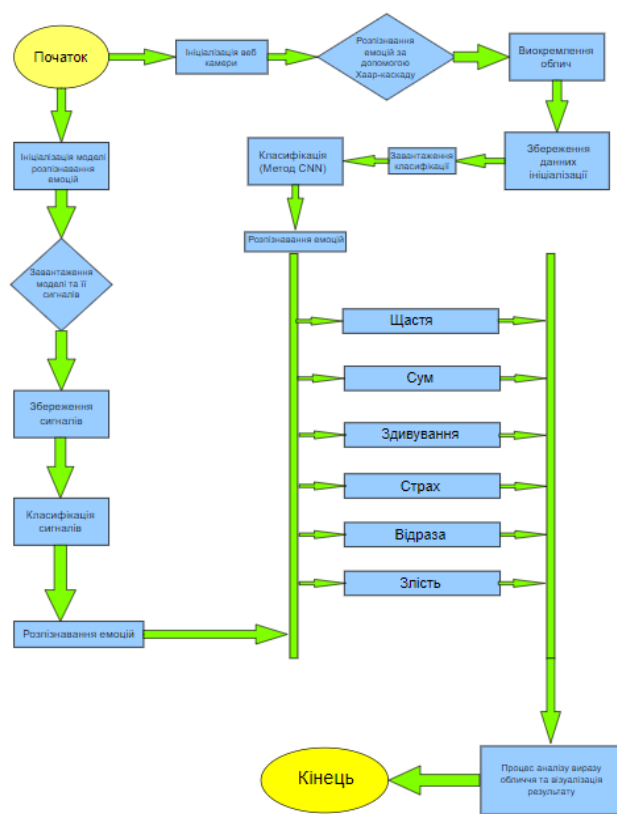


Рис. 1. Процес ініціалізації програми

зображення у відтінках сірого(додатково до наявних зображень у системі Fer2013). Це спростило процес розпізнавання виразу обличчя при розпізнаванні міміки. Потім, використовуючи зображення у відтінках сірого, визначаються очі суб'єкта, і десять віртуальних маркерів (одиниць дії) розміщуються на обличчі суб'єкта у визначених місцях за допомогою математичної моделі.

Алгоритм читання емоцій із зображення ґрунтується на алгоритмі їхнього розпізнавання, який використовується, наприклад, у FaceID. Для вирішення цих завдань із зображень витягуються спеціальні дескриптори облич, або вектори ознак. Спочатку мережа глибокого навчання тренується розпізнавати риси обличчя, роблячи це за допомогою карти точок навколо основних його частин (очі, брови, ніс, губи). Відстань між цими точками і є вектор ознак. Подібно до відбитка пальця, вектор ознак буде індивідуальний для кожної людини. Хоча риси обличчя й індивідуальні, вираз одних і тих самих емоцій однаковий у більшості людей. Для їхнього розпізнавання використовують групи точок із вектора ознак. Так, наприклад, підйом групи точок біля куточків рота свідчить про посмішку, зведення групи точок на бровах – про насупленість тощо. Процес розпізнавання зображень на рис. 3.

Алгоритм оптичного потоку Лукаса-Канде використовується для передачі положення кожного віртуального маркера, щоб відстежити його положення під час емоційного вираження суб'єкта. Десять ознак виводяться як відстань між кожним маркером і точкою. У поточному дослідженні всі дані про відстані обчислюються за допомогою теореми Піфагора. Потім вони зберігаються у форматі CSV під час процесу збору даних для подальшої обробки.

Основна функція цього методу – це дослідження відстаней між очима та носом, носом та ротом тощо. Це дає змогу зрозуміти почуття людини внаслідок дослідження її міміки та особливостей виразу обличчя. Звісно, у кожної людини свій власний вираз обличчя, але за допомогою цього методу можна дослідити різні емоції та навчити модель розпізнавати усі ознаки міміки та виразів обличчя людини. Цей процес є досить складним, але внаслідок розвитку технологій, цю систему можна досить



Рис. 2. Фрагмент системи баз даних FER2013[16]



Рис. 3. Процес розпізнавання емоцій [2]

легко імплементувати за допомогою штучного інтелекту та машинного навчання. Увесь процес розпізнавання обличчя людини зображений на рис. 4, де вказано більшість особливостей обличчя людини та процес його ідентифікації.

