

УДК 004.932+004.056.53

DOI [https://doi.org/10.24144/2616-7700.2025.47\(2\).312-319](https://doi.org/10.24144/2616-7700.2025.47(2).312-319)**С. В. Шкіря¹, Ю. В. Андрашко²**

¹ ДВНЗ «Ужгородський національний університет»,
аспірант кафедри системного аналізу та теорії оптимізації
serhii.shkiria@uzhnu.edu.ua
ORCID: <https://orcid.org/0009-0006-7129-2823>

² ДВНЗ «Ужгородський національний університет»,
доцент кафедри системного аналізу та теорії оптимізації,
кандидат технічних наук, доцент
yurii.andrashko@uzhnu.edu.ua
ORCID: <https://orcid.org/0000-0003-2306-8377>

ПРОЄКТУВАННЯ ТА ОПТИМІЗАЦІЯ ГЛИБОКОЇ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ НА ПРИКЛАДІ ДОСЛІДЖЕННЯ ДАТАСЕТУ CIFAR-10

У даній статті розглянуто розробку та дослідження згорткової нейронної мережі (CNN), яка призначена для класифікації зображень на наборі даних CIFAR-10. Було створено архітектуру, що складається з трьох згорткових блоків, використовуючи пакетну нормалізацію та проріджування, щоб підвищити ефективність і уникнути перенавчання. Модель навчалася з використанням технік аугментації даних та оптимізатора Nadam. Під час експериментів було досягнуто точності класифікації 91.28% на тестовому наборі даних. Результати порівнюються з іншими відомими підходами, що демонструє високу ефективність запропонованого методу.

Ключові слова: розпізнавання зображень, згорткова нейронна мережа, CIFAR-10, глибоке навчання, аугментація даних, пакетна нормалізація.

1. Вступ. Класифікація зображень є однією з ключових задач комп'ютерного зору та штучного інтелекту. Вона полягає у визначенні належності вхідного зображення до одного з наперед визначених класів. Незважаючи на очевидну простоту постановки, ця задача має значний вплив на широкий спектр прикладних систем: від автономних транспортних засобів і медичної діагностики до промислового контролю якості та моніторингу безпеки.

У багатьох інформаційних системах класифікація може виконувати допоміжну роль у ширшій задачі автентифікації цифрових об'єктів. Зокрема, якщо необхідно перевірити, чи відповідають дані, отримані від користувача, тим, що збережені у базі, класифікація дозволяє визначити, чи належить об'єкт до очікуваної категорії. Такий попередній етап зменшує невизначеність і забезпечує додатковий рівень контролю перед застосуванням складніших процедур верифікації або порівняння ознак.

Глибокі згорткові нейронні мережі (CNN) стали одним із найефективніших інструментів для розв'язання задач класифікації зображень завдяки здатності автоматично формувати ієрархічні ознаки та виявляти характерні патерни об'єктів. Подальше підвищення точності та стабільності моделей пов'язане з оптимальним проектуванням їхньої архітектури, вибором методів регуляризації та налаштуванням процесу навчання.

Метою даної роботи є проектування та оптимізація глибокої згорткової нейронної мережі для задачі класифікації зображень. Для тестування застосовано

набір CIFAR-10 як репрезентативний приклад, однак акцент робиться не на специфіці цього набору, а на принципах побудови ефективної моделі, що може бути адаптована й до інших задач класифікації та автентифікації цифрових об'єктів

2. Аналіз останніх досліджень і публікацій. Задача класифікації на наборі даних CIFAR-10 активно досліджується вже багато років. Класичні архітектури, такі як AlexNet і VGG, стали основою для багатьох сучасних рішень, але з часом з'явилися й більш ефективні підходи.

AlexNet, одна з перших глибоких згорткових мереж, досягла значного успіху на конкурсі ImageNet. Коли її адаптували для CIFAR-10, спрощені версії показали точність у межах 80-87%. Це зазвичай вважається базовим результатом, з яким порівнюють більш складні моделі [1].

Архітектура VGGNet, представлена у 2014 році, стала значним кроком уперед завдяки використанню дуже малих згорткових фільтрів (3x3), що дозволило досягти високої продуктивності завдяки збільшенню глибини мережі. Вона стала основою для багатьох подальших досліджень. Наприклад, згідно з роботою [2], неоптимізована модель VGG-16 показує точність 90.73% на CIFAR-10.

