

УДК 620.186:623.746-519 (55)

DOI: [https://doi.org/1034169/2414-0651.2024.1\(41\).128-141](https://doi.org/1034169/2414-0651.2024.1(41).128-141)

І. Б. ЧЕПКОВ, доктор технічних наук, професор
<https://orcid.org/0000-0002-4294-4152>

В. І. СЛЮСАР, доктор технічних наук, професор
<https://orcid.org/0000-0002-2912-3149>

А. В. КУЧИНСЬКИЙ, кандидат технічних наук,
старший науковий співробітник
<https://orcid.org/0000-0003-1687-9393>

МЕТОДИКА ПОБУДОВИ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ІДЕНТИФІКАЦІЇ КОНСТРУКЦІЙ СПЕЦІАЛЬНОГО ПРИЗНАЧЕННЯ ЗА КЛАСИФІКАЦІЄЮ ТА КРАЇНОЮ ПОХОДЖЕННЯ НА ПІДСТАВІ ЕЛЕКТРОННО-МІКРОСКОПІЧНОГО АНАЛІЗУ ЇХ ФРАГМЕНТІВ

У статті представлено методику візуального аналізу мікроструктур металевих зразків для їх класифікації на основі застосування технологій штучного інтелекту. Для формування датасету було використано серію фотографій, зроблених за допомогою мікроскопа, які представляли різні класи металів. Кожен клас металу мав унікальні візуальні характеристики, такі як: розмір зерен, форма та розподіл фаз, наявність інклюдцій та ознаки механічної обробки. Для ідентифікації мікроструктур тестових зразків використовувалися згорткові нейронні мережі. В результаті проведених експериментів підтверджено ефективність використання нейромереж для класифікації двох та трьох типів металів. На додаток до нейромереж в роботі описано особливості залучення до класифікації металів великої мовної моделі GPT-4v і продемонстровано її ефективність.

Ключові слова: датасет, нейромережа, аугментація, матриця переплутувань, GPT-4v.

Серед відомих методів автоматизації процесів класифікації та ідентифікації конструкцій спеціального призначення на основі електронно-мікроскопічного аналізу найбільш перспективним підходом слід вважати застосування нейронних мереж. Їх сучасні спроможності дозволяють забезпечити класифікацію зображень за різними ознаками [1–5].

Разом з тим, слід зазначити, що вказана ідентифікація конструкцій спеціального призначення на підставі ана-

лізу їх фрагментів за допомогою штучних нейронних мереж є складним завданням, яке вимагає комплексного підходу. В якості базової методики для вирішення такої задачі пропонується багатоетапна послідовність дій, яка є типовою для застосування нейронних мереж (рис. 1).

Перший крок запропонованої методики полягає у зборі та підготовці якісного датасету. Для цього необхідно здійснити електронно-мікроскопічний аналіз різноманітних конструкцій спеціального призначення з різних країн і отримати високоякісні зображення фрагментів конструкцій. Зібрані фотознімки далі систематизуються та поділяються на класи за типом матеріалу конструкції і країною походження, що стають мітками для подальшого навчання.

На стадії підготовки даних зображення перетворюються до потрібного розміру, наприклад, 224x224 пікселів для застосування нейромережі типу VGG16 [6]. Щоб розширити набір даних і зменшити можливість перенавчання, слід провести аугментацію даних, яка включає повороти, зміщення, зум, віддзеркалення, часткове розфокусування зображень тощо.

Далі здійснюється вибір архітектури нейронної мережі, що залежить від специфіки завдання та доступних ресурсів. При цьому можуть бути використані готові моделі, такі як VGG16 [6], EfficientNet [7] або MobileNet [8, 9], або ж можна розробляти власну архітектуру. Суттєво, що найкращих результатів вдається отримати у разі використання типових нейромереж, попередньо навчених на датасеті ImageNet [1, 10].

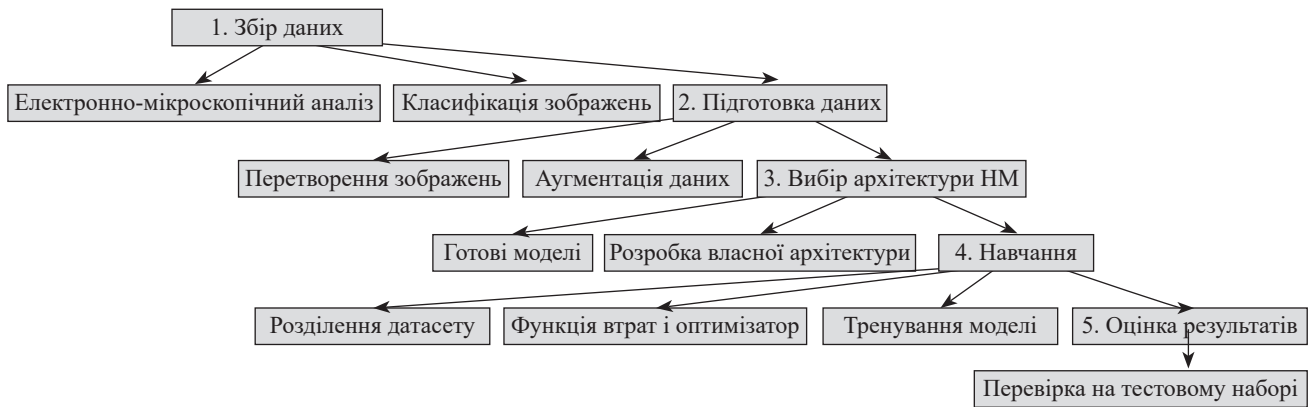
На етапі навчання датасет розділяється на тренувальний, валідаційний та тестовий масиви. Визначаються функція втрат, оптимізатор та його параметри налаштування, крок навчання та розмір батчів. Після цього модель тренується на тренувальному наборі даних, використовуючи валідаційний набір для налаштування параметрів.

По завершенні навчання необхідно перевірити ефективність нейромережі на тестовому наборі, використовуючи так звану матрицю переплутувань, показники точності, повноту та інші метрики для оцінки якості ідентифікації. З метою ефективного впровадження, модель нейромережі інтегрується з інтерфейсом системи аналізу або в застосунок і регулярно оновлюється шляхом додавання нових даних або внесення змін до архітектури для підвищення точності функціонування.

Описана методика є лише загальним підходом. В реальних умовах потрібно буде звертати увагу на багато додаткових деталей, таких як якість зображень, відсутність артефактів на поверхні конструкцій, рівномірний розподіл освітлення, відсутність затінення та ін.

З розвитком технологій та появи нових матеріалів і конструкцій спеціального призначення важливість адаптації та модернізації моделей штучних нейронних мереж буде зростати. Щоб підтримувати актуальність моделі, вона має бути гнучкою до введення нових даних і параметрів. Також рекомендується систематично аналізувати результати ідентифікації, вивчаючи можливі аномалії або неточності в висновках.

Ефективність системи ідентифікації може бути збільшена шляхом інтеграції з іншими аналітичними



Р и с . 1. Методика навчання нейронної мережі з класифікації зображень

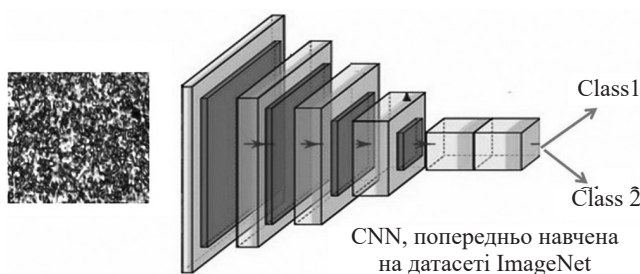
інструментами. Наприклад, більш глибоке розуміння властивостей матеріалу та його походження може бути забезпечене комбінуванням результатів електронно-мікроскопічного аналізу з іншими методами дослідження, такими як спектральний аналіз, ультразвукове сканування та ін.

Побудова ефективної системи ідентифікації на основі нейронних мереж для аналізу конструкцій спеціального призначення вимагає не лише технічної експертизи, але й стратегічного планування, постійного моніторингу та адаптації.

Для виконання описаних завдань може знадобитися спеціалізована архітектура згорткових нейронних мереж. Розглянемо їх більш докладно.

