

**СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ/SYSTEM ANALYSIS,  
MANAGEMENT AND PROCESSING OF INFORMATION**DOI: <https://doi.org/10.60797/IRJ.2026.168.100> EDN: LDXJSN**ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ МЕТОДОВ КЛАССИФИКАЦИИ ЭМОЦИОНАЛЬНЫХ  
СОСТОЯНИЙ ПО ИЗОБРАЖЕНИЯМ ЛИЦ НА ОСНОВЕ РУЧНЫХ ПРИЗНАКОВ И СВЕРТОЧНЫХ  
НЕЙРОННЫХ СЕТЕЙ**

Научная статья

**Аликрицкий М.С.<sup>1,\*</sup>, Захарова О.И.<sup>2</sup>**<sup>1,2</sup> Поволжский государственный университет телекоммуникаций и информатики, Самара, Российская Федерация

\* Корреспондирующий автор (mixail195222[at]gmail.com)

Предложена: 06.04.2026; Принята: 02.06.2026; Опубликовано: 17.06.2026

**Аннотация**

В работе проведён анализ эффективности методов классификации эмоциональных состояний по изображениям лиц с использованием ручных признаков и сверточных нейронных сетей. Сопоставляются традиционные подходы, основанные на извлечении дескрипторов (LBP, HOG), а также современные методы глубокого обучения. Цель работы — выполнить сравнительное исследование качества классификации при применении различных методов и определение наиболее эффективного подхода для решения задачи распознавания эмоций. В качестве экспериментальной базы использован набор изображений лиц с разметкой эмоциональных состояний. Оценка результатов проводилась с использованием метрик качества классификации. Установлено, что модели на основе сверточных нейронных сетей обеспечивают более высокую точность по сравнению с методами, основанными на ручных признаках, однако требуют больших вычислительных ресурсов. Обоснована целесообразность применения гибридных или глубоких моделей для повышения эффективности систем распознавания эмоций.

**Ключевые слова:** распознавание эмоций, анализ изображений лиц, классификация изображений, сверточные нейронные сети, ручные признаки, LBP, HOG, компьютерное зрение.

**A STUDY OF THE EFFECTIVENESS OF METHODS FOR CLASSIFYING EMOTIONAL STATES FROM FACIAL  
IMAGES BASED ON MANUAL FEATURES AND CONVOLUTIONAL NEURAL NETWORKS**

Research article

**Alikrickiy M.S.<sup>1,\*</sup>, Zakharova O.I.<sup>2</sup>**<sup>1,2</sup> Povolzhskiy State University of Telecommunications and Informatics, Samara, Russian Federation

\* Corresponding author (mixail195222[at]gmail.com)

Suggested: 06.04.2026; Accepted: 02.06.2026; Published: 17.06.2026

**Abstract**

The work analyses the effectiveness of methods for classifying emotional states from facial images using manual features and convolutional neural networks. Traditional approaches based on descriptor extraction (LBP, HOG) are compared with modern deep learning methods. The aim of the study is to conduct a comparative analysis of classification quality when applying different methods and to identify the most effective approach for solving the problem of emotion recognition. A dataset of facial images annotated with emotional states was used as the experimental basis. The results were evaluated using classification quality metrics. It was found that models based on convolutional neural networks provide higher accuracy compared to methods based on manual features, but require greater computational resources. The feasibility of using hybrid or deep models to improve the performance of emotion recognition systems is demonstrated.

**Keywords:** emotion recognition, facial image analysis, image classification, convolutional neural networks, manual features, LBP, HOG, computer vision.

**Введение**

Задача автоматического распознавания эмоций по изображениям лиц остается актуальной для систем человеко-машинного взаимодействия, дистанционного обучения и медицинской диагностики [1], [2], [3]. Потребность в таких системах возрастает с развитием интерфейсов, адаптивных образовательных платформ и инструментов психологической поддержки, где анализ мимики позволяет получать обратную связь о состоянии пользователя в реальном времени. Основные трудности связаны с вариативностью мимических проявлений, условиями освещения и шумами, а также дисбалансом классов в реальных данных [4], [5].