Підхід Network in Network (NIN) став справжнім проривом, замінивши традиційні багатопарові перцептрони та глобальне усереднення, щоб зменшити кількість параметрів і боротися з перенавчанням. Ці мережі демонструють вражаючу точність на CIFAR-10, досягаючи 89-91% [3].

Револьюційна архітектура ResNet (Residual Networks) представила концепцію residual blocks, що дозволяє навчати надзвичайно глибокі мережі. Навіть прості версії ResNet, такі як ResNet-20 і ResNet-32, досягають точності 90-92%, що є вражаючим результатом, хоча й вимагає спеціалізованої архітектури блоків [4].

У дослідженні [5] автори вивчали, як техніки регуляризації, такі як Dropout і Batch Normalization, впливають на якість навчання. Вони виявили, що ці методи можуть взаємодіяти не найкращим чином, викликаючи зміщення дисперсії, що негативно позначається на стабільності та кінцевій точності моделі. Автори підкреслили, що правильне використання цих технік є критично важливим для запобігання перенавчанню та покращення здатності моделі до узагальнення, а також надали рекомендації щодо їх оптимального поєднання.

Не менш важливою є аугментація даних – штучне збільшення тренувального набору шляхом застосування випадкових трансформацій до зображень, таких як повороти та зсуви. Як показано в [6], це дозволяє моделі краще узагальнювати та бути більш стійкою до змін у вхідних даних.

Сучасні підходи, такі як DenseNet [7], де кожен шар отримує на вхід карти ознак з усіх попередніх шарів, та EfficientNet [8], що пропонує методи масштабування глибини, ширини та роздільної здатності мережі, демонструють вражаючі результати, часто перевищуючи 95%. Проте ці архітектури є значно складнішими і вимагають більше обчислювальних ресурсів.

Аналіз цих досліджень показує, що ключовими факторами для досягнення високої точності є глибина та ширина мережі, ефективні методи регуляризації (такі як Dropout та пакетна нормалізація) та використання архітектурних інновацій, таких як залишкові з'єднання. Метод, запропонований у цій статті, базується на класичній архітектурі глибокої CNN, але з ретельно підібраними параметрами для досягнення конкурентоспроможного результату.

3. Розробка архітектури моделі глибокої згорткової нейронної мережі для задачі класифікації зображень. Для розв'язання задачі класифікації зображень було розроблено архітектуру глибокої згорткової нейронної мережі. Модель побудована таким чином, щоб поетапно формувати інформативні ознаки з вхідних зображень, переходячи від простих локальних структур (контури, градієнти, текстури) до складніших семантичних патернів. Така здатність до ієрархічного навчання представлень є ключовою перевагою згорткових мереж у порівнянні з класичними алгоритмами машинного навчання.

Архітектура моделі включає три послідовні згорткові блоки, кожен з яких складається з двох згорткових шарів із нелінійною активацією. Після кожного шару згортки виконується пакетна нормалізація, що стабілізує статистику активацій та прискорює збіжність. Для контролю складності моделі та запобігання перенавчанню використовується операція прорідження (dropout), яка випадковим чином вимикає частину нейронів під час навчання. Підвибірка (max pooling), розміщена наприкінці кожного блоку, зменшує просторову розмірність тензора ознак, скорочує кількість параметрів і концентрує інформацію на найбільш виразних компонентах.

Після проходження згорткових блоків тензор ознак перетворюється у вектор фіксованої довжини, який подається до повнозв'язного шару. Завершальний шар моделі застосовує функцію softmax, що формує ймовірнісний розподіл належності зображення до одного з класів.

Реалізація моделі здійснена з використанням бібліотеки TensorFlow та її високорівневого API Keras, що забезпечує декларативний опис архітектури та ефективну організацію процесу навчання.