Згорткові нейронні мережі (CNN) для класифікації зображень металів з метою їх ідентифікації

Важливо відзначити, що ефективність нейронних мереж при вирішенні завдань класифікації зображень значною мірою залежить від якості та кількості даних, на яких вони навчаються. Більш того, для кожного конкретного випадку може знадобитися спеціалізована архітектура нейронної мережі. Однак найбільш придатним варіантом нейронних мереж для вирішення такої задачі є згорткові нейронні мережі (CNN) (рис. 2). В загальному випадку для цього доцільно використовувати нейромережі, попередньо навчені на датасеті ImageNet [1]. Вони можуть обробляти двовимірні зображення поверхневих шарів металів, реагуючи на пористість, структурні візерунки, зернистість та деформації з залученням як візуальних зображень у видимому діапазоні оптичного спектру, так і в інфрачервоному або навіть з застосуванням результатів ультразвукового сканування тощо.



Р и с . 2. Типова структура CNN для класифікації зображень

Оскільки розміри зображень можуть перевищувати розміри сцен, на які налаштовані попередньо навчені нейромережі, необхідно провести згорткове зменшення початкових зображень на кількох прошарках вхідного сегменту нейромережі, до якої інтегрується попередньо навчена нейромережна модель. Результатом роботи CNN в контексті завдань класифікації буде розподіл зображень на такі, що містять той чи інший клас металу (рис. 2).

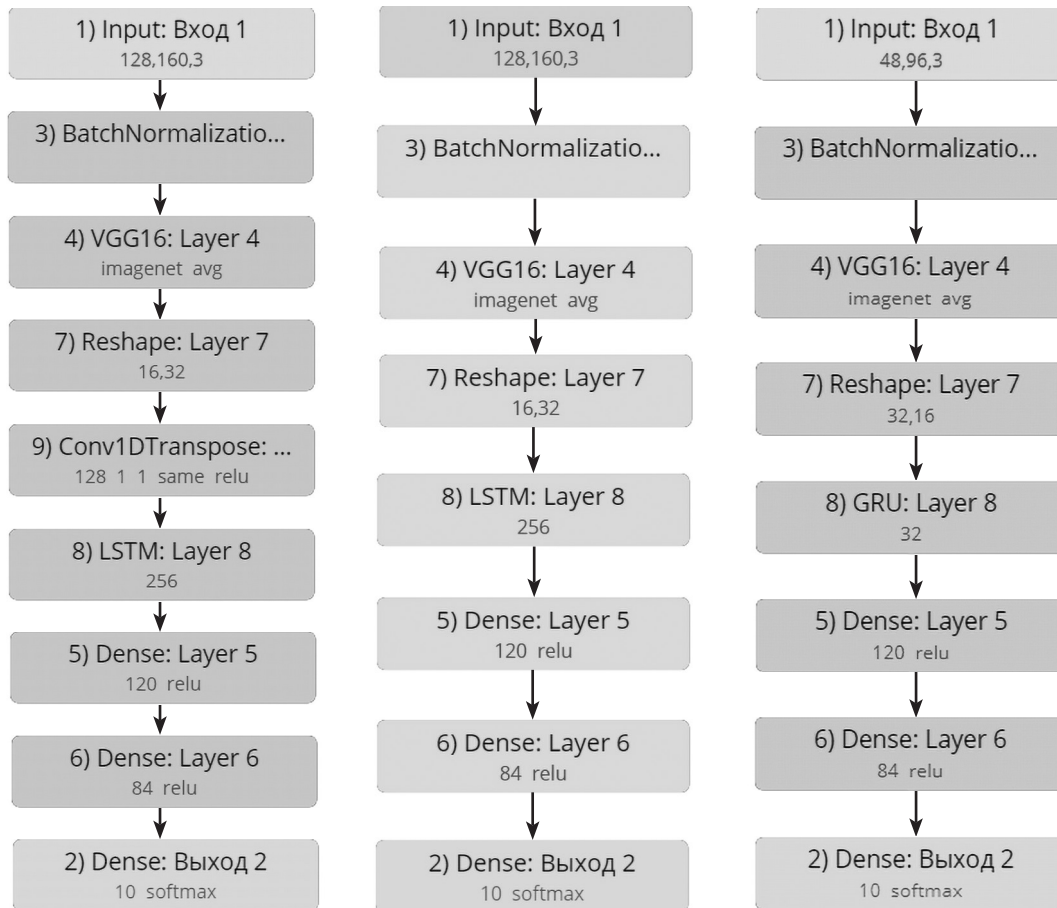
Альтернативний варіант використання CNN-мереж для класифікації зображень полягає у їхній інтеграції з рекурентними нейронними мережами типу LSTM або GRU. Зокрема, на рис. 3 наведено архітектури нейромереж, що містять VGG-16 та LSTM [1] або GRU. Разом з тим слід зазначити, що такі архітектури не відрізняються високою точністю, яка, наприклад, на датасеті CIFAR10 становить для них 88,5–90 %.

Як свідчать результати досліджень, представлені в [1], найбільшої точності класифікації можливо досягти шляхом використання попередньо навчених на датасеті ImageNet нейромереж Xception [11] або InceptionV3 [12] з попереднім масштабуванням вхідних зображень у бік їхнього збільшення (рис. 4, 5). Однак для ефективного застосування такого підходу в контексті ідентифікації типів металів необхідно провести масштабування архітектур Xception [11] або InceptionV3 [12] для обробки зображень великої розмірності, а не передбачених в них обмежених розмірів сцени $299 \times 299 \times 3$ та $229 \times 229 \times 3$ пікселів відповідно.

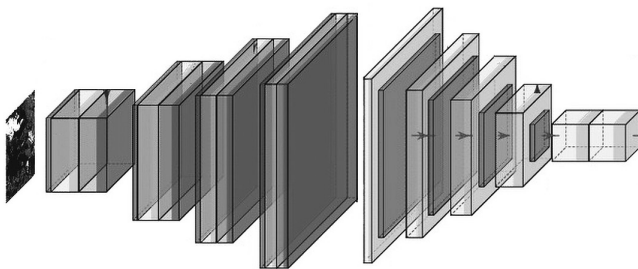
Крім того, такі відмасштабовані варіанти нейромереж необхідно знову попередньо навчити на датасеті типу ImageNet, але з великими за розмірами фото. Для цього необов'язково накопичувати мільйони зображень з розмірами 4288×2848 пікселів чи більше, а можливо відмасштабувати вже наявні в ImageNet фото розмірами 256×256 та 500×500 пікселів за допомогою нейромереж збільшення та поліпшення якості зображень.

Експериментальна перевірка пропонованої методики

Для експериментальної перевірки працездатності пропонованого підходу на першому етапі розглядалося завдання бінарної класифікації. Авторами був підготовлений датасет з фотографіями двох класів металів з загальною кількістю 438 знімків, з яких для навчання нейромережі було виділено 395, а для тестування – 43 зображення.



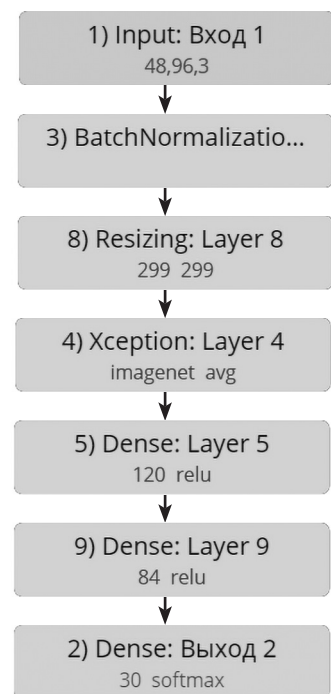
Р и с . 3. Варіанти поєднання VGG-16 з LSTM [1] та GRU для класифікації зображень



Р и с . 4. Ефективна архітектура згорткової мережі для класифікації [1]

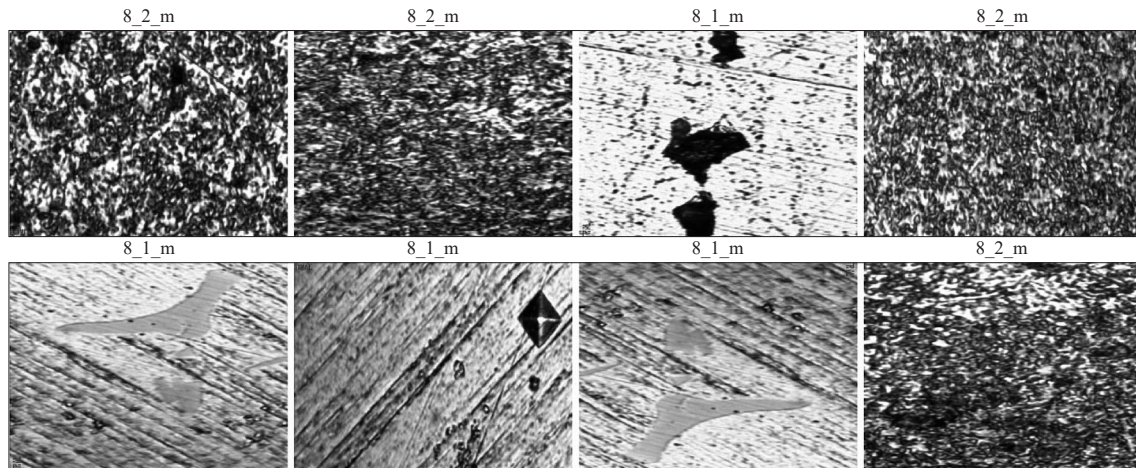
При цьому для одного класу було отримано 246 фото, з яких для тестів відводилося 24 файли, а для другого класу – 192 фото, у тому числі 19 знімків – для тестування ефективності навчання. Таким чином, спостерігався незначний дисбаланс у наповненні датасетів різних класів, однак в цілому рівень дисбалансу був в межах допустимого.

