

## FACIAL EMOTION RECOGNITION USING ARTIFICIAL INTELLIGENCE <sup>17</sup>

---

**Eng. Radostin Kolev, MSc, PhD Student**

Department of Telecommunications,  
“Angel Kanchev” University of Ruse  
Tel.: +359 886 107 646  
E-mail: rado\_789@yahoo.com

**Eng. Kristian Velkovski, MSc, PhD Student**

Department of Telecommunications,  
“Angel Kanchev” University of Ruse  
Tel.: +359 898 641 128  
E-mail: kvelkovski66@gmail.com

**Abstract:** This paper reviews the facial emotion recognition paradigm, as a subfield of computer vision and machine learning technologies. Facial emotion recognition harnesses the power of AI algorithms and employs it to analyze facial expressions and to decode the underlying emotions. This paper explores the significance of facial emotion recognition in facilitating more natural and empathetic human-computer interactions, with potential applications spanning over diverse domains including healthcare, entertainment, etc.

**Keywords:** Facial Emotion Recognition, Human-Computer Interaction, Affective Computing, Computer Vision, Machine Learning.

### ВЪВЕДЕНИЕ

Благодарение на изкуствения интелект, взаимодействието между хората и компютрите навлиза в ера на дълбока трансформация. Тъй като технологиите продължават активно да се вплитат във всички дейности от човешкото ежедневие, възможността машините да разбират и интерпретират човешките емоции се превръща в първостепенна задача. Разпознаването на лицеви емоции – сфера, намираща се в пресечната точка между компютърното зрение, машинното обучение и психологията, се очертава като една от най-важните и значими области за провеждане на научни изследвания и експерименти.

Този научен доклад проследява тенденциите в областта на технологиите за разпознаване на лицеви емоции (Facial Expression Recognition – FER) и е фокусиран върху техните особености, технологичните им принципи, приложенията им в реалния свят и предизвикателствата, пред които те са изправени. Докладът проследява симбиозата на човешките емоции и изкуствения интелект и стремежа да преведем богатството на нашите изрази в компютърни алгоритми. Разгледани и анализирани са видовете невронни мрежи, архитектурите и технологиите за дълбоко обучение и наборите от данни, които доведоха до значителния напредък в тази предметна област. Анализирани и представени са методите и решенията, с които тази технология излезе от изследователските лаборатории и оказва незаличимо въздействие върху различни индустрии – от здравеопазването и маркетинга до развлечението и образованието.

От съществено значение е да бъдат взети предвид не само техническите аспекти на разпознаването на лицеви емоции, но и по-широките социални последици. Този доклад се стреми да предостави цялостно разбиране на текущото състояние в предметната област, като очертава неговите перспективи и опасности. В крайна сметка, това е изследване за това как

---

<sup>17</sup> Докладът е представен на научната сесия на 27.10.2023 в секция „Комуникационна и компютърна техника“ с оригинално заглавие на български език: РАЗПОЗНАВАНЕ НА ЛИЦЕВИ ЕМОЦИИ С ИЗПОЛЗВАНЕ НА ИЗКУСТВЕН ИНТЕЛЕКТ

машините се учат да разчитат емоциите и какво означава тази дълбока трансформация за настоящето и бъдещето на взаимодействието между хората и компютърните системи.