У процесі експериментального добору гіперпараметрів було проведено серію навчань із різними комбінаціями глибини мережі, кількості фільтрів та коефіцієнтами регуляризації. Порівняння моделей показало, що найкращих результатів вдалося досягти при використанні багаторівневої архітектури з трьома згортковими блоками, кожен з яких містить по два шари згортки з подальшою нормалізацією, підвибіркою та прорідженням. Така конфігурація забезпечує збалансоване співвідношення між виразністю ознак та здатністю до узагальнення, що є критичним для класифікації зображень на обмежених за розміром наборах даних. Нижче наведено структурні параметри моделі, що були відібрані як оптимальні за результатами експериментів.

Архітектура нейронної мережі:

- **Вхід:** зображення розміром $32 \times 32 \times 3$.
- **Блок 1:** два послідовні шари Conv2D(64, 3×3 , ReLU, padding=same), після кожного згорткового шару застосовується BatchNormalization; далі — MaxPooling2D(2×2) та Dropout(0.3).
- **Блок 2:** два шари Conv2D(128, 3×3 , ReLU, padding=same) з BatchNormalization після кожного; потім — MaxPooling2D(2×2) та Dropout(0.4).
- **Блок 3:** два шари Conv2D(256, 3×3 , ReLU, padding=same) з BatchNormalization після кожного; далі — MaxPooling2D(2×2) і Dropout(0.4).
- **Класифікація:** Flatten $\rightarrow$ Dense(1024, ReLU) $\rightarrow$ BatchNormalization $\rightarrow$ Dropout(0.5) $\rightarrow$ Dense(10, softmax).

Загальна кількість параметрів моделі становить 5 358 666, з яких 5 354 826 є навчаючими, тобто їх значення змінюються в процесі градієнтної оптимізації під час навчання мережі. Решта параметрів належить до категорії ненавчаючих і використовується службовими компонентами моделі, зокрема механізмами пакетної нормалізації, які зберігають статистики активацій для стабільної роботи мережі під час інференсу.

4. Опис експерименту. Процес навчання моделі здійснювався у кілька етапів, що включали попередню підготовку даних, налаштування гіперпараметрів та контроль збіжності під час оптимізації. Для забезпечення стабільного навчання було обрано пакетний розмір 64, що є компромісом між точністю оцінки градієнта та ефективністю використання обчислювальних ресурсів. Мережу навчали із використанням оптимізатора *Nadam*, який поєднує адаптивне коригування швидкості навчання (*Adam*) з методом Нестерова, що прискорює збіжність. Початкова швидкість навчання становила 0.001, функцією втрат виступала *sparse_categorical_crossentropy*, а основною метрикою оцінювання – *accuracy*.

Для підвищення здатності моделі до узагальнення та зменшення ризику перенавчання застосовувалася аугментація даних. На рівні вхідного зображення вносились контрольовані випадкові трансформації: повороти до 15°, горизонтальні відображення, масштабування до 10% та зсуви по ширині/висоті до 10%. Це дозволило штучно збільшити розмаїття даних і зробити модель менш чутливою до варіативності вхідних прикладів.

Додатково застосовувалися механізми адаптивного контролю навчання. Метод *ReduceLROnPlateau* автоматично зменшував швидкість навчання при відсутності покращення на валідаційній вибірці, а *EarlyStopping* завершував навчання, коли протягом визначеної кількості епох не спостерігалось зменшення функції втрат, із відновленням найкращих ваг мережі (*best model restore*).

Набір даних CIFAR-10 використано як тестовий приклад для демонстрації працездатності розробленої архітектури. Він містить 60 000 кольорових зображень розміром 32×32 у 10 класах, з яких 50 000 використано для навчання та 10 000 – для тестування. Кожне зображення представлено трьома кольоровими каналами RGB, де інтенсивність кожного каналу є цілим числом у діапазоні від 0 до 255 (8-бітне кодування на канал). Перед подаванням зображень до моделі значення каналів було нормалізовано до інтервалу [0, 1] шляхом ділення на 255, що зменшує варіацію вхідних даних та сприяє швидшому і стабільнішому процесу навчання.

Хід навчання та результати. Навчання моделі тривало 200 епох з розміром батчу 64 з валідацією на тестовому наборі після кожної епохи. Завдяки механізму ранньої зупинки (*EarlyStopping*), навчання автоматично завершилося на 68-й епосі, оскільки показник втрат на валідаційній вибірці перестав покращуватися. Модель відновила ваги з 65-ї епохи, яка показала найкращий результат.