Перед поданням на нейронну мережу початкові фотографії надвисокого розрізнення розрізалися на 12 частин за допомогою програми Image Splitter 1.1 (www.vovsoft.com). Далі з метою збагачення датасету ці фрагменти піддавалися віддзеркаленню, обертанню на 90 градусів вліво та вправо тощо. Потім усі фото приводилися до розміру 320x240 пікселів. Приклад зображень, які підлягали класифікації, наведено на рис. 6. На рис. 7 надано результати застосування процедур аугментації зображень, що вбудовані в мову програмування Python. Як видно, аугментація фактично полягала в частковому

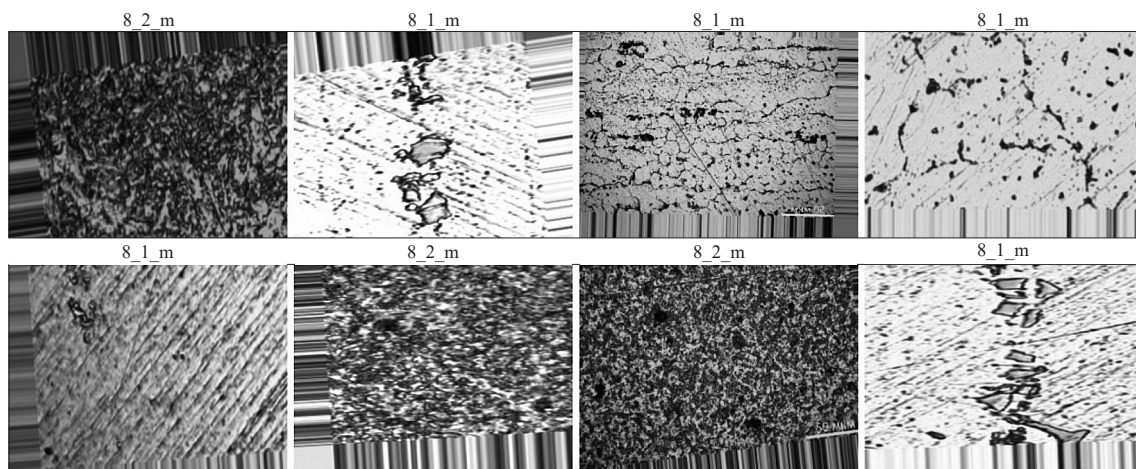


Р и с . 5. Схема підключення попередньо навченої нейромережі Xception

обертанні зображень та обрамленні їх пікселями, розтягнутими за рахунок удаваного зсуву. Однак через такі викривлення частини зображення результати аугментації до остаточного датасету не включалися.



Р и с . 6. Приклади датасету з двома класами металів (умовні назви класів «8_1_m» та «8_2_m»)



Р и с . 7. Результати аугментації зображень

Для навчання використовувалася нейронна мережа, архітектура якої представлена у табл. 1. Вона має розмір 735,19 МБ та 192724946 параметрів, з яких не підлягають тренуванню лише 2048. Архітектура представленої нейронної мережі є послідовною та складається з 19 шарів, кожен з яких виконує унікальну роль у процесі навчання. Початковий шар, `conv2d` (`Conv2D`), є згортковим і призначений для обробки вхідних зображень з розмірами 240x320 пікселів, з використанням 256 фільтрів. Кожен фільтр проходить по вхідному зображенні з ядром згортки 3x3 без зміни розмірів зображення для виявлення певних характеристик. Виходом згортки є функція активації ReLU. Кількість параметрів у цьому шарі становить 7168, що є запорукою формування спроможності виявляти складні особливості вхідних даних.

Другий шар, `batch_normalization` (`BatchNormalization`), застосовує процес нормалізації партій зображень (батчів), що допомагає підвищити стабільність та швидкість навчання, регулюючи розподіл вхідних даних. Нормалізація зменшує проблему внутрішнього зсуву коваріації, дозволяючи кожному шару навчатися більш незалежно.

Третій шар, `conv2d_1` (`Conv2D`), також є згортковим і має аналогічну структуру до першого згорткового шару, пропонуючи додатковий рівень абстракції для виявлення особливостей зображення. Кількість параметрів цього

шару 590080, підкреслює його складність та здатність до деталізації.

Четвертий шар, `max_pooling2d` (`MaxPooling2D`), застосовує операцію максимального згортання, що допомагає зменшити розміри вхідних даних для зниження обчислювальної складності та перевантаження наступних шарів, одночасно зберігаючи важливі особливості оброблюваних сцен.

Подальші шари продовжують альтернацію між згортковими шарами, нормалізацією партій та згортанням максимального пулінгу, додаючи шари відкидання (`dropout`) для запобігання перенавчанню шляхом випадкового ігнорування деяких нейронних з'єднань під час тренування. Завершальні шари мережі складаються з повнозв'язних шарів (`Dense`), які забезпечують класифікацію на основі виявлених особливостей та абстракцій.

Кожен шар у цій архітектурі сприяє поступовій обробці та абстракції вхідних даних, дозволяючи мережі ефективно вчитися розпізнавати складні шаблони та властивості вхідних зображень для точної класифікації.

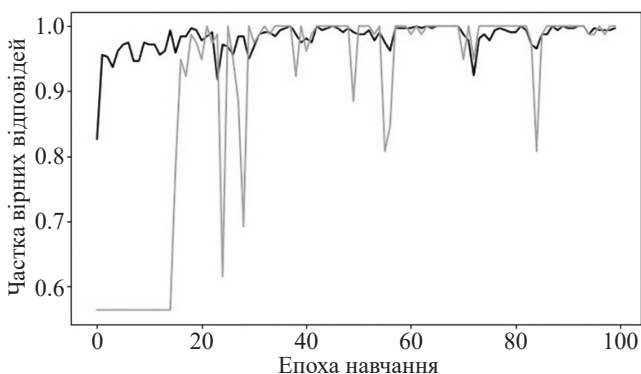
Структура розглянутої нейронної мережі була описана мовою програмування Python й інтегрована до середовища Google Colab Pro+, в межах якого для навчання нейронної мережі було орендовано GPU-прискорювач A100 з 40 ГБ оперативної пам'яті графічного процесора, 166,8 ГБ

Т а б л и ц я 1. Архітектура нейронної мережі бінарної класифікації металів

N п/п	Шар (Тип)	Формат виведення	Кількість параметрів
1	conv2d (Conv2D)	(None, 240, 320, 256)	7168
2	batch_normalization (BatchNormalization)	(None, 240, 320, 256)	1024
3	conv2d_1 (Conv2D)	(None, 240, 320, 256)	590080
4	max_pooling2d (MaxPooling2D)	(None, 80, 106, 256)	0
5	conv2d_2 (Conv2D)	(None, 80, 106, 256)	590080
6	batch_normalization_1 (BatchNormalization)	(None, 80, 106, 256)	1024
7	dropout (Dropout)	(None, 80, 106, 256)	0
8	conv2d_3 (Conv2D)	(None, 80, 106, 256)	590080
9	max_pooling2d_1 (MaxPooling2D)	(None, 26, 35, 256)	0
10	dropout_1 (Dropout)	(None, 26, 35, 256)	0
11	conv2d_4 (Conv2D)	(None, 26, 35, 512)	1180160
12	batch_normalization_2 (BatchNormalization)	(None, 26, 35, 512)	2048
13	conv2d_5 (Conv2D)	(None, 26, 35, 1024)	4719616
14	max_pooling2d_2 (MaxPooling2D)	(None, 8, 11, 1024)	0
15	dropout_2 (Dropout)	(None, 8, 11, 1024)	0
16	flatten (Flatten)	(None, 90112)	0
17	dense (Dense)	(None, 2048)	184551424
18	dense_1 (Dense)	(None, 240)	491760
19	dense_2 (Dense)	(None, 2)	482

дискового простору та отримано у користування 83,5 ГБ оперативної пам'яті системи.

