

## **ФОРМУВАННЯ АУДІОВІЗУАЛЬНОГО ПОРТРЕТУ ЕМОЦІЙНИХ СТАНІВ**

Кравченко Е.Л., Ярмілко А.В.

*Черкаський національний університет імені Богдана Хмельницького*

У 2019 році ринок розпізнавання лиця штучним інтелектом оцінювався у 3,2 мільярда доларів [1]. Станом на сьогодні, люди можуть використовувати сканер свого обличчя для покупок в магазині, доступу до об'єктів, тощо. Але багато де така технологія є забороненою в цілях конфіденційності. Розпізнавання обличчя може бути корисним і шкідливим явищем водночас.

Розпізнавання обличчя, за визначенням, означає технологію, яка здатна розпізнати людину за зображенням її обличчя. Розпізнавання обличчя ґрунтується на складних математичних алгоритмах штучного інтелекту та машинного навчання, які фіксують, зберігають та аналізують риси обличчя, щоб поєднати їх із зображеннями людей у вже існуючій базі даних і інформацією про них у цій базі даних. У цей процес також можна включити багато інших додаткових підпроцесів, наприклад – розпізнавання емоції людини. Розпізнавання обличчя є частиною загального технічного терміну біометрії, який також включає розпізнавання відбитків пальців, відбитків долонь, сканування очей, розпізнавання голосу та підписів.

Разом з цим, постійно розвивається музична індустрія. Вченими доведено, що обрана музика та мелодія конкретного музичного напрямку впливає на емоції, почуття та психіку людини. Крім того, слід враховувати окремі звуки, які також впливають на людину. Наприклад, шум моря допомагає людині налаштуватись на продуктивну роботу, звуки лісу допоможуть людині зняти стрес або відпочити [2]. Біометричні методи аналізу надають засоби для зв'язування певного музичного контенту з конкретними персоналіями за встановленими правилами, зокрема – за емоціями. Така задача є актуальною як для пересічних меломанів, так і для спеціалістів з психології, реабілітації, управління трудовими кондиціями виробничого персоналу.

Існує багато готових алгоритмів розпізнавання обличчя, написаних на мовах програмування Python, C++, Lisp, Prolog або Java, реалізація яких відрізняється трудоміскістю, бюджетними вимогами, орієнтацією на конкретні дослідницькі чи бізнес-цілі [3]. Аналіз використаних у них методів дозволяє вважати перспективними для сервісу з розпізнавання емоції лиця людини та формування набору їх музичного супроводу такі три моделі:

- згортова нейронна мережа (Convolution Neural Network – CNN);
- Vision Transformer;
- залишкова нейронна мережа (Residual Neural Network – ResNet).

Згортова нейронна мережа (CNN) – це клас штучних нейронних мереж, який став домінуючим у різних завданнях що стосуються комп'ютерного зору.

CNN призначений для автоматичного та адаптивного вивчення просторових ієрархій функцій шляхом зворотного розповсюдження за допомогою декількох будівельних блоків, таких як згорткові шари, агрегувальні шари та повноз'єднані шари [4].

Роботу Vision Transformer можна описати наступним чином: першим кроком моделі Vision Transformer є розподіл вхідного зображення на послідовність патчів зображення. Відповідно до методу, зображення ділиться на сегменти 16 x 16; Потім ці ділянки зображення проходять через тренувальний лінійний проєкційний шар. Цей шар виконує роль шару вбудовування і виводить вектори фіксованого розміру; Потім вбудовані позиції додаються до послідовності патчів зображення [5]. Ідея того, як працює вбудовування позиції, продемонстрована на зображенні нижче (рис. 1):



Рисунок 1 – Робота Vision Transformer на прикладі фото міста

Залишкова нейронна мережа (ResNet) – це така штучна нейронна мережа (ANN), яка базується на конструкціях, відомих з досліджень пірамідальних клітин кори головного мозку. Залишкові нейронні мережі використовують пропускання з'єднань, або ярлики, для переходу через деякі шари. Типові моделі ResNet реалізовані з дво- або тришаровими пропусками. Додаткова матриця ваги може бути використана для вивчення пропуску ваг – ці моделі відомі як HighwayNets. Моделі з декількома паралельними пропусками називаються DenseNets [6].

Кожна з оглянутих моделей має свої переваги та недоліки, що, в свою чергу, несе загрозу неточності розпізнавання емоцій та, відповідно, хибного аудіовізуального портрету особистості. Тому є потреба у проведенні ряду дослідів, в яких використовувалися б наведені вище моделі, з метою вибору найбільш ефективної з них щодо ідентифікації емоційних станів. Загальна ж структура запропонованої системи формування аудіовізуального портрету за емоційними станами зображена на рис. 2.

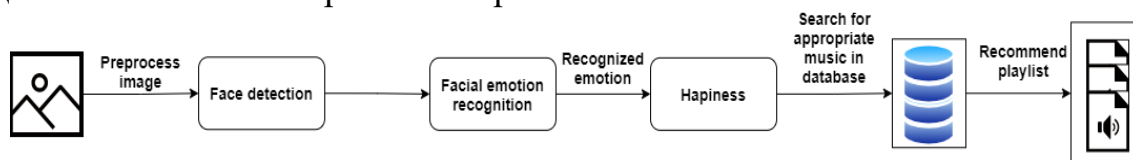


Рисунок 2 – Структура системи формування аудіовізуального портрету за емоційними станами

Вхідне зображення обличчя формату JPG або PNG обробляється нейронною мережею, яка розпізнає лице людини. Далі система ідентифікує емоцію лица, яке вона розпізнала на попередньому кроці. Після визначення всіх необхідних даних, система формує плейлист за набором музичних творів у відповідній базі даних, доповнює його відповідними візуальними атрибутами та пропонує даний контент користувачеві.

Таким чином, проведений етап досліджень дозволив окреслити коло перспективних методів обробки зображень обличчя людини для аналізу її емоцій. Оцінка ефективності цих методів при вирішенні поставленої задачі потребує залучення експериментальних методів. Вирішення задачі оптимального вибору моделі визначення емоційного стану людини має забезпечити належний рівень адекватності аудіовізуального портрету, що створює перспективи використання такої технології в наукових та комерційних цілях.

#### Список літератури:

1. Facial recognition: top 7 trends (tech, vendors, markets, use cases & latest news). Електронний ресурс – Режим доступу: <https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition>
2. Ynpress - Роль музики в житті людей. Електронний ресурс – Режим доступу: <https://ynpress.com/archives/56659#:~:text=B%20различных%20случаях%20музыка%20влияет,поднимать%20настроение%2C%20когда%20становится%20тоскливо.>
3. AI Programming: 5 Most Popular AI Programming Languages. Електронний ресурс – Режим доступу: <https://ncube.com/blog/ai-programming-languages>
4. Convolutional neural networks: an overview and application in radiology. Електронний ресурс – Режим доступу: <https://insightsimaging.springeropen.com/articles/10.1007/s13244-018-0639-9>
5. Vision Transformer (ViT) - Using Transformers for Image Recognition. Електронний ресурс – Режим доступу: <https://www.section.io/engineering-education/vision-transformer-using-transformers-for-image-recognition/>
6. Wikipedia - Residual\_neural\_network. Електронний ресурс – Режим доступу: [https://en.wikipedia.org/wiki/Residual\\_neural\\_network](https://en.wikipedia.org/wiki/Residual_neural_network)