Згорткова нейронна мережа (CNN) була використана в нашій системі для покращення розпізнавання емоцій обличчя, оскільки вона застосовується в інших комп'ютерних галузях, таких як розпізнавання обличчя та виявлення об'єктів. Крім того, передбачення ґрунтуються на інформації, наданій у певний момент часу.

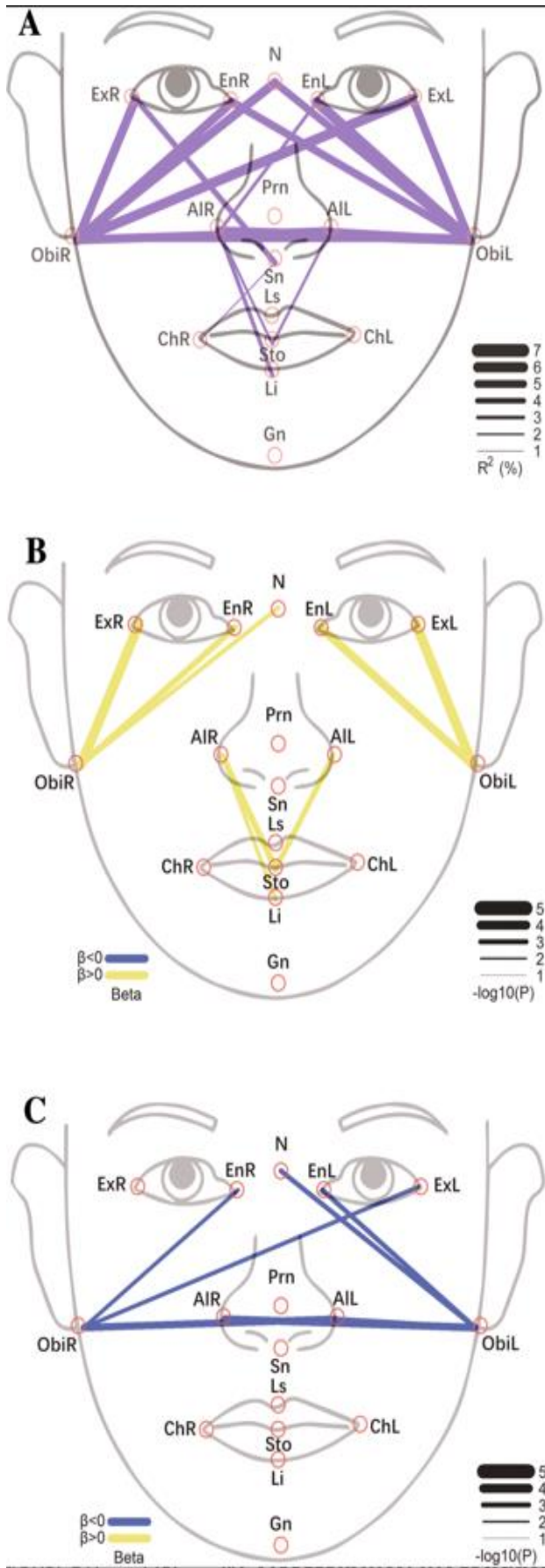


Рис. 4. Процес розпізнання обличчя людини [2]

На рис. 5 показано структуру мережі, яка використовується для розпізнання емоцій за допомогою орієнтирів на обличчі. Ця мережа отримує вхідне зображення і намагається передбачити вихідну емоцію. Вона має вісім етапів, включаючи згортки, об'єднання і повністю з'єднані шари з операціями випрямлених лінійних одиниць (ReLU), які зберігають хорошу якість, але роблять збіжність набагато швидшою. Кількість фільтрів становила 32, 64 і 128 з розміром фільтра 5×5 для згорткових шарів, а кількість вихідних вузлів у повністю з'єднаному шарі становила 6 з оптимізатором “Adam” [19] і коефіцієнтом відсіву 0,3.