На рис. 8 наведено графіки нормованої точності навчання (вертикальна вісь) в залежності від кількості епох.

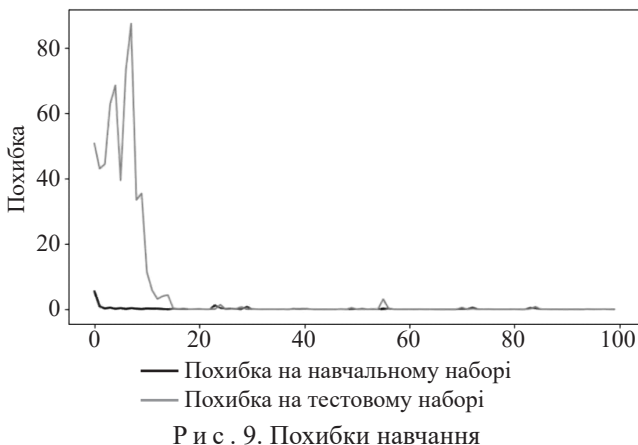


— Частка правильних відповідей на навчальному наборі
— Частка правильних відповідей на тестовому наборі
Р и с . 8. Точність навчання в межах 100 епох

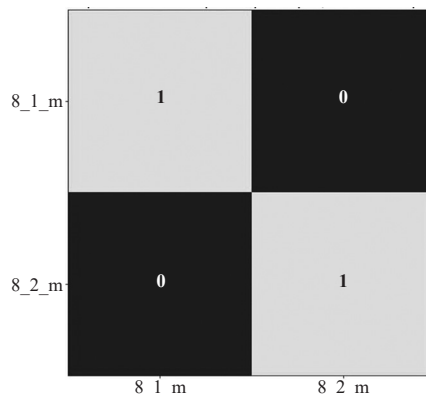
На рис. 9 проілюстровано відповідні похибки навчання. Як видно з рис. 8, максимальна досягнута точність навчання на валідаційній вибірці становить 100 %. До речі, такий же показник був досягнутий і на тестовій вибірці, по тим зображенням, які нейромережі були надані для аналізу вперше.

Ще однією характеристикою якості навчання є нормована матриця переплутувань, яка схематично відображена на рис. 10 і в наведеному випадку ілюструє відсутність прийняття хибних рішень (нулі на побічній діагоналі матриці).

На наступному етапі досліджень завдання класифікації було ускладнене умовою ідентифікації трьох різ-

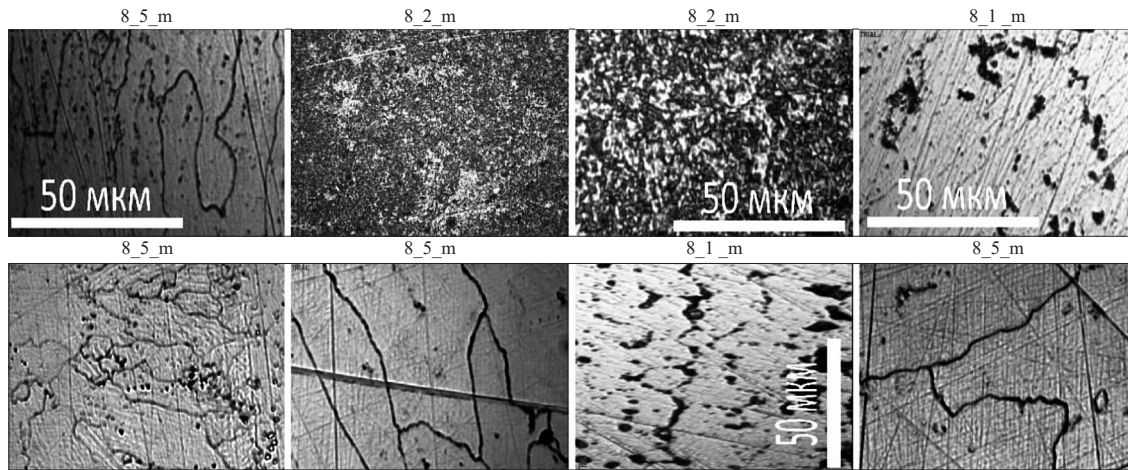


Р и с . 9. Похибки навчання

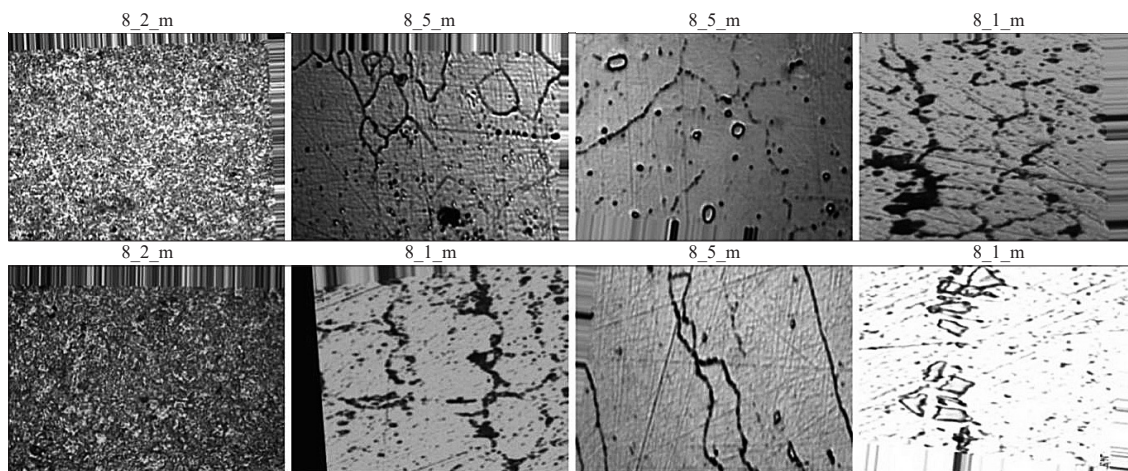


Р и с . 10. Нормована матриця переплутувань

них класів металів, умовно названих «8_1_m», «8_2_m» та «8_5_m». Відповідний датасет був збільшений до 637 зображень, з яких для навчання нейромережі було виділено 575, а для тестування – 62 фотознімки. При цьо-



Р и с . 11. Приклади датасету з трьома класами металів (умовні назви класів «8_1_m», «8_2_m» та «8_5_m»)



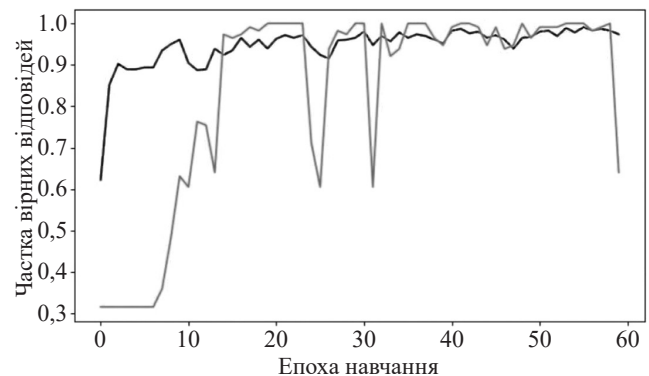
Р и с . 12. Додаткові зображення, що можна отримати з трикласового датасету за допомогою операції аугментації зображень

му сегмент першого класу «8_1_m» налічував 246 фото, а для теста відводилося 24 відповідних зображень. Клас «8_2_m» налічував 192 фото з додатковими 10 % (19) до цієї кількості для тестової вибірки. Обсяг зображень класу «8_5_m» сягав 199 фото й був представлений у тестовій вибірці також 19 знімками. Розмірність фото обмежувалась форматом 320x240 пікселів, розрізнявальна здатність мікроскопу була задана на рівні 50 мкм. Приклад вибірки з вихідного масиву зображень описаного варіанту датасету та результатів їх аугментації наведено на рис. 11, 12.

Структура нейромережі була залишена такою ж, як і для розглянутої бінарної класифікації двох класів металів: 19 шарів, об'єднаних у послідовну архітектуру. Відмінність полягала лише у розмірності згортки та вихідного шару Dense, який отримав три виходи.