В настоящее время применяются два основных подхода: ручное извлечение признаков (LBP, HOG) с последующей классификацией и глубокое обучение, автоматически формирующее иерархические признаки. Классические методы отличаются низкими вычислительными затратами, но ограничены в обобщении [7], [8]. Ряд современных работ указывает на преимущество CNN [9], [10], [11].

*Цель работы* — сравнительное экспериментальное исследование эффективности LBP+SVM, HOG+SVM и CNN на наборе FER2013 с учетом дисбаланса классов. Научная новизна заключается в сравнительном анализе

устойчивости классических и нейросетевых методов к дисбалансу классов в условиях ограниченного разрешения изображений.

### Материалы и методы

Экспериментальной базой послужил открытый набор данных FER2013 в стандартном разбиении: 28709 изображений для обучения, 3589 — для валидации, 3589 — для тестирования, все изображения в градациях серого размером 48x48 пикселей. Изображения отнесены к семи классам эмоциональных состояний: angry, disgust, fear, happy, neutral, sad, surprise. Распределение классов неравномерно, наиболее представлен класс «happy» (895 примеров в тестовой выборке), наименее — «disgust» (56 примеров). Подобная неравномерность распределения типична для реальных выборок и может приводить к смещению модели в сторону многочисленных классов [6], [11]. Перед обучением все изображения были нормализованы в диапазон от 0 до 1. Дополнительная предобработка, такая, как выравнивание лиц по глазам или удаление фона не применялась, чтобы сохранить соответствие реальным условиям работы системы, где такие операции не всегда возможны.

В исследовании реализованы три подхода к классификации. Первый подход использует извлечение локальных бинарных шаблонов (LBP), для каждого изображения строилась гистограмма LBP-признаков, которая затем подавалась на вход классификатора опорных векторов (SVM) с линейным ядром. Линейное ядро было выбрано из соображений интерпретируемости и меньшей склонности к переобучению на признаках небольшой размерности. Второй метод использует признаки ориентированных градиентов (HOG): вычислялся вектор HOG-дескриптора, после чего выполнялась классификация линейным SVM. Оба подхода относятся к классическим методам компьютерного зрения и отличаются вычислительной эффективностью, но имеют ограничения при анализе сложных мимических выражений. Аугментация для классических методов не применялась, так как они работают на фиксированных дескрипторах.

Третий подход представлен сверточной нейронной сетью (CNN), обучаемая непосредственно на исходных изображениях. Архитектура используемой сверточной нейронной сети представлена на рисунке 1. Сеть принимает на вход изображения размером 48x48x1 и состоит из трех последовательных сверточных блоков. Первый блок включает два сверточных слоя с 32 фильтрами размером 3x3 (padding="same", шаг свертки 1), каждый из которых сопровождается функцией активации ReLU, после первого сверточного слоя идет пакетная нормализация, завершающий этап блока — применение слоя подвыборки MaxPooling (2x2) и Dropout с вероятностью 0,25. Второй блок имеет аналогичную структуру, но с увеличением числа фильтров до 64 и Dropout 0,25. Третий блок содержит 128 фильтров и Dropout 0,3. После сверточных блоков используется слой глобального усредняющего объединения (GlobalAveragePooling2D), далее полносвязный слой из 256 нейронов с функцией активации ReLU и Dropout 0,5. Выходной слой реализован с использованием функции softmax и содержит 7 нейронов, соответствующих числу классов. Выбор архитектурных параметров обусловлен особенностями решаемой задачи.

