процесів створюють можливість виконувати багатоваріантний аналіз і оптимізацію параметрів друку до початку експериментів, а їх інтеграція з ML-моделями формує основу замкненого контуру керування якістю. Запропонована модель інтелектуальної системи визначення характеристик якості охоплює всі етапи – від сировини до постпроцесної обробки – й дає змогу розглядати металеве адитивне виробництво як елемент кіберфізичної виробничої системи Індустрії 4.0 [1].

Список використаних джерел

1. Григоренко Я. С., Кузмічов М. Ф., Дмитрієв Д. О. Інтелектуальні технології визначення характеристик якості металевих виробів в процесах пошарового утворення. *Вісник Херсонського національного технічного університету.* 2025. № 4(95). С. 84–91.

2. Yeshiwas T. A., Tiruneh A. B., Sisay M. A. A review article on the assessment of additive manufacturing. *Journal of Materials Science: Materials in Engineering.* 2025. Vol. 20, № 85.

3. Gao B., Wu D., Wang Y. A review of research progress in selective laser melting (SLM). *Micromachines.* 2023. Vol. 14(1). P. 57.

4. Machine learning assisted quality control in metal additive manufacturing: A review / Z. Hu, X. Gao, Y. Li et al. *Advanced Powder Materials.* 2025. URL: https://www.sciencedirect.com/science/article/pii/S2772834X25000788 (дата звернення: 16.11.2025).

5. Ben Amor S., Brissaud D., Rivette M. Digital twin implementation in additive manufacturing: A comprehensive review. *Processes.* 2024. Vol. 12, № 6. P. 1062. URL: https://www.mdpi.com/2227-9717/12/6/1062 (дата звернення: 16.11.2025).

УДК 004.001

Зелінський О. О, здобувач 1 курсу
ОС «Магістр» ФЗ Комп'ютерні науки,
Січко Т. В., канд. техн. наук, доцент
кафедри інформаційних технологій

ІНТЕРПРЕТАЦІЇ РЕЗУЛЬТАТІВ ШТУЧНОГО ІНТЕЛЕКТУ ПІД ЧАС ІДЕНТИФІКАЦІЇ ЗОБРАЖЕНЬ

Донецький національний університет імені Василя Стуса, м. Вінниця

Ідентифікація зображень є одним із найбільш динамічних напрямів штучного інтелекту, у якому глибинні нейронні мережі демонструють здатність до високоточної класифікації та розпізнавання візуальних об'єктів [1]. Проте разом із ефективністю ці методи приносять і суттєву проблему: користувачі не можуть безпосередньо оцінити логіку прийняття рішення моделлю. Більшість сучасних архітектур – CNN, ResNet, Vision Transformers – працюють за принципом «чорної скриньки», коли рішення є точним, але його причини залишаються непрозорими [1; 4].

Така непрозорість створює нові виклики: помилки в ідентифікації можуть залишатися непоміченими, модель може навчитися на випадкових, а не релевантних ознаках, а в критично важливих сферах (медицина, безпека, промисловість) відсутність пояснень може становити прямий ризик. Саме тому зростає значення пояснюваних систем – Explainable Artificial Intelligence (XAI), що забезпечують інтерпретацію процесу прийняття рішень [5].

Глибинні моделі працюють із багатовимірними представленнями, тому їх внутрішні обчислення недоступні для прямої інтерпретації [5]. Результати класифікації можуть бути арифметично правильними, але водночас базуватися на нерелевантних патернах – наприклад, на фоновому шумі, кольорі зображення або сторонніх артефактах [1].

У реальних задачах це призводить до проблем:

– у медичних зображеннях модель може фокусуватися не на патології, а на текстових позначках на знімку [1];

– у системах відеоспостереження алгоритм іноді орієнтується на фон, а не на обличчя;

– у промислових системах контролю якості модель може «навчитися» реагувати на освітлення, а не на структуру поверхні.

Отже, інтерпретованість стає критично важливою, оскільки без неї неможливо оцінити коректність логіки моделі [5].

Методологія XAI включає низку підходів, які надають користувачам можливість отримувати докладні пояснення до кожного рішення моделі [5]. Серед основних методів:

1. Візуальні методи:

– Grad-CAM – будує теплові карти важливих областей зображення, підсвічуючи ті ділянки, що найбільше вплинули на класифікацію [2];

– CAM-варіації (Score-CAM, Guided Backpropagation, SmoothGrad) – забезпечують уточнені або згладжені візуальні пояснення [2];

– Attention Maps у Vision Transformer – дають змогу безпосередньо візуалізувати механізм уваги, що є природною особливістю трансформерів [4].

Ці методи дають змогу побачити, чи дійсно модель «дивиться» на необхідний об’єкт.

2. Локальні пояснення:

– LIME – створює локальні апроксимації навколо конкретного прогнозу, даючи змогу зрозуміти, які регіони зображення вплинули на рішення [3];

– SHAP – оцінює внесок кожної ознаки, побудований на ігровій теорії [3].

Ці методи допомагають пояснювати окремі рішення, а не модель загалом.

3. Глобальні пояснення:

– аналіз шарів нейронних мереж;

– побудова інтерпретованих сурогатних моделей;

– дослідження чутливості моделі до змін у даних [5].

Глобальні пояснення допомагають оцінити загальну логіку моделі та її надійність.