Щоб виконати поставлене завдання, необхідно створити власну програму, яка дозволить натренувати систему та перевірити отримані знання у реальному часі способом підключення машинного зору.

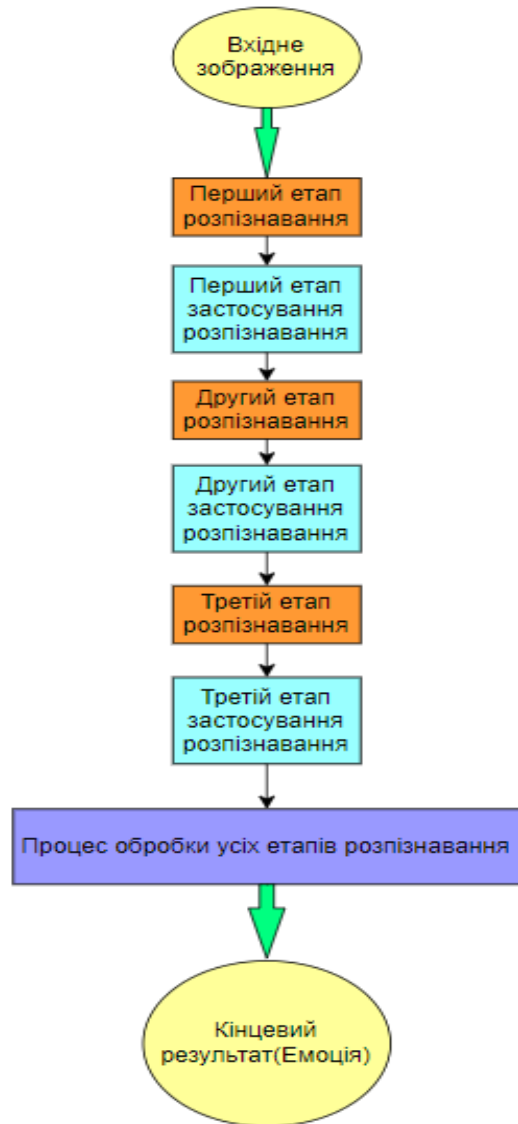


Рис. 5. Процес розпізнання системою CNN

Для тренування було використано середовище Jupyter Notebook [17] та мову програмування Python.

Для ефективної роботи з кодом потрібно завантажити додатки до Python – пакети, які спеціально розроблені для специфічних задач та значно спрощують роботу з кодом та його швидкість. Було використано такі пакети: Tensorflow, Keras, Opencv, pandas, numpy, os. Завантаження додатків можна побачити на рис. 6.

Після підключення усіх додатків, необхідно завантажити зображення всередину моделі для подальшого навчання. Процес завантаження зображений на рис. 7, де картинки вже розділені на навчальні та тренувальні (навчальні для процесу навчання моделі, а тренувальні – самоперевірки моделі та оцінки якості тренування).

Після завантаження зображень потрібно налаштувати модель та характеристики тренування для більш точного підходу системи до навчання. Налаштування моделі для нашої системи зображено на рис. 8. Було специфіковано налаштування нашої моделі та задаємо параметри, необхідні для навчання.

Після налаштування характеристик для тренування, відбувається саме тренування моделі. Цей процес є довготривалим та може зайняти дуже багато часу. Час тренування залежить від характе-

ристик пристрою, на якому відбувається процес. По закінченню ми отримуємо графіки щодо точності та коректності моделі (рис. 9).