На графіку точності (рис. 1) видно, що точність на навчальній вибірці стабільно зростає, досягаючи ~ 93%. Точність на валідаційній вибірці також підвищується, досягаючи максимуму близько 50–60 епохи, після чого стабілізується на рівні ~ 91%.

Графік втрат (рис. 2) показує, що втрати на навчальній вибірці постійно

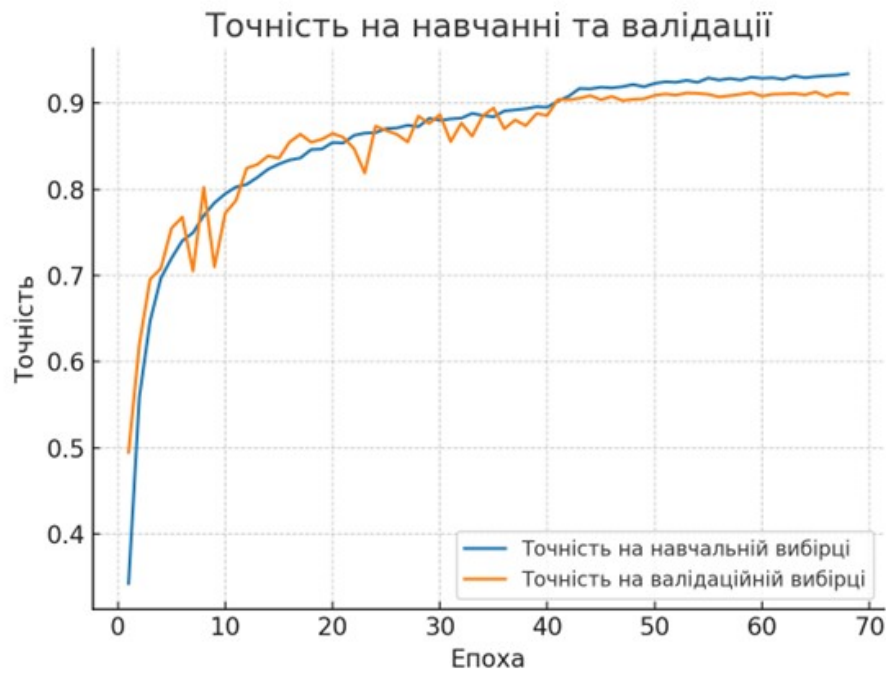


Рис. 1. Графік точності на навчальній та валідаційній вибірках.

зменшуються. Валідаційні втрати досягають мінімуму приблизно на 65-й епосі, після чого починають коливатися, що і стало причиною для ранньої зупинки.

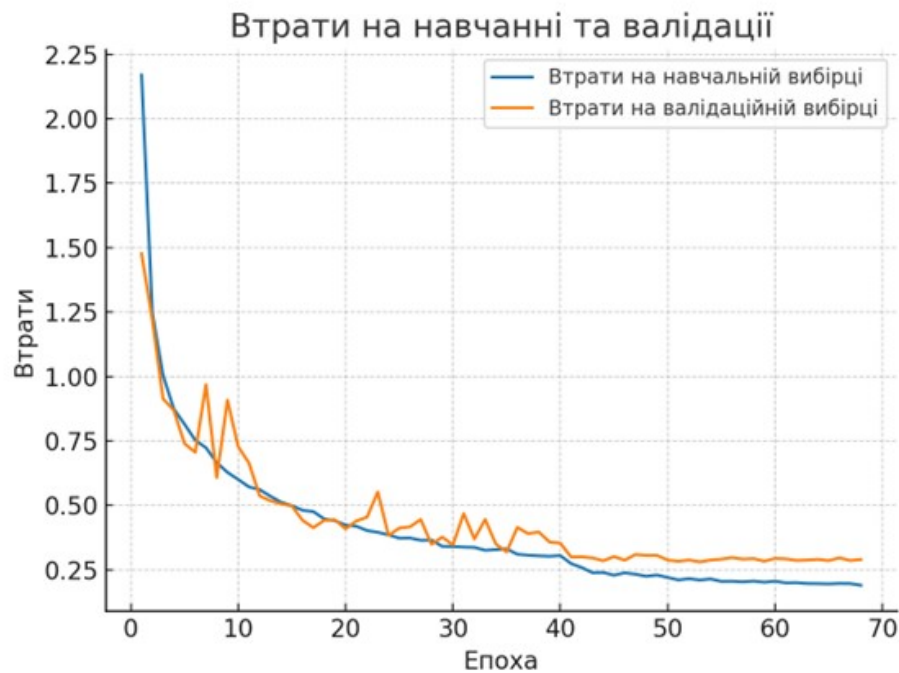


Рис. 2. Графік втрат на навчальній та валідаційній вибірках.