## ИЗЛОЖЕНИЕ

Изражението на лицето е основен метод и характерна черта на хората при предаването и определянето на настроението им. Съществуват много опити да се състави инструмент за автоматичен анализ на различните изражения на човешкото лице, като това има приложение в много области – роботика, медицина, системи за подпомагане на шофирането, детектори на лъжата и др. През двадесети век, Пол Екман дефинира седем основни емоции, които са независими от културата, етноса, възрастта или други различия на хората (Ekman, P., 1993). Тези емоции са гняв, страх, щастие, тъга, презрение, отвращение и изненада. В своите разработки, Пол Екман представя обширно проучване за идентификацията на лицевите емоции, включително характеристиките за набора от данни и класификаторите за изследване и разпознаване на лицевите емоции. Пример за тези техники за класифициране са показани на Фиг. 1 и те са полезни при по-нататъшното разглеждане на методите за разпознаване на емоции, защото този доклад изследва възможностите за прогнозиране на бъдещи реакции от изображения въз основа на разпознаването на емоции, използвайки различни класове класификатори. Класификационните алгоритми, които са разгледани и обсъдени, са базирани на опорните векторни машини и на невронните мрежи или по-точно на конволюционните невронни мрежи (Convolutional Neural Networks – CNN).



Фиг. 1. Седемте основни емоции според Пол Екман (Ekman, P., 1993)

При разпознаването на лицевите емоции може да възникнат множество проблеми, като преекспониране на емоциите, демонстриране на различни изражения, затруднено разпознаване на емоциите заради грим, лицеви травми, лицево окосмяване и т.н., които могат да се решат с помощта на изкуствен интелект и конволюционни мрежи.

Развитието на компютърното зрение и машинното обучение прави разпознаването на емоциите много по-точно и достъпно за широката публика. В резултат на това, алгоритмите и механизмите за разпознаване на лицеви изражения, които са част от технологиите за обработката на изображения, придобиват все по-голяма популярност и все по-бързо се разширяват и разпространяват. Някои от възможните приложения на тези решения са при взаимодействията между хората и компютърните системи, за психиатрични наблюдения, за разпознаване на шофьори, които са употребили алкохол или наркотици, както и може би най-важното – при детекторите на лъжата.

Основните методи за разпознаване на лицевите емоции (FER) са следните:

### *Разпознаване на емоции*

Разпознаването на лица е аспект на компютърните науки, който се занимава с методите и стратегиите за откриване на емоции в израженията на лицето. Очаква се, в близко бъдеще, това да бъде следващата стъпка в комуникацията с компютрите, като по-голямата част от изследванията в тази област се фокусират върху разпознаването на човешки емоции от аудио и видео източници (Kahou, S.E., et. al., 2015). Много от изследванията са фокусирани върху разпознаването и съпоставянето на лица, но без да се използват конволюционни невронни

мрежи за анализ и класификация на емоции в снимките. Разпознаването на емоциите е метод за изследване и идентифициране на емоциите, както и на стратегиите и процедурите, използвани за това. Емоциите могат да бъдат открити чрез изражения на лицето, вербални сигнали и други индикатори. Машинното обучение, невронните мрежи, изкуственият интелект и емоционалната интелигентност са само някои от методите, използвани за откриване и определяне на емоции (Sajjad, M., et. al., 2023). Разпознаването на емоциите е все по-силно изследвана област, което е от голямо значение за решаването на различни социални, икономически и научни предизвикателства.

#### *Разпознаване на лицеви емоции*

Разпознаването на лицеви емоции е изследователска област, която се опитва да идентифицира емоцията от изражението на човешкото лице. Проучванията сочат, че развитието в разпознаването на емоции прави сложните системи по-прости. Разпознаването на емоции е труден процес, тъй като емоциите могат да се различават в зависимост от средата, външния вид, културата и реакцията на лицето, което води до неясни данни.

#### *Дълбоко обучение*

Дълбокото обучение е техника за машинно обучение, която моделира данните, използвани при изпълнението на определена задача. Дълбокото обучение на невронни мрежи има приложение в областта на разпознаването на изображения, класификацията, вземането на решения, разпознаването на образи и др.

Един от факторите за успеха на дълбокото обучение е обучението на невронната мрежа с примери. Множество FER бази данни са на разположение на изследователите за изпълнението на тази задача, като те се различават по отношение на броя и размера на изображенията и видеоклиповете, вариациите на осветеността, позата на лицето и т.н. По-популярните бази данни за изследване на лицевите емоции са представени в Таблица 1.

