

## АНАЛІЗ АРХІТЕКТУР ТА КОМПОНЕНТІВ ДЛЯ ЕФЕКТИВНОГО НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ СИСТЕМ РОЗПІЗНАВАННЯ ЕМОЦІЙ, ВІКУ ТА СТАТІ

*М.А. Кравченко<sup>1</sup>, С.М. Коваленко<sup>2</sup>*

*<sup>1</sup> магістрант кафедри ПІІТУ, НТУ «ХПІ», Харків, Україна*

*<sup>2</sup> доцент кафедри ПІІТУ, канд. техн. наук, НТУ «ХПІ», Харків, Україна  
[maksym.kravchenko@cs.khpi.edu.ua](mailto:maksym.kravchenko@cs.khpi.edu.ua)*

Системи розпізнавання емоцій, віку та статі є важливими компонентами сучасних прикладних рішень у галузі комп'ютерного зору. Основним інструментом для таких систем є нейронні мережі, що потребують ефективного навчання на великих датасетах. Вибір архітектури та компонентів для навчання відіграє критичну роль у досягненні високої точності моделей. У дослідженні розглядається процес навчання нейронних мереж для розпізнавання емоцій, віку та статі, а також проводиться порівняльний аналіз основних бібліотек: Caffe, PyTorch і TensorFlow [1].

Навчання нейронної мережі полягає в ітеративному оновленні вагових коефіцієнтів моделі за допомогою алгоритму зворотного поширення помилки (backpropagation) та градієнтного спуску. Для систем розпізнавання емоцій, віку та статі важливими етапами навчання є збір і підготовка даних, які повинні бути збалансованими за кількістю класів (емоцій, віку та статі) та мати достатню кількість зразків для кожного з класів. Використання глибинних нейронних мереж, зокрема Convolutional Neural Networks (CNN), є стандартним підходом у задачах комп'ютерного зору. Мережі мають містити кілька згорткових шарів для автоматичного виділення ознак. Важливо також вибрати правильну функцію втрат: для задач класифікації емоцій, віку та статі зазвичай використовуються категоріальна перехресна ентропія або середньоквадратична похибка для регресійних задач, таких як визначення віку. Оптимізація моделі забезпечується за рахунок адаптивних оптимізаторів, таких як Adam або RMSprop, що підвищують швидкість та стабільність навчання [2]. Ключовими аспектами ефективного навчання є налаштування гіперпараметрів, зокрема швидкості навчання (learning rate), а також використання технік регуляризації, таких як Dropout, для уникнення перенавчання (overfitting). Крім того, важливо проводити валідацію на кожній ітерації для контролю за прогресом і точністю моделі.

Серед інструментів, які використовуються для навчання нейронних мереж, Caffe вирізняється високою швидкістю виконання, що робить його ефективним для задач комп'ютерного зору. Однак, ця бібліотека орієнтована переважно на операційні системи Linux, що може бути обмеженням для користувачів інших платформ. Крім того, Caffe є менш гнучкою у порівнянні з іншими бібліотеками, коли йдеться про складні архітектури або індивідуальні налаштування моделей. PyTorch є однією з найпопулярніших бібліотек завдяки своїй гнучкості та зручності у використанні, зокрема через динамічну обчислювальну графіку, яка дозволяє змінювати модель "на льоту". Це особливо корисно при дослідженнях і експериментах з новими архітектурами. PyTorch підтримується на всіх основних платформах, таких як Windows, Linux та MacOS, що робить його універсальним інструментом. TensorFlow також є потужною платформою, що підтримує навчання та розгортання нейронних мереж на різних пристроях, від серверів до мобільних телефонів. Використання статичної обчислювальної графіки робить TensorFlow менш зручним для досліджень, ніж

PyTorch, однак ця особливість є корисною для великих виробничих проектів, що вимагають стабільності та масштабованості [3].