Слід зазначити, що збільшення кількості класів цілком очікувано мало призвести до погіршення точності навчання. Однак через обмежений обсяг датасету все одно була зафіксована максимальна середньокласова точність класифікації 100 % (рис. 13), хоча на окремих епохах похибка навчання й мала викиди до відносного рівня 3,5 (рис. 14).

Разом з тим, тестова вибірка дозволила з'ясувати реальну спроможність класифікації. Нормована матриця

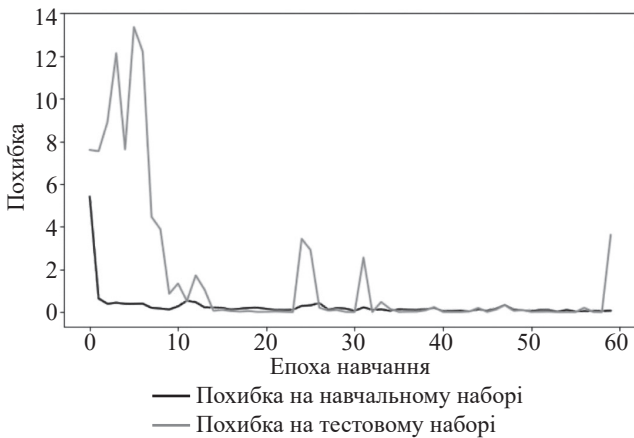


— Частка правильних відповідей на навчальному наборі
 - - Частка правильних відповідей на тестовому наборі

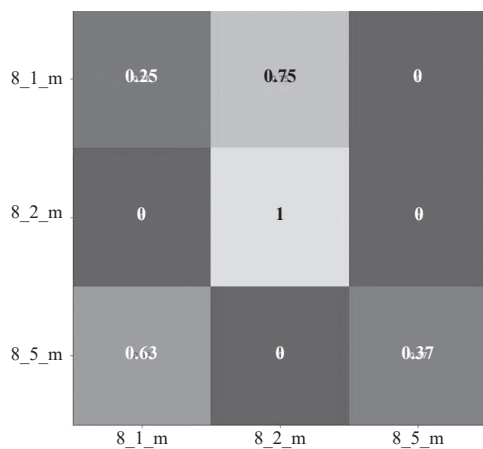
Р и с . 13. Усереднена для трьох класів точність навчання на тренувальній та валідаційній вибірках

переплутувань (рис. 15), що побудована за результатами застосування навченої нейромережі для обробки зображень трикласової тестової вибірки, засвідчила факт наявності хибних рішень.

Зокрема у 75 % випадків нейромережа невірно сприйняла клас «8_1_m» як метал класу «8_2_m» та, навпаки, віднесла 63 % зображень класу «8_5_m» до класу «8_1_m». При цьому середня по усіх класах точність



Р и с . 14. Усереднена по трьох класах похибка навчання на тренувальній та валідаційній вибірках



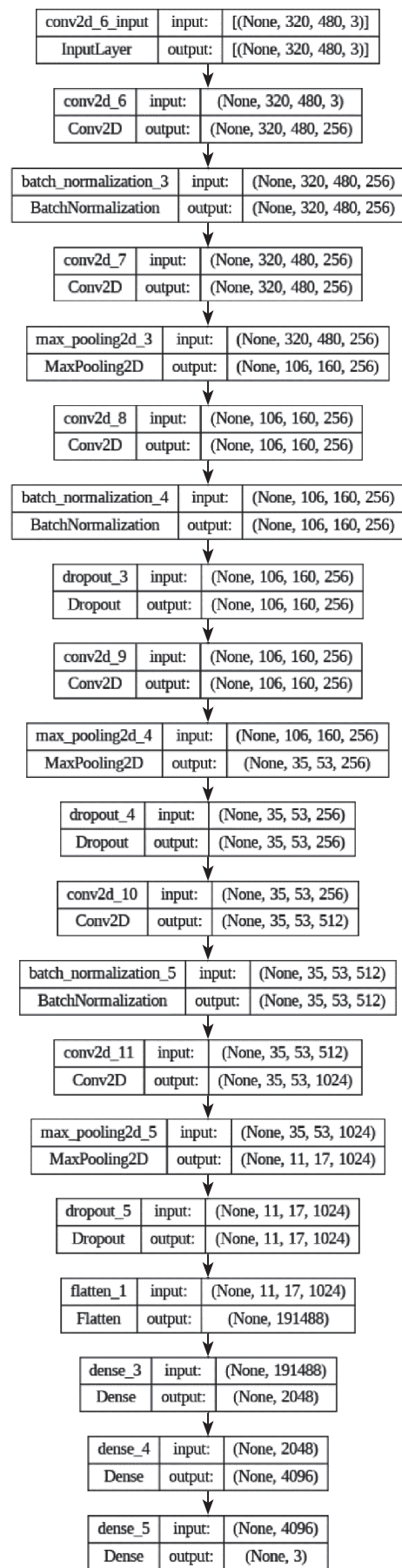
Р и с . 15. Нормована матриця переплутувань для трикласової тестової вибірки

розпізнавання металів становила 54 % завдяки відсутності похибок в ідентифікації класу «8_2_m».

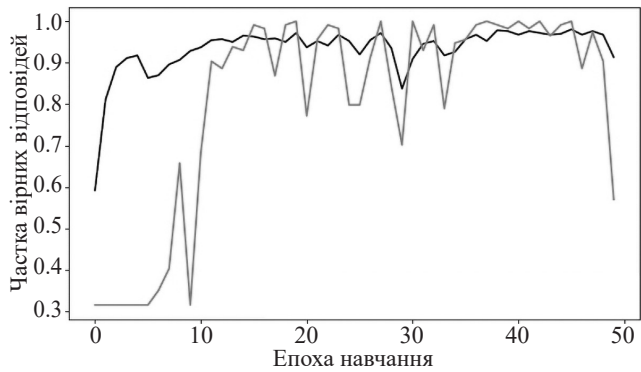
Окрім потреби у збільшенні кількості зображень в датасеті, для підвищення ефективності багатокласової ідентифікації металів доцільно також оперувати зображеннями більшого формату. При цьому дефіцит наявного масиву зображень фактично компенсується більшою площею аналізованої поверхні металу. Для підтвердження цього висновку було проведено аналогічне навчання неймережі по фотознімках формату 480x320 пікселів. Налаштована під такий розмір зображень структура неймережі з зазначенням параметрів шарів наведена на рис. 16.

Результати навчання, проілюстровані на рис. 17 і 18, засвідчили зменшення рівня викидів у похибках та в точності навчання на більшості з несприятливих епох. Це, в свою чергу, призвело до зниження інтенсивності переплутувань класів у відповідній матриці переплутувань по тестовій частині датасету (рис. 19). Як і раніше, без похибок ідентифікувався метал «8_2_m», а правильна ідентифікація металів «8_1_m» та «8_5_m» зросла до 58 % та 53 % відповідно. В результаті усереднена по всіх класах точність розпізнавання металів на тестовій вибірці збільшилася до 70 %.

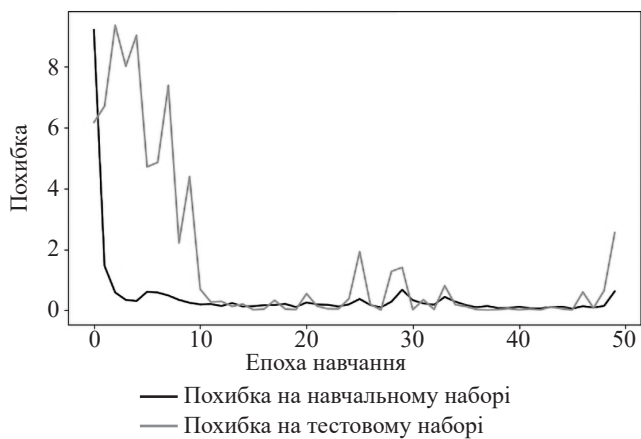
Для закріплення отриманого ефекту поліпшення точності класифікації при переході до більших за розміра-



Р и с . 16. Структура неймережі класифікації зображень поверхонь трьох класів металів формату 480x320 пікселів



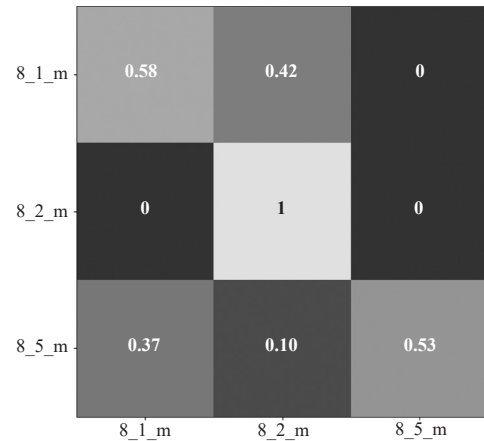
— Частка правильних відповідей на навчальному наборі
 — Частка правильних відповідей на тестовому наборі
 Р и с . 17. Зменшення рівнів погіршення точності навчання між 15-ю та 40-ю епохами при форматі зображень 480x320 пікселів



— Похибка на навчальному наборі
 — Похибка на тестовому наборі
 Р и с . 18. Ілюстрація зниження рівнів аномальних викидів похибки навчання при розмірі зображень 480x320 пікселів

ми зображень аналогічні дослідження були проведені для датасету з розмірністю зображень 640x480 пікселів (рис. 20).