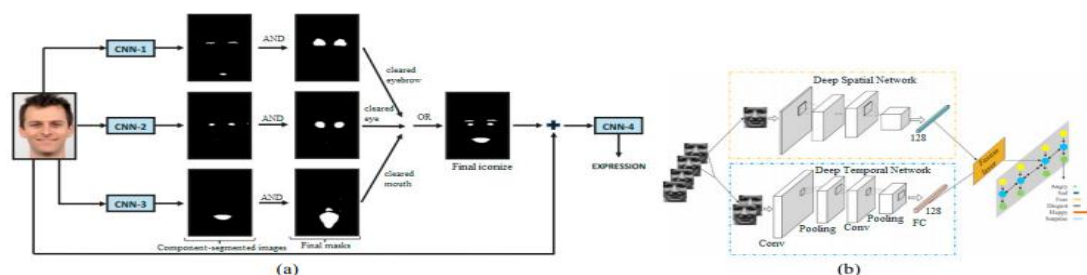
Таблица 1. Свободни за достъп FER бази данни

| Бази данни | Описание                                                                            | Емоции                                              |
|------------|-------------------------------------------------------------------------------------|-----------------------------------------------------|
| MultiPie   | Повече от 750 000 изображения, заснети от 15 изгледа и 19 условия на осветяване     | Гняв, Отвращение, Неутрален, Щастлив, Вик, Изненада |
| MMI        | 2900 видеоклипа, показват неутрален, начало, връх и отместване                      | Шест основни емоции и неутрални                     |
| GEMEP FERA | 289 последователности от снимки                                                     | Гняв, страх, тъга, облекчение, щастие               |
| SFEW       | 700 изображения с различна възраст, оклузия, осветеност и поза на главата           | Шест основни емоции и неутрални                     |
| CK+        | 593 видеоклипа за позирани и непозирани изражения                                   | Шест основни емоции и неутрални                     |
| FER2013    | 35 887 изображения в нива на сивото се събират от търсенето на изображения в Google | Шест основни емоции и неутрални                     |
| BU-3DFE    | 2500 3D изображения на лицето, заснети в два изгледа -45°, +45°                     | Шест основни емоции и неутрални                     |
| Oulu-CASIA | 2880 видеоклипа, заснети при три различни условия на осветеност                     | Радост, отвращение, изненада, регресия и други      |
| AffectNet  | Повече от 440 000 Интернет изображения                                              | Шест основни емоции и неутрални                     |
| RAFD-DB    | 30 000 изображения от реалния свят                                                  | Шест основни емоции и неутрални                     |

Въпреки забележителния успех на традиционните методи за лицево разпознаване чрез извличане на характеристики, през последното десетилетие изследванията са основно насочени към подхода за дълбоко обучение поради големия му капацитет за автоматично разпознаване. В този контекст в следващата част на тази публикация ще бъдат разгледани проучвания в областта на FER, показващи предложените методи за дълбоко обучение, чрез които се получава много по-точно разпознаване на емоциите.

Първото такова проучване предлага използването на конволюционни невронни мрежи за FER на изображения от няколко различни бази данни (Walecki, R. & Rudovic, O., 2017). След извличане на лицевите ориентир от данните, изображенията се намаляват до размери от 48 x 48 пиксела и се прилага техника за увеличаване на данните. Използваната архитектура се състои от два слоя за обединяване, след което се добавят два модула за начални стилове, които съдържат конволюционни слоеве с размери 1 x 1, 3 x 3 и 5 x 5. Те представят способността да се използва техниката „мрежа в мрежата“, позволяваща увеличаване на локалната производителност.