На графіку, який зображений на рис. 9, відображена точність під час навчання та самоперевірки моделі внаслідок багатьох етапів перевірки та навчання. Для покращення точності потрібно сформулювати нові методи тренування, більш потужніші пристрої для проведення процесу навчання та новіші дані/зображення для покращення розуміння міміки людей системою. Перевірка результату тренування та передбачення моделі зображено на рис. 10. Спочатку, була вказана емоція, яку бажаємо отримати від моделі, а потім сама модель намагається передбачити емоцію, аналізуючи картинку та міміку людини яка зображена на фото. Очікуваний результат – це сум. Модель вказала, що це сум, а отже вона досить добре відрізняє емоції по фото. Ще один приклад емоції зображено на рис. 11. Очікуваний результат – відраза, що система нам і показала.

Отже, результат тренування є досить високим так, як модель здатна розпізнати емоції на картинках.

Ще один спосіб перевірити модель – це протестувати її у реальному часі, щоб упевнитись чи відповідає модель нашим вимогам, чи ні. Результат перевірки зображений на рис. 12.

```
[6]: from keras.utils import to_categorical
      from keras_preprocessing.image import load_img
      from keras.models import Sequential
      from keras.layers import Dense, Conv2D, Dropout, Flatten, MaxPooling2D
      import os
      import pandas as pd
      import numpy as np
```

Рис. 6. Завантаження пакетів для Python

```
[10]: train_features = extract_features(train['image'])
      100% ██████████ 28821/28821 [08:34<00:00, 80.38it/s]
      C:\Users\timha\AppData\Local\Programs\Python\Python311\Lib\site-packages\keras
      ease use color_mode = "grayscale"
      warnings.warn('grayscale is deprecated. Please use '
[11]: test_features = extract_features(test['image'])
      100% ██████████ 7066/7066 [02:04<00:00, 70.87it/s]
```

Рис. 7. Процес завантаження зображень

```

model = Sequential()
# convolutional layers
model.add(Conv2D(128, kernel_size=(3,3), activation='relu', input_shape=(48,48,1)))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Conv2D(256, kernel_size=(3,3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Conv2D(512, kernel_size=(3,3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Conv2D(512, kernel_size=(3,3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.4))

model.add(Flatten())
# fully connected layers
model.add(Dense(512, activation='relu'))
model.add(Dropout(0.4))
model.add(Dense(256, activation='relu'))
model.add(Dropout(0.3))
# output layer
model.add(Dense(7, activation='softmax'))
    
```

Рис. 8. Налаштування моделі для тренування

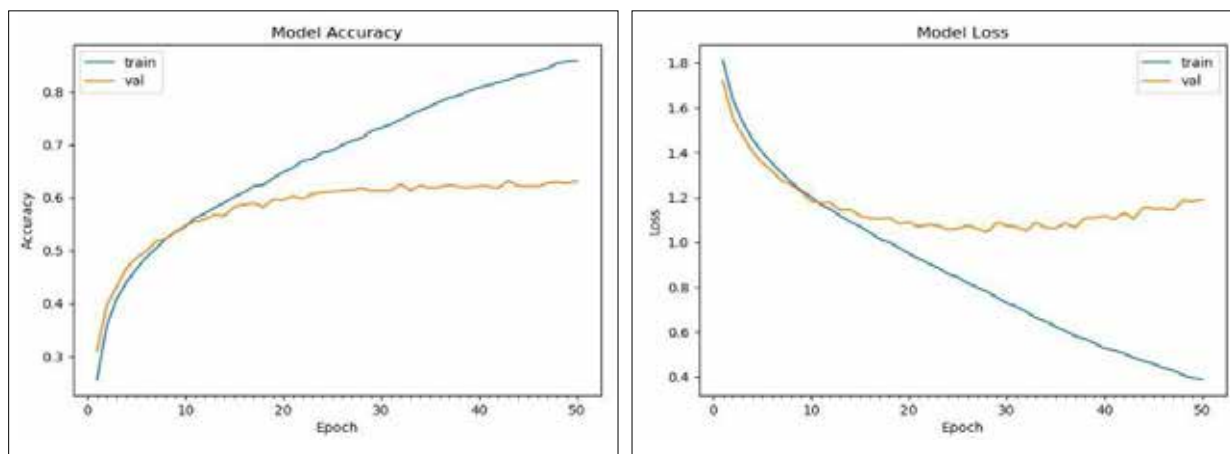


Рис. 9. Точність моделі



Рис. 10. Передбачення емоції (Сум)