Структура згорткової мережі, що була задіяна, наведена в табл. 2. Її особливістю є зростання загальної кількості параметрів до 836074755, з яких тренуванню підлягало 836072707. При цьому об'єм нейромережі до-

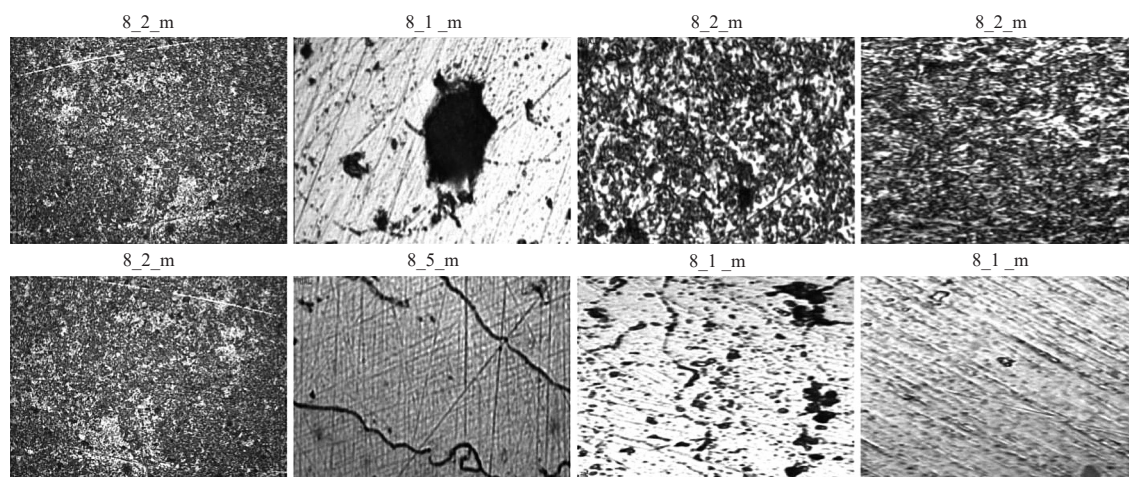


Р и с . 19. Нормована матриця переплутувань при класифікації тестових зображень 480x320 пікселів

сяг 3,11 ГБ. Це збільшило вимоги до апаратних ресурсів, необхідних для проведення обчислень в процесі навчання та інференсу. В результаті на тому ж орендованому сервісі Google Colab Pro+ на основі прискорювача A100 вдалося завершити етапи навчання та валідації його результатів (рис. 21, 22), тоді як при побудові матриці переплутувань відповідна програма розрахунків дала збій через нестачу оперативної пам'яті обчислювальної системи.

Разом з тим, поведінка залежностей точності та похибок навчання, представлених на рис. 21 і 22, наочно продемонструвала подальше зростання якості процесу налаштування гіперпараметрів нейромережі. Зокрема, на валідаційній вибірці, починаючи з 15-ї епохи і до завершення процесу навчання, не було жодного випадку зниження середньокласової точності класифікації нижче рівня 70 %.

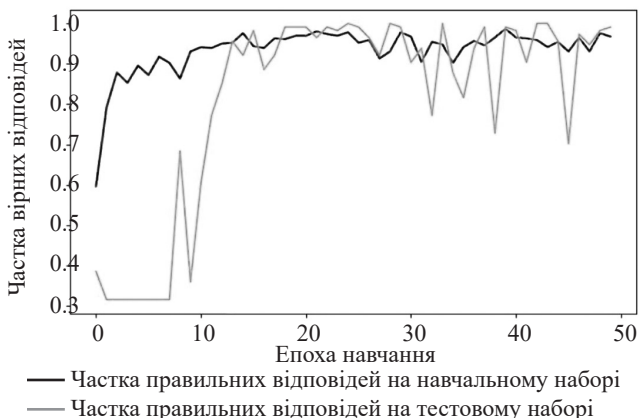
Таким чином, в рамках проведеного дослідження візуальний аналіз мікроструктур на основі згорткових нейромереж виявився ефективним у вирішенні завдання класифікації металів. Незважаючи на обмеження цього методу, він може застосовуватися як експрес-інструмент першої оцінки перед проведенням більш комплексного аналізу. Подальші дослідження доцільно зосередити на



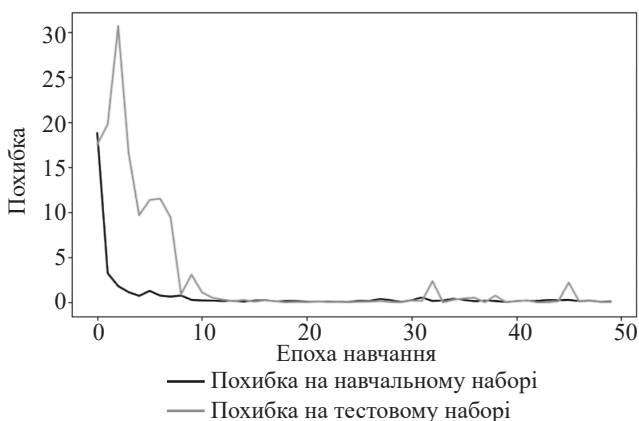
Р и с . 20. Фрагмент датасету з зображеннями формату 640x480 пікселів

Таблиця 2. Структура нейронної мережі з ідентифікації трьох класів металів по зображеннях їх поверхні у форматі 640×480 пікселів

N п/п	Шар (тип)	Формат виведення	Кількість параметрів
1	conv2d (Conv2D)	(None, 480, 640, 256)	7168
2	batch_normalization (BatchNormalization)	(None, 480, 640, 256)	1024
3	conv2d_1 (Conv2D)	(None, 480, 640, 256)	590080
4	max_pooling2d (MaxPooling2D)	(None, 160, 213, 256)	0
5	conv2d_2 (Conv2D)	(None, 160, 213, 256)	590080
6	batch_normalization_1 (BatchNormalization)	(None, 160, 213, 256)	1024
7	dropout (Dropout)	(None, 160, 213, 256)	0
8	conv2d_3 (Conv2D)	(None, 160, 213, 256)	590080
9	max_pooling2d_1 (MaxPooling2D)	(None, 53, 71, 256)	0
10	dropout_1 (Dropout)	(None, 53, 71, 256)	0
11	conv2d_4 (Conv2D)	(None, 53, 71, 512)	1180160
12	batch_normalization_2 (BatchNormalization)	(None, 53, 71, 512)	2048
13	conv2d_5 (Conv2D)	(None, 53, 71, 1024)	4719616
14	max_pooling2d_2 (MaxPooling2D)	(None, 17, 23, 1024)	0
15	dropout_2 (Dropout)	(None, 17, 23, 1024)	0
16	flatten (Flatten)	(None, 400384)	0
17	dense (Dense)	(None, 2048)	819988480
18	dense_1 (Dense)	(None, 4096)	8392704
19	dense_2 (Dense)	(None, 3)	12291



Р и с . 21. Еволюція усередненої по трьох класах точності класифікації для тренувального та валідаційного датасетів



Р и с . 22. Зміна рівня похибки навчання залежно від номера епохи

застосуванні попередньо навчених нейромереж більш складної структури за умови розширення кількості класів металів та збільшення обсягів датасету. Розвиток запропонованого підходу має спиратися на комбінування різних архітектур нейронних мереж, що дозволить сполучити сильні сторони задіяних архітектур та зменшить вплив слабких місць окремих моделей і призведе до більш точної та надійної ідентифікації металевих структур.