В друго изследване се изучава въздействието на предварителната обработка на данни преди обучението на мрежата, за да има по-добра класификация на емоциите (Ngo, Q. & Yoon, S., 2020). Увеличаване на данните, корекция на завъртането, изрязване, семплиране с 32 x 32 пиксела и нормализиране на интензитета са стъпките, които са приложени преди CNN, която се състои от два слоя за вземане на решение с напълно свързани 256 и 7 неврона. Изследователите показват, че комбинирането на всички тези стъпки по предварителна обработка е много по-ефективно от прилагането им поотделно. На Фиг. 2 е представен този метод за разпознаването на емоции (Ngo, Q. & Yoon, S., 2020).



Фиг. 2. Метод за разпознаване на лицеви емоции

През последните години се наблюдава задълбочен интерес към механизмите за FER чрез дълбоко обучение и невронни мрежи. Задачата на автоматизираното разпознаване на лицеви емоции преминава през няколко различни стъпки – обработка на данните, избор на архитектура на модела и процес по разпознаване на емоциите.

Предварителната обработка е важна стъпка, която включва няколко техники, като преоразмеряване и изрязване на изображенията за намаляване на времето за обучение на мрежата, нормализиране на пространствените и интензивни пиксели, усилване на данните за увеличаване на разнообразието на изображенията и премахването на различни проблеми (Ko, B.C., 2018). В изследванията, които са обобщени в Таблица 2, се вижда, че е постигната висока точност на FER чрез прилагане на CNN мрежи с пространствени и последователни данни. Използването на комбинация между CNN и RNN, особено при мрежата LSTM (Lindemann, B., et. al., 2021), ясно показва, че невронните мрежи са сред най-точните методи за разпознаване на лицеви емоции. Също така може да бъде отбелязано, че за да бъде тествана ефективността на невронните мрежови архитектури, трябва предварително да бъдат обучени и тествани модели на всяка една мрежа, като за целта може да се използват изображения от няколко бази данни. Както е видимо от Таблица 2, степента на разпознаване на емоциите варира при изследвания с различни бази данни, дори при един и същ модел за дълбоко обучение (Marrero-Fernández, P.D., 2020).

Таблица 2. Данни от проведени изследвания и резултати от комбинирането на различни бази данни за обучение на невронни мрежи за разпознаване на лицеви емоции

| База данни                                                            | Използвана архитектура | Успеваемост на разпознаване                   |
|-----------------------------------------------------------------------|------------------------|-----------------------------------------------|
| MultiPie (Gross, et. al., 2008), MMI, DISFA, FERA, SFEW, CK+, FER2013 | CNN                    | 94.7%, 77.9%, 55%, 76.7%, 47.7%, 93,2%, 61.1% |
| CK+, JAFFE, BU-3DFE                                                   | CNN                    | 96.76% for CK+                                |
| CK+ (Lucey et. Al, 2010)                                              | CNN                    | 97.01%                                        |
| JAFFE, CK+                                                            | SBN-CNN                | 95.24%, 96.87%                                |
| RAF-DB, AffectNet                                                     | ACNN                   | 80.54%, 54.84%                                |
| RafD                                                                  | CNN                    | 94.44%                                        |
| FER2013                                                               | CNN                    | 65%                                           |
| JAFFE, CK+                                                            | CNN                    | 95.23%, 93.24%                                |
| MMI, CASME II                                                         | CNN-LSTM               | 78.61%, 60.98%                                |
| CK+, Oulu-CASIA, MMI, BP4D                                            | STC-NLSTM              | 99.8%, 93.45%, 84.53%                         |
| CK+, Oulu-CASIA, MMI                                                  | DCBiLSTM               | 99.6%, 91.07%, 80.71%                         |

## ЗАКЛЮЧЕНИЕ

Този доклад представя обзор на тенденциите и изследванията от областта на лицевото разпознаване на емоции. Анализирани и представени са архитектури на CNN, предложени наскоро в различни изследвания, както и различни бази данни, съдържащи изображения, събрани от реалния свят или в лабораторна среда, и използвани за откриване и разпознаване на човешки емоции. FER е един от най-важните начини за предоставяне на информация за емоционалното състояние, но методът е органичен и позволява разпознаването само на неутралната и шестте основни емоции. Този механизъм не отговаря изцяло на това, което присъства в ежедневието ни, а именно голямо множество разнообразни други емоции. Това определя нуждата за съставяне на бази данни с изображения за всички човешки емоции и води до необходимостта да бъдат създадени мощни архитектури за дълбоко обучение, които да могат да разпознават всички основни и вторични емоции.