Після завершення навчання модель була оцінена на тестовому наборі даних, який не використовувався у процесі тренування та підбору гіперпараметрів. Це дозволяє оцінити здатність моделі до узагальнення та перевірити, наскільки добре вона класифікує раніше невидані зображення. Для оцінювання використовувалися дві основні метрики: функція втрат (Loss), яка характеризує відхилення прогнозів від істинних міток класів, та точність класифікації (Accuracy), що відображає відсоток правильно класифікованих зображень. Підсумкові результати наведено в табл. 1.

Таблиця 1.

Результати оцінювання моделі на тестовому наборі CIFAR-10

Метрика	Значення
Втрати	≈ 0.2858
Точність, %	≈ 91.28

Після отримання результатів оцінки модель була порівняна з низкою відомих архітектур згорткових нейронних мереж, які традиційно використовуються для класифікації зображень на наборі даних CIFAR-10. У порівняння включено як класичні моделі (такі як AlexNet та VGG), так і вдосконалені архітектури, що застосовують більш глибокі або інноваційні блоки (NIN, ResNet). Наведені значення точності відповідають результатам, опублікованим у відповідних роботах або репрезентативних реалізаціях моделей. Порівняння демонструє, що запропонована архітектура досягає точності на рівні сучасних моделей, перевищуючи VGG-подібні підходи та показуючи результат, співставний із ResNet-20 (табл. 2). Порівняльні результати з класичними і сучасними архітектурами наведено у табл. 2.

Таблиця 2.

Порівняння точності різних моделей на CIFAR-10

Модель / Метод	Точність (%)
AlexNet-подібна архітектура	~ 85.54
VGG-подібна архітектура	~ 90.73
Network in Network (NIN)	~ 91.19
ResNet-20	~ 91.25
Розроблена модель	~ 91.28

Отримані результати підтверджують ефективність запропонованої архітектури згорткової нейронної мережі для задачі класифікації зображень на прикладі CIFAR-10. Модель демонструє точність на рівні сучасних підходів, зберігаючи при цьому відносно просту структуру без використання глибоких резидуальних блоків або спеціалізованих механізмів оптимізації. Порівняння з іншими моделями свідчить, що досягнута точність перевищує результати класичних архітектур (AlexNet, VGG) та є співставною з точністю ResNet-20 і Network-in-Network, які належать до більш складних моделей. Це вказує на те, що ретельно підібране поєднання згорткових блоків, нормалізації та регуляризації дозволяє

забезпечити високий рівень узагальнення навіть для відносно компактних мереж.

Окремим напрямом застосування розробленої моделі є використання класифікації зображень як допоміжного компонента у задачах автентифікації користувачів. У попередніх дослідженнях авторів було показано, що автентифікація може базуватися на аналізі порівняння нових даних користувача з уже збереженими еталонними представленнями [9]. У цьому контексті класифікаційна модель може виконувати функцію первинної фільтрації або виявлення належності вхідних даних до певної категорії або користувача перед подальшим аналізом спеціалізованими методами. Таким чином, згорткова нейронна мережа може бути використана як підзадача більш складної системи автентифікації, сприяючи підвищенню точності та зменшенню кількості хибних спрацювань на етапі попередньої перевірки.

5. Висновки. У роботі представлено підхід до побудови та оптимізації згорткової нейронної мережі для задачі класифікації зображень на прикладі набору даних CIFAR-10. Запропонована архітектура поєднує три послідовні згорткові блоки з пакетною нормалізацією та регуляризацією Dropout, що забезпечує баланс між глибиною мережі та здатністю моделі до узагальнення. Використання аугментації даних, механізмів **EarlyStopping** та **ReduceLROnPlateau** дозволило уникнути перенавчання та оптимізувати процес тренування.