Перевірка можливості використання великих мовних моделей

Слід зазначити, що окрім розглянутих варіантів застосування згорткових нейромереж для ідентифікації металів на основі зображень їх поверхонь можна також використати мультимодальні великі мовні моделі (LLM), зокрема, GPT-4v від компанії OpenAI [13].

Суттєвою перевагою GPT-4v є можливість завантажити у відповідний чат серію картинок для аналізу. Такою спроможністю серед інших відомих LLM наділена також Claude 3 Opus, до якої одночасно можна подати на розгляд 20 зображень обсягом до 5 МБ кожне. Однак, як засвідчили проведені тести, за якістю аналізу зображень Claude 3 Opus поступається GPT-4v. В інших же відомих LLM, зокрема Gemini Ultra 1.0 від Google, LLava [14] та інших, існує обмеження на завантаження одночасно тільки однієї картинки для опису, що позбавляє можливості проводити навчання подібних LLM по серії зображень.

Щоб експериментально підтвердити здатність GPT-4v класифікувати зображення відповідно до набору даних користувача, було вирішено повернутися до завдання бінарної класифікації. При цьому налаштування GPT-4v здійснювалося за допомогою системного запиту

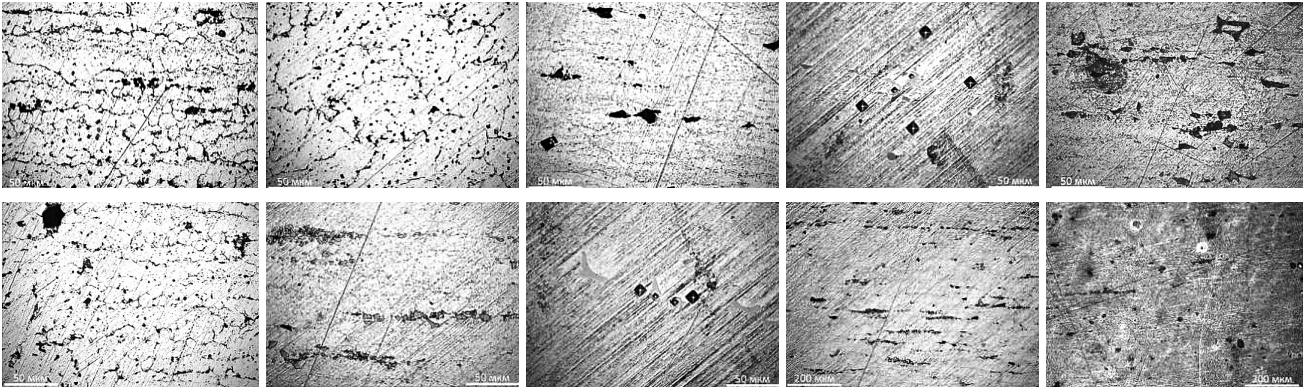


Рис. 23. Навчальна вибірка одночасно звантажених у GPT-4v зображень metal 1

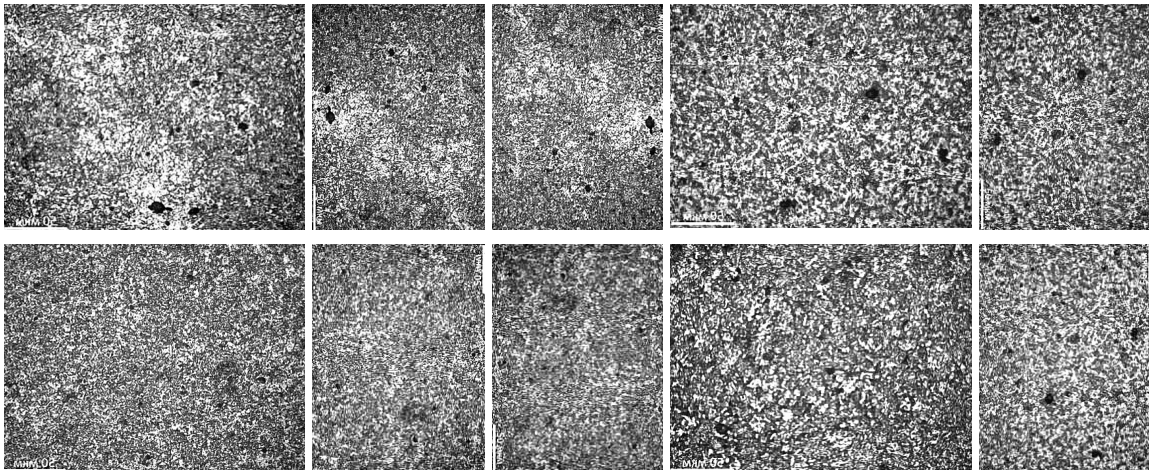


Рис. 24. Навчальна вибірка одночасно звантажених у GPT-4v зображень metal 2

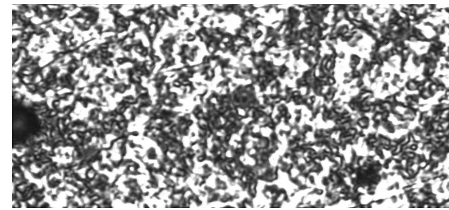
наступного змісту: «Далі буде надано кілька зображень металу 1, а потім – зображення металу 2. Необхідно вивчити усі відмінності між цими класами металів. Після цього буде надано кілька тестових зображень, щоб оцінити здатність класифікувати метали на основі зображень».

Навчальний набір зображень металу 1 (metal 1), які були одночасно звантажені в GPT4v, показано на рис. 23. Вони ілюструють різні мікроскопічні види металевих поверхонь з видимими дефектами та шорсткостями. Далі так само в GPT4v були звантажені зображення для 2-го класу металу metal 2 (рис. 24). Ці зображення показують різні мікроскопічні види металу з менш вираженою текстурою, ніж на попередніх знімках metal 1 (рис. 23), й більш однорідну структуру.

Після виконання зазначених налаштувань кульмінація експерименту полягала у зверненні до GPT-4v з проханням класифікувати невідомі зображення металів, використовуючи здобуті знання. GPT-4v було проінформовано, що він ознайомився з двома класами металевих зразків, і далі йому будуть звантажені окремі зображення, які він раніше не аналізував, щоб отримати відповідь, до якого класу вони належать.

Першим для аналізу було звантажено зображення поверхні metal 2. Його вигляд, налаштування та відповідь GPT-4v представлені на рис. 25. Слід зазначити, що GPT-4v зробив чіткий і правильний висновок, що наданий зразок належить до класу metal 2 і обґрунтував,

у чому полягає специфіка розглянутої ним металевої поверхні. Ось дослівний переклад відповіді GPT-4v на тестове зображення metal 2:



You got some pictures of metal 1 and then - pictures for metal 2. You should be study all difference between these metals classes. After this I will test you and will give few test pictures to assesment your ability to classification of metals based on pictures. Please classified what class of metal (Metal 1 or metal 2) is depicted on thus picture

ChatGPT

Based on the previously shared images of 'metal 1' and 'metal 2', I can make an informed guess about the classification of the new image.

The image you've uploaded appears to have a microstructure that is more similar to the images provided for 'metal 2'. It exhibits a fine and homogenous grain structure with a uniform phase distribution and without the clear directional texture that was present in 'metal 1'.

The consistent, fine-grained appearance and the lack of elongated grains or evidence of directional processing suggest that this is aligned with the characteristics of 'metal 2'.

Рис. 25. Налаштування та відповідь GPT-4v на тестове зображення metal 2

«На основі раніше наданих зображень «metal 1» і «metal 2» я можу зробити обґрунтоване припущення щодо класифікації нового зображення.

Зображення, яке ви завантажили, має мікроструктуру, більш схожу на зображення, надані для «metal 2». Він демонструє дрібну і однорідну зернисту структуру з рівномірним розподілом фаз і без чіткої спрямованої текстури, яка була присутня в «metal 1».

Послідовний, дрібнозернистий вигляд і відсутність подовжених зерен або доказів спрямованої обробки припускають, що це більше відповідає характеристикам «metal 2». Однак важливо зазначити, що точну класифікацію металів за допомогою мікроструктур в ідеалі має виконувати матеріалознавець, який також може врахувати інші важливі фактори, такі як хімічний склад і механічні властивості. Ця класифікація є найкращим припущенням на основі візуальної схожості з попередніми зображеннями».