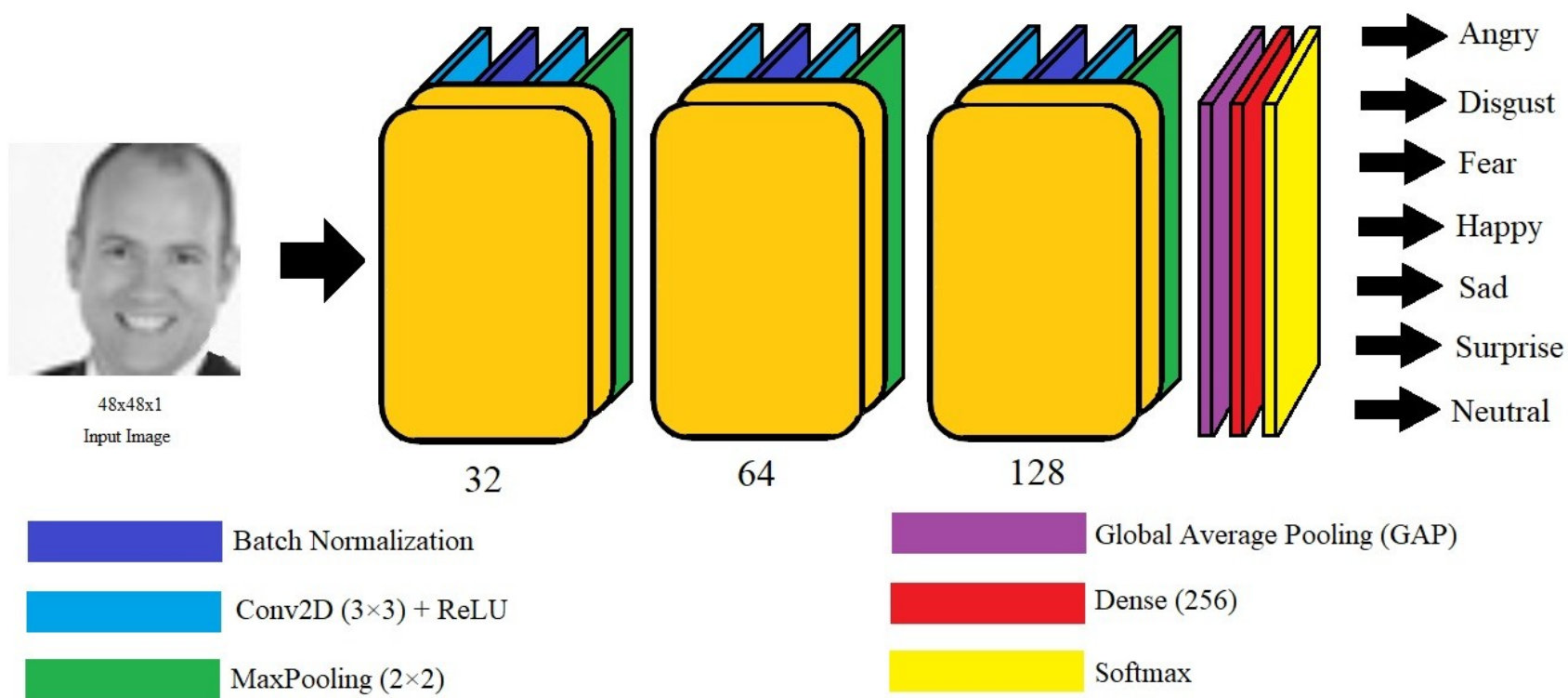


Рисунок 1 - Архитектура сверточной нейронной сети  
DOI: <https://doi.org/10.60797/IRJ.2026.168.100.1>

Обучение проводилось с оптимизатором Adam (начальная скорость обучения 0,0005) и функцией потерь categorical crossentropy в течение 60 эпох. Для контроля переобучения применялась валидационная выборка, на которой отслеживалась динамика потерь. Для расширения обучающей выборки применялась аугментация, такие как случайные повороты, сдвиги, масштабирование и горизонтальные отражения. Для компенсации дисбаланса классов использовались весовые коэффициенты, обратно пропорциональные частоте встречаемости каждого класса. Качество моделей оценивалось по метрикам accuracy, precision, recall и F1-score на независимой тестовой выборке.