В настоящия момент от време, разпознаването на емоции разчита все по-малко на едномодалните методи за анализ и все повече на прилагането на мултимодални подходи. Тази мултимодалност е една от предпоставките за точното и ефективно разпознаване на човешките емоции. Разработването и създаването на мултимодални архитектури, като например, чрез сливане на аудио-визуални и физиологични модалности, както и на нови и подробни бази данни с изображения на лицеви емоции, ще позволи създаването на нови решения и методи за автоматизирано разпознаване на човешките емоции.

## БЛАГОДАРНОСТИ

Публикацията отразява резултати от работата по Национална научна програма „Интелигентно растениевъдство“ на Министерството на образованието и науката (МОН), одобрена с Решение на МС № 866/26.11.2020 г.

Публикацията е реализирана с подкрепата на проект № 2023 – ФЕЕА – 03 „Симулационно и експериментално изследване на методите и механизмите за предоставяне конфиденциалност и интегритет на данните в съвременните локални и безжични комуникационни мрежи“, финансиран от ФНИ на Русенски университет „Ангел Кънчев“.

## REFERENCES

- Ekman, P. (1993). *Facial expression and emotion*. American Psychologist, 48(4), pp. 384-392, <https://doi.org/10.1037/0003-066X.48.4.384>
- Gross, R., Matthews, I., Cohn, J., Kanade, T. & Baker, S. (2008). *Multi-PIE*, 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, Netherlands, pp. 1-8, DOI: 10.1109/AFGR.2008.4813399.
- Kahou, S.E., Michalski, V. & Konda, K. (2015). *Recurrent Neural Networks for Emotion Recognition in Video*. Proceedings of the ACM on International Conference on Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015, pp. 467-474.
- Ko, B.C. (2018). *A Brief Review of Facial Emotion Recognition Based on Visual Information*. Sensors (Basel). Vol. 18(2):401. DOI: 10.3390/s18020401.
- Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). *A Survey on Long Short-Term Memory Networks for Time Series Prediction*. Procedia CIRP, Volume 99, pp. 650-655, ISSN 2212-8271, <https://doi.org/10.1016/j.procir.2021.03.088>.
- Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z. & Matthews, I. (2010). *The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, 13-18 June 2010, pp. 94-101.
- Marrero-Fernández, P.D., Tsang, I., Tsang, I.J., Guerrero-Peña, F.A., & Cunha, A. (2020). *Deep Metric Structured Learning for Facial Expression Recognition*. ArXiv, abs/2001.06612.
- Ngo, Q. & Yoon, S. (2020). *Facial Expression Recognition Based on Weighted-Cluster Loss and Deep Transfer Learning Using a Highly Imbalanced Dataset*. Sensors vol. 20(9):2639. DOI: 10.3390/s20092639.
- Sajjad, M., Ullah, F.U.M., Ullah, M., Christodoulou, G., Cheikh, F.A., Hijji, M., Muhammad, K. & Rodrigues, J.J.P.C. (2023). *A Comprehensive Survey on Deep Facial Expression Recognition: Challenges, Applications, and Future Guidelines*, Alexandria Engineering Journal, Volume 68, pp. 817-840, ISSN 1110-0168, <https://doi.org/10.1016/j.aej.2023.01.017>.
- Walecki, R. & Rudovic, O. (2017). *Deep Structured Learning for Facial Expression Intensity Estimation*. Journal on Image and Vision Computing, vol. 259, pp. 143-154.