Рис. 11. Передбачення емоції (Відраза)

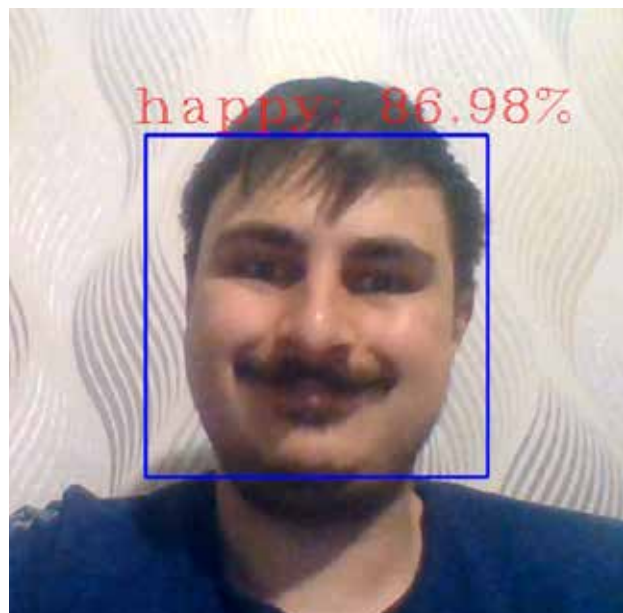


Рис. 12. Передбачення емоції у реальному часі

Висновки. У даному дослідженні було розглянуто проблематику розпізнавання міміки людини та проаналізовано спосіб, що допоможе вирішити цю проблему та дозволить покращити взаємодію людини та машини у майбутньому. Було створено модель, яка розпізнає емоції людини у реальному часі. Модель було навчено за допомогою нейромережі та відкри-

тої бази даних із зображенням емоцій різних людей. Систему було протестовано на статичних зображеннях та у форматі реального відеопотоку.

Було проведено аналіз точності моделі та варіанти її можливого покращення. Загалом, модель показала досить непогані результати у розпізнаванні емоцій.

Список літератури:

1. Zhang Ke, et al. Real-time video emotion recognition based on reinforcement learning and domain knowledge. *IEEE Transactions on Circuits and Systems for Video Technology*. 2021. № 32.3. P. 1034–1047.
2. Hassouneh Aya, Mutawa A. M., Murugappan M. Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*. 2020. Vol. 20. P. 1057–1120.
3. Ozdemir, Mehmet Akif et al. Real time emotion recognition from facial expressions using CNN architecture. In: *2019 medical technologies congress (tiptekno)*. IEEE. 2019. P. 1–4.