### Результаты и обсуждение

LBP+SVM показал точность 0,25. Наибольшее число ошибок наблюдается между эмоциями fear, sad и neutral из-за сходства текстур. Применение HOG+SVM позволило увеличить точность до 0,44 за счет учета геометрии лица, но сохранил чувствительность к дисбалансу и ошибки между визуально схожими классами. Модель CNN была обучена дважды на одной архитектуре, без использования весовых коэффициентов точность составила 0,5453, применение весовых классов увеличило точность до 0,6077. Лучшие показатели — для happy (precision=0,83) и surprise (0,72), самые низкие — для disgust (recall=0,52) из-за малого числа примеров. Fear, sad и neutral частично смешиваются из-за перекрытия признаков. Это указывает на то, что часть ошибок обусловлена не только ограничениями модели, но и объективной близостью визуальных проявлений данных эмоциональных состояний. Применение весов классов повысило recall для редких классов. Результаты сравнительного анализа методов распознавания эмоций представлены в таблице 1.

Таблица 1 - Сравнение точности классификации различных методов

DOI: <https://doi.org/10.60797/IRJ.2026.168.100.2>

Метод	Признаки	Accuracy, доли	Precision (macro), доли	Recall (macro), доли	F1-score (macro), доли
SVM	LBP	0,25	0,1418	0,1452	0,0647
SVM	HOG	0,44	0,3923	0,3888	0,3830
CNN	Автоматически извлекаемые	0,6077	0,5566	0,6152	0,5677

Примечание: значения метрик приведены в долях от 1

Результаты для метода LBP+SVM характеризуются крайне низкими значениями F1-score (0,0647), что связано с выраженным смещением модели в сторону наиболее представленного класса (happy). Это указывает на высокую чувствительность метода к дисбалансу классов и ограниченную способность LBP-признаков описывать сложную мимику. В противовес, HOG+SVM демонстрирует более сбалансированные значения метрик (F1-score = 0,3830), что объясняется учетом пространственной структуры изображения. Прирост accuracy от HOG к CNN составляет более 16%, что свидетельствует о более высокой эффективности автоматического извлечения признаков. Полученные значения соответствуют диапазону 55–65%, указанным в обзоре Li и Deng [6]. Классические методы LBP и HOG, несмотря на свою интерпретируемость, уступают нейросетевым из-за неспособности моделировать сложные нелинейные зависимости в мимике.

Проблема дисбаланса классов сохраняет существенное влияние: recall для класса disgust почти на 30% ниже, чем для happy. Применение весовых коэффициентов смягчило, но не устранило полностью этот эффект. Дополнительный анализ показал, что даже после балансировки модель часто путает disgust с fear и angry, что может быть связано с анатомическим сходством отдельных мимических паттернов. Ограничение достигнутой точности CNN может быть связано с низким пространственным разрешением изображений (48×48 пикселей), что приводит к потере мелких мимических деталей, критичных для различения сходных эмоциональных состояний. Дополнительным ограничением является использование датасета FER2013, содержащего изображения низкого качества и шумные метки, что может ограничивать верхнюю границу достигаемой точности моделей. С практической точки зрения, даже 60% точности может быть достаточно для некоторых приложений, но для индивидуальной диагностики требуется дальнейшее повышение надёжности. Перспективными направлениями являются генеративные методы дообучения (синтез редких классов) и использование метрического обучения для улучшения разделения визуально сходных эмоций (fear, sad, neutral). Также многообещающим выглядит применение трансформеров для анализа мимики и интеграция с видео-последовательностями, где доступна временная динамика выражений.