Датасети відіграють ключову роль у процесі навчання нейронних мереж, оскільки від їх якості та розміру залежить точність моделі. Для розпізнавання емоцій, віку та статі важливо використовувати великі, збалансовані та різноманітні датасети, що охоплюють різні категорії. Чим більше варіацій даних представлено в наборі, тим краще мережа зможе узагальнювати інформацію, а отже, краще справлятиметься з новими, раніше невідомими їй прикладами.

Середня кількість даних для навчання моделі залежить від складності завдання та архітектури моделі. Для завдань класифікації емоцій, віку та статі рекомендується використовувати десятки тисяч зображень, хоча в ідеалі кількість зразків може досягати кількох сотень тисяч для більшої точності. Наприклад, для класифікації емоцій потрібен збалансований датасет, що містить від 10 000 до 50 000 зображень для кожного класу емоцій, залежно від кількості емоцій, що розпізнаються. Що стосується віку та статі, важливо мати представництво різних вікових груп та обидвох статей, щоб уникнути перекосів у навчанні моделі [4].

Крім кількості, якість даних також є критичним фактором. Датасети повинні бути чітко анотовані, з мінімумом шумів або некоректних міток. Наявність великої кількості "брудних" даних може суттєво знизити ефективність навчання та призвести до неправильних прогнозів моделі. Бажано, щоб зображення мали високу роздільну здатність, особливо для задач, пов'язаних із розпізнаванням облич, де важливо точно відтворити деталі. Загалом, збалансований датасет, що включає різноманітні умови освітлення, різні ракурси та індивідуальні особливості осіб, забезпечує найкращі результати в процесі навчання нейронних мереж [5].

Отже, ефективність навчання нейронних мереж для систем розпізнавання емоцій, віку та статі значною мірою залежить як від обраних компонентів і архітектур, так і від якості та кількості даних у датасеті. Вибір між бібліотеками, такими як Caffe, PyTorch чи TensorFlow, визначається специфікою завдання, доступністю ресурсів та вимогами до платформи. Проте навіть найсучасніша архітектура не забезпечить високої точності без належного датасету. Великий, різноманітний і добре анотований набір даних є необхідною умовою для узагальнення інформації та коректного функціонування моделі в реальних умовах. Збалансований підхід до вибору компонентів та якісних даних є ключем до створення потужних і точних систем розпізнавання.

### **Список літератури:**

1. *Kravchenko, M. A. Analysis of toolkits for facial recognition system / M.A. Kravchenko, S.M. Kovalenko // Інформаційні технології: наука, техніка, технологія, освіта, здоров'я: тези доповідей XXXII міжнародної науково-практичної конференції MicroCAD-2024, 22-25 травня 2024 р. / за ред. проф. Сокола Є.І. – Харків: НТУ «ХПІ». – Р. 1219.*
2. *Taigman, Y. DeepFace: Closing the Gap to Human-Level Performance in Face Verification / Y. Taigman, M. Yang, M. Ranzato, L. Wolf // 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA. – 2014. – Р. 1701 – 1708.*
3. *Dinghofer, K. Analysis of Criteria for the Selection of Machine Learning Frameworks / K. Dinghofer, F. Hartung // International Conference on Computing, Networking and Communications (ICNC) 57. – 2020, P. 2325 – 2626.*
4. *Luo, P. Hierarchical face parsing via deep learning / P. Luo, X. Wang, X. Tang // In Proc. Conf. Comput. Vision Pattern Recognition. – 2012. – P. 2480–2487.*
5. *Kovalenko, S. M. Approach to the automatic creation of an annotated dataset for the detection, localization and classification of blood cells in an image / S. M. Kovalenko, O. S. Kutsenko, S. V. Kovalenko, A. S. Kovalenko // Radio Electronics, Computer Science, Contro. – 2024. – No 1. –P. 128 - 139. doi: <https://doi.org/10.15588/1607-3274-2024-1-12>*