4. Lech Margaret et al. Real-time speech emotion recognition using a pre-trained image classification network: Effects of bandwidth reduction and companding. *Frontiers in Computer Science*. 2020. Vol. 2. P. 1–14.
5. Pathar Rohit, et al. Human emotion recognition using convolutional neural network in real time. In: *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. IEEE. 2019. P. 1–7.
6. Pang Bo, Nijkamp Erik, WU Ying Nian. Deep learning with tensorflow: A review. *Journal of Educational and Behavioral Statistics*. 2020. vol. 45.2. P. 227–248.
7. Grattarola, Daniele, Alippi, Cesare. Graph neural networks in tensorflow and keras with spektral [application notes]. *IEEE Computational Intelligence Magazine*. 2021. Vol. 16.1. P. 99–106.
8. Keshri, Ashish, et al. Automatic detection and classification of human emotion in real-time scenario. *Journal of IoT in Social, Mobile, Analytics, and Cloud*. 2022. Vol. 4.1. P. 41–53.
9. Raschka, Sebastian, Patterson, Joshua, Nolet, Corey. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*. 2020. Vol. 11.4. P. 40–193.
10. Khare, Smith K., Bajaj, Varun. Time–frequency representation and convolutional neural network-based emotion recognition. *IEEE transactions on neural networks and learning systems*. 2020. Vol. 32.7. P. 2901–2909.
11. Muller-Putz, Gernot R. Electroencephalography. *Handbook of Clinical Neurology*. 2020. Vol. 168. P. 249–262.
12. Kaya, Davut, et al. Swin-Phog-LPQ: An accurate computed tomography images classification model using Swin architecture with handcrafted features. *Biomedical Signal Processing and Control*. 2023. Vol. 86. P. 105–108.
13. Deabes, Wael, Sheta, Alaa, Braik, Malik. ECT-LSTM-RNN: An electrical capacitance tomography model-based long short-term memory recurrent neural networks for conductive materials. *IEEE Access*. 2021. Vol. 9. P. 76325–76339.
14. Hosseini, Sepidehsadat; CHO, Nam Ik. GF-CapsNet: Using gabor jet and capsule networks for facial age, gender, and expression recognition. In: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE. 2019. P. 1–8.
15. Jin, Xuesong; Du, Xin; Sun, Huiyuan. VGG-S: Improved Small Sample Image Recognition Model Based on VGG16. In: *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*. IEEE, 2021. P. 229–232.
16. Khairuddin, Yousif; Chen, Zhuofa. Facial emotion recognition: State of the art performance on FER2013. *arXiv preprint arXiv:2105.03588*, 2021. P. 170–175.
17. Wang, Jiawei; Li, Li; Zeller, Andreas. Better code, better sharing: on the need of analyzing jupyter notebooks. In: *Proceedings of the ACM/IEEE 42nd international conference on software engineering: new ideas and emerging results*. 2020. P. 53–56.
18. Larradet, Fanny, et al. Toward emotion recognition from physiological signals in the wild: approaching the methodological issues in real-life data collection. *Frontiers in psychology*. 2020. Vol. 11. P. 1111.
19. Bock, Sebastian; WEIß, Martin. A proof of local convergence for the Adam optimizer. In: *2019 international joint conference on neural networks (IJCNN)*. IEEE. 2019. P. 1–8.

Isaiev T.S., Kysil T.M. A METHOD FOR RECOGNIZING HUMAN EMOTIONS IN REAL TIME USING MACHINE VISION

Nowadays, there are different methods and different features that can be used to recognize faces and emotions, such as face, text, speech, and so on. Facial expression is the best among them. Emotion detection has also been widely used in many fields such as understanding human behavior, detecting mental disorders, detecting the emotions of people in a crowd, and so on. In the proposed concept, face emotion recognition can be divided into three stages. At the first stage, a human face is detected in the camera image, then at the second stage, the input data is analyzed based on features using a convolutional neural network (CNN) model. At the last stage, human emotions are classified into 7 main categories: fear, anger, disgust, sadness, surprise, neutral, and joy. Thus, the proposed model is divided into three main goals: face detection, face recognition, and emotion classification. Face recognition has many important applications in various fields, such as identity verification, security, biometrics, smart cards, surveillance systems, etc. Facial expressions convey non-verbal signals that play an important role in interpersonal relationships. The Cognitive Emotion AI system is a process of identifying a person's emotional state. The main goal of our research is to develop a robust system that can detect and recognize human emotions from live broadcasts. The methodology of this system is based on two stages – face recognition is performed by extracting the Haar cascade and recognizing facial features using the Viola-Jones algorithm, and then emotion verification and recognition is performed using artificial intelligence methods. The system takes an image or a frame as input, and by analyzing the

model image, preprocessing and facial feature extraction is performed, after which the system predicts the emotional state. We have developed a convolutional neural network to classify human emotions from dynamic facial expressions in real time. We trained our system using the Tensorflow model, which was trained to classify human emotions. We used an open database of images to train the model, as well as our own photos to improve the results. A live video stream was connected to display the results. The results demonstrate the possibility of using neural networks in real time to detect human emotions.

Key words: *emotion recognition, analytics, model training, real-time, machine learning, artificial intelligence.*