Модель досягла точності 91.28% на тестовому наборі, що є співставним з точністю сучасних архітектур, таких як Network-in-Network та ResNet-20, та перевищує або не поступається класичним моделям рівня AlexNet та VGG. Це демонструє, що ретельне проектування архітектури та правильний підбір гіперпараметрів можуть забезпечити високі результати без необхідності застосовувати надмірно складні глибокі мережі.

Практична значимість отриманих результатів полягає у можливості інтеграції моделі в прикладні системи, де класифікація зображень використовується як допоміжний етап більш складних процесів обробки даних. Зокрема, модель може виконувати функцію попереднього відсіву нерелевантних даних або перевірки належності вхідних об'єктів до еталонних категорій у системах цифрової автентифікації, що узгоджується з підходами, описаними у роботі [9]. Таким чином, запропонована архітектура може виступати ефективною складовою ширших рішень з підтримкою інтелектуальних методів перевірки автентичності.

Подальший розвиток дослідження може передбачати розширення набору даних, використання методів перенесення навчання або адаптацію архітектури для задач з обмеженими обчислювальними ресурсами. Отримані результати підтверджують, що оптимізація структури мережі та гіперпараметрів дозволяє забезпечити високу точність класифікації при помірній складності моделі, що є важливим для її практичного застосування.

Дослідження здійснено в рамках кафедральної науково-дослідної роботи «Моделі і методи системного аналізу в міждисциплінарних дослідженнях» (державний обліковий номер 0125U003246)

References

1. Shin, S., Dong Ho, S., & Kim, J. (2024). Implementation and comparison of CNN models on CIFAR-10 dataset. In *Proceedings of the 1st World Congress on Industrial Engineering and Operations Management* (pp. 496–503). IEOM Society International.

- <https://doi.org/10.46254/WC01.20240168>
2. Guo, K., Lin, Z., Chen, C., Xing, X., Liu, F., & Xu, X. (2024). Compact model training by low-rank projection with energy transfer. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4), 6708-6722. <https://doi.org/10.1109/TNNLS.2024.3400928>
 3. Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv*. arXiv:1312.4400
 4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778). <https://doi.org/10.1109/CVPR.2016.90>
 5. Li, X., Chen, S., & Yang, J. (2019). Understanding the disharmony between dropout and batch normalization by variance shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 268-277). <https://doi.org/10.1109/CVPR.2019.00279>
 6. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).
 7. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2261-2269). <https://doi.org/10.1109/CVPR.2017.243>
 8. Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv*. <https://doi.org/10.48550/arXiv.1905.11946>
 9. Shkiria, S. V., & Korniyk, O. V. (2024). Analysis of digital object authentication methods. *Scientific Bulletin of Uzhhorod University. Series of Mathematics and Informatics*, 45(2), 286-291. [https://doi.org/10.24144/2616-7700.2024.45\(2\).286-291](https://doi.org/10.24144/2616-7700.2024.45(2).286-291).

Shkiria S. V., Andrashko Yu. V. Design and optimization of a deep convolutional neural network in a CIFAR-10 case study.

The article addresses the problem of digital object authentication, which is of great importance for many modern technologies and studies, particularly in image recognition and processing. It describes the challenges associated with image analysis and classification in the context of increasing volumes of digital data, as well as the need to develop efficient methods capable of operating in real-time.

Modern technologies, such as deep learning and neural networks, have opened new opportunities for solving these tasks, providing high accuracy and performance. However, unresolved issues remain, including image processing under poor lighting conditions, object diversity, the need for large training datasets, and high computational complexity.

The article analyzes contemporary face recognition methods, including DeepFace, DeepID, FaceNet, VGG Face, and OpenFace. A comparative analysis of their accuracy, features, and computational resource requirements is provided. Particular attention is given to the use of deep convolutional neural networks (DCNN) in combination with specialized loss functions, which ensure high recognition accuracy.

The proposed directions can significantly expand the applicability of these technologies in applied problems.

Keywords: Digital object authentication, Image recognition, Deep learning, Neural networks, Face recognition methods.

Одержано 12.09.2025