### Заключение

В рамках работы выполнено экспериментальное исследование трёх подходов к классификации эмоций по лицам. Классические методы (LBP, HOG) обеспечивают базовый уровень точности (0,25-0,44), но их эффективность ограничена при сложной мимике и дисбалансе данных. Сверточная нейронная сеть с аугментацией и весами классов достигла accuracy 0,6077, что указывает на целесообразность применения глубокого обучения для данной задачи. Результаты сопоставимы с современными работами [6], [9], [10], [11]. Перспективными направлениями являются разработка более сложных стратегий балансировки выборки и архитектурных модификаций CNN.



### Конфликт интересов

Не указан.

#### Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

### Conflict of Interest

None declared.

#### Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

### Список литературы на английском языке / References in English

1. Ko B.C. A brief review of facial emotion recognition based on visual information / B.C. Ko // *Sensors*. — 2018. — Vol. 18, No. 2. — DOI: 10.3390/s18020401
2. Shan C. Facial expression recognition based on local binary patterns: A comprehensive study / C. Shan, S. Gong, P.W. McOwan // *Image and Vision Computing*. — 2009. — Vol. 27, No. 6. — P. 803–816. — DOI: 10.1016/j.imavis.2008.08.005.
3. Murthy M.K.L. Efficient Feature Extraction for Recognition of Human Emotions through Facial Expressions Using Image Processing Algorithms / M.K.L. Murthy, M. Divyanjali, S.M. Basha [et al.] // *E3S Web of Conferences*. — 2023. — Vol. 391. — Article ID 01182. — DOI: 10.1051/e3sconf/202339101182.
4. Khaireddin Y. Facial Emotion Recognition: State of the Art Performance on FER2013 / Y. Khaireddin, Z. Chen // arXiv preprint arXiv:2105.03588. — 2021. — URL: <https://arxiv.org/abs/2105.03588> (accessed: 06.04.2026).
5. Wang X. Optimizing Class Imbalance in Facial Expression Recognition Using Dynamic Intra-Class Clustering / X. Wang, B. Li, J. Yu [et al.] // *Biomimetics*. — 2025. — Vol. 10, No. 5. — P. 296–311. — DOI: 10.3390/biomimetics10050296.
6. Li S. Deep Facial Expression Recognition: A Survey / S. Li, W. Deng // *IEEE Transactions on Affective Computing*. — 2022. — Vol. 13, No. 3. — P. 1195–1215. — DOI: 10.1109/TAFFC.2020.2981446.
7. Goodfellow I.J. Challenges in Representation Learning: A Report on Three Machine Learning Contests / I.J. Goodfellow, D. Erhan, P.L. Carrier [et al.] // *Neural Information Processing*. — Berlin, Heidelberg: Springer, 2013. — P. 117–124. — DOI: 10.1007/978-3-642-42051-1\_16.
8. Dalal N. Histograms of oriented gradients for human detection / N. Dalal, B. Triggs // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). — San Diego, CA, USA, 2005. — P. 886–893. — DOI: 10.1109/CVPR.2005.177.
9. Telceken M. Can artificial intelligence understand our emotions? Deep learning applications with face recognition / M. Telceken, D. Akgun, S. Kacar [et al.] // *Current Psychology*. — 2025. — Vol. 44. — P. 7946–7956. — DOI: 10.1007/s12144-025-07375-0.
10. Chinnasamy P. Improving facial emotion recognition through dataset merging and balanced training strategies / P. Chinnasamy, W. Jeberson, K. Singh // *Computers and Electrical Engineering*. — 2025. — Vol. 123. — Article 110208. — DOI: 10.1016/j.compeleceng.2025.110208.
11. Safarov F. Enhanced AlexNet with Gabor and Local Binary Pattern Features for Improved Facial Emotion Recognition / F. Safarov [et al.] // *Sensors*. — 2025. — Vol. 25, No. 12. — P. 3832–3847. — DOI: 10.3390/s25123832.