Попри значний розвиток XAI, існує низка проблем [5]:

– пост-хок-пояснення не гарантують істинності. Деякі методи лише наближають пояснення, не розкриваючи справжньої логіки моделі [3];

– нечіткість теплових карт. Інколи карти активностей не дають змоги однозначно визначити, на які саме області спиралася модель [2];

– висока обчислювальна складність. Деякі методи, наприклад, SHAP, потребують великої кількості обчислень [3];

– відсутність універсальних метрик оцінки якості пояснень. Сьогодні не існує загальноприйнятого способу виміряти, наскільки інтерпретація «правильна» [5].

Ці недоліки зумовлюють потребу у вдосконаленні інтерпретованих моделей та формуванні більш стандартизованих підходів.

Інтерпретація результатів не є лише теоретичною задачею – вона має прикладне значення у багатьох галузях [5]:

– медицина: пояснюваність дає змогу лікарю переконатися, що модель аналізує саме клінічно значущі області;

– промисловість: полегшує виявлення причин помилкових рішень у системах виявлення дефектів;

– транспорт: дає змогу аналізувати рішення систем автономного керування та виявляти помилкові тригери;

– безпека: пояснюваність допомагає уникнути помилкових спрацювань та знижує ризик упередженості;

– агросектор: дає змогу зрозуміти, які візуальні патерни використовуються під час діагностики стану рослин чи ґрунтів.

У кожній із цих сфер інтерпретація допомагає підвищити довіру до системи та мінімізувати вплив помилок.

Сучасні тенденції свідчать про рух у напрямі створення моделей, які є інтерпретованими за своєю природою, а не лише за допомогою пост-хок-методів [5]. До перспектив належать:

– розробка архітектур, що поєднують точність і прозорість;

– використання генеративних моделей для кращого аналізу важливих ознак [3];

– поєднання XAI із мультимодальними системами, де текстові пояснення можуть супроводжувати візуальні;

– автоматична діагностика упередженості моделей на основі пояснень;

– розвиток стандартизованих метрик для оцінювання якості інтерпретацій.

Отже, XAI перетворюється на самостійний напрям, що визначає майбутнє систем комп’ютерного зору.

Інтерпретація результатів штучного інтелекту під час ідентифікації зображень є фундаментальним складником побудови надійних, прозорих і безпечних систем комп’ютерного зору. Вона дає змогу не лише покращити якість моделей, а й забезпечити довіру користувачів, зрозуміти логіку роботи алгоритмів та уникнути критичних помилок. Розвиток методології Explainable AI сприятиме створенню моделей, здатних не лише правильно класифікувати об’єкти, а й аргументувати власні рішення. Це відкриває нові можливості для впровадження ШІ у відповідальні сфери та формує підґрунтя для подальших наукових досліджень.

Список використаних джерел

1. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*. 2015. Vol. 521, № 7553. P. 436–444.

2. Grad-CAM: visual explanations from deep networks via gradient-based localization / R. R. Selvaraju, M. Cogswell, A. Das et al. *Proceedings of the IEEE International Conference on Computer Vision (ICCV).* 2017. P. 18–26.

3. Ribeiro M., Singh S., Guestrin C. Why should I trust you?: explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016. P. 115–124.

4. An image is worth 16×16 words: transformers for image recognition at scale / A. Dosovitskiy, L. Beyer, A. Kolesnikov et al. *Advances in Neural Information Processing Systems.* 2020. Vol. 33. P. 58–68.

5. Кульчицька О. Ю., Січко Т. В. Цифрова обробка зображень та відео. *Прикладні інформаційні технології:* тези доповідей Всеукраїнської науково-практичної конференції (м. Вінниця, 29 квітня 2020 р.). Вінниця, 2020. С. 110–111.

УДК 004.8

Грибаніна А. О., студентка 4 курсу спеціальності 122 «Комп'ютерні науки»,
Ніколюк П. К., д-р фіз.-мат. наук, професор кафедри інформаційних технологій

МОДЕЛІ НЕЙРОННИХ МЕРЕЖ

Донецький національний університет імені Василя Стуса, м. Вінниця

Стрімкий розвиток технологій штучного інтелекту (ШІ) зумовлює зростання попиту на адаптивні моделі, здатні обробляти складні структури даних у режимі реального часу. У сучасних дослідженнях центральне місце займають нейронні мережі – обчислювальні системи, що успішно вирішують задачі класифікації, прогнозування, аналізу образів, керування автономними системами та моделювання складних процесів.

Нейронні мережі, на відміну від класичних алгоритмів, дають змогу виявляти приховані залежності, працювати з великими обсягами даних і формувати узагальнені представлення, що підвищує їх практичну ефективність. Мета цієї роботи – проаналізувати сучасні архітектури нейронних мереж (2017–2024) та продемонструвати апроксимацію функції в MATLAB як приклад їх застосування. Особливу увагу приділено автоенкодерам, згортковим мережам (CNN), рекурентним структурам (RNN) та трансформерам.

Сучасний етап розвитку нейронних мереж характеризується переходом від простих багатошарових моделей до архітектур, які здатні опрацьовувати просторові, часові та гібридні дані (рис. 1).

До ключових ознак сучасних ШНМ належать:

– ієрархічність структури, що дає змогу поетапно формувати ознаки;
– механізми уваги, які підсилюють релевантні елементи даних;
– паралельна обробка інформації, характерна для трансформерних моделей;