Далі GPT-4v було надано картинку з metal 1 і він так само доказово та без помилки відповів, що це метал 1 (рис. 26): «Нове зображення, яке ви надали, демонструє мікроструктуру з подовженими зернами та спрямованою текстурою, подібну до зображень, які раніше ідентифікувалися як «metal 1». Лінійні особливості та наявність більш темних включень або виділень уздовж меж зерен вказують на механічну обробку, таку як прокатка чи кування, яка має тенденцію подовжувати зерна в напрямку деформації».

Ці візуальні характеристики відрізняються від більш рівновісної та однорідної мікроструктури «metal 2». Таким чином, на основі візуального порівняння з наданими раніше зображеннями, це зображення буде класифіковано як «metal 1»».

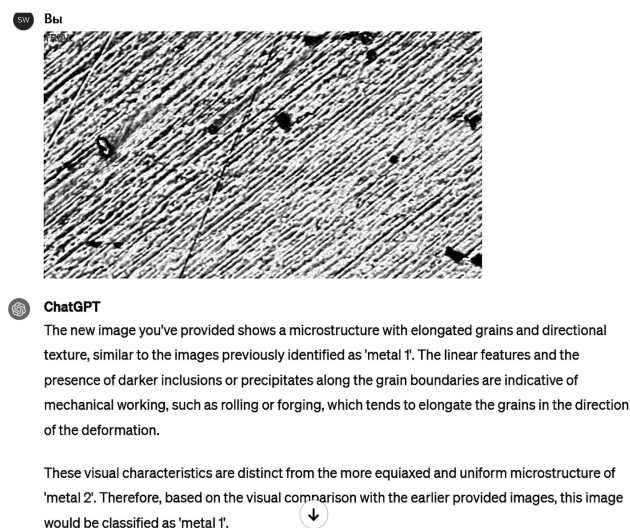


Рис. 26. Відповідь GPT-4v на зображення класу metal 1

Таким чином, GPT-4v відмінно впорався з завданням класифікації і продемонстрував свою здатність до вивчення та розпізнавання різних видів металевих мікроструктур. При цьому враховувалися загальні принципи мікроструктурного аналізу та емпіричні дані, отримані з літературних джерел і попереднього досвіду.

Продемонстрована ефективність використання великої мовної моделі для класифікації мікроструктур різних металів є новим підходом, який дозволяє не тільки ви-

значати клас металу, а й отримати доказове обґрунтування та опис результатів аналізу. Дослідники можуть завантажувати масиви зображень, і LLM використовує внутрішні алгоритми для аналізу та класифікації цих зображень згідно з навчальними даними. Цей процес охоплює вивчення текстур, зернистості, шорсткості та інших характеристик поверхні, які є критичними для розпізнавання і виявлення різниці між різними типами металів.

Суттєво, що GPT-4v не лише класифікує зображення, але й надає можливість користувачам перевіряти своє розуміння матеріалів, дозволяючи їм самостійно класифікувати зображення, а потім перевіряти свої відповіді по результатах, наданих GPT-4v. Це свідчить про потенціал LLM як навчального інструменту та їх здатність підтримувати технічні дослідження і контроль якості в індустрії металів. Ключовою перевагою є можливість автоматизації процесів і забезпечення об'єктивності класифікації завдяки машинному навчанню. Ідеальною стратегією є використання різних великих мовних моделей для доповнення їх мультимодальних спроможностей на основі переваг кожної LLM для заповнення прогалин інших.

ВИСНОВКИ

Методика, представлена в цій роботі, може бути корисною у ситуаціях, коли доступ до складніших аналітичних інструментів обмежений або неможливий. Для підвищення точності класифікації рекомендується комбінувати візуальний аналіз з іншими методами матеріалознавства, враховуючи вплив термічної обробки, легування та інших виробничих процесів на мікроструктуру металів. Крім того, вагомим внеском у цю галузь могла б стати розробка алгоритмів машинного навчання, які здатні автоматизувати процес класифікації на основі аналізу мікроструктурних зображень.

Включивши зазначені стратегії та методи в інтелектуальну систему ідентифікації конструкцій спеціального призначення на основі нейромережних технологій, можливо підвищити продуктивність і забезпечити високу точність аналізу. Завдяки цьому вдасться знизити ризики, а також зекономити час і ресурси, які б традиційно витрачалися на цей складний і відповідальний процес.

Подальші дослідження можуть включати в себе застосування LLM та нейромереж для дифракційного аналізу за допомогою рентгенівських променів (XRD) для визначення кристалічної структури матеріалів та скануючу електронну мікроскопію (SEM) з метою детальнішого вивчення їх мікроструктури.

СПИСОК ПОСИЛАНЬ

1. Слюсар В.І. Архітектурно-математичні основи удосконалення нейронних мереж з класифікації зображень. Штучний інтелект. 2022. № 1. С. 127—138. <https://doi.org/10.15407/jai2022.01.127>.
2. Slyusar, V., Protsenko, M., Chernukha, A., Gornostal, S., Rudakov, S., Shevchenko, S., Chernikov, O., Kolpachenko, N., Timofeyev, V. & Artiukh, R. (2021). Construction of an advanced method for recognizing monitored objects by a convolutional neural network using a discrete wavelet transform. Eastern-European J. of

- Enterprise Technologies. No 4(9(112)). Pp. 65—77. <https://doi.org/10.15587/1729-4061.2021.238601>.
3. Слюсар В.І., Проценко М.М., Докучаєв О.В. Методика побудови нейронних мереж для ідентифікації озброєння та військової техніки. *Озброєння та військова техніка*. 2022. № 2(34). С. 99—107. [https://doi.org/1034169/2414-0651.2022.2\(34\).99-107](https://doi.org/1034169/2414-0651.2022.2(34).99-107).
 4. Vadym Slyusar, Mykhailo Protsenko & Mariia Bondarchuk. (2022). Methodology for Armaments Identification Using a Neural Network. *Problems of Infocommunications. Science and Technology: 2022 Intern. Scient.-Pract. Conf. PIC S&T*. 5 p. <https://doi.org/10.1109/PICST57299.2022.10238568>.
 5. Slyusar, V.I. & Sliusar, I.I. (2021). Lions of Neural Networks Zoo, Neyromerezhni tehnologii ta yih zastosuvannya NMTIZ-2021: zb. nauk. pr. XX Mizhn. Nauk. konf. «Neiromerezhny tehnologii ta yih zastosuvannya NMTIZ-2021». Kramatorsk: DDMA. Pp. 129—133. <https://doi.org/10.13140/RG.2.2.17187.58405>.
 6. Qassim, H., Verma, A. & Feinzimer, D. (2018). Compressed residual-VGG16 CNN model for big data places image recognition. *Computing and Communication Workshop and Conf. (CCWC) 2018 IEEE 8th Annual*. Pp. 169—175.
 7. M. Tan & Q. Le. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks; in *Proc. of Intern. Conf. on Machine Learning*. Pp. 6105—6114.
 8. Sandler, M., Howard, A. & Zhu, M. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, 18-23 June 2018*. Pp. 4510—4520. <https://doi.org/10.1109/CVPR.2018.00474>.
 9. Howard, A., Sandler, M. & Chu, G. (2019) Searching for Mobilenetv3. *Proc. of the IEEE Intern. Conf. on Computer Vision, Seoul, 27 October – November 2019*. Pp. 1314—1324. <https://doi.org/10.1109/ICCV.2019.00140>.
 10. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conf. on computer vision and pattern recognition*. IEEE. Pp. 248—255.
 11. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. Pp. 1800—1807. <https://doi.org/10.1109/CVPR.2017.195>.
 12. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. Pp. 2818—2826. <https://doi.org/10.1109/CVPR.2016.308>.
 13. GPT-4. Technical Report by OpenAI, 27 March 2023. Available at: <https://arxiv.org/pdf/2303.08774v3.pdf>.
 14. Liu, H., Li, C., Wu, Q. & Lee, Y.J. (2023). Visual Instruction Tuning. 19 p. Available at: <https://arxiv.org/abs/2304.08485>.
- chenko, N., Timofeyev, V. & Artiukh, R. (2021). Construction of an advanced method for recognizing monitored objects by a convolutional neural network using a discrete wavelet transform. *Eastern-European J. of Enterprise Technologies*. No 4(9(112)). Pp. 65—77. <https://doi.org/10.15587/1729-4061.2021.238601>.
3. Slyusar, V.I., Protsenko, M.M. & Dokuchaev O.V. (2022). “Metodyka pobudovy neironnykh merezh dlia identyfikatsii ozbroennia ta viiskovoi tekhniki” [The technique of building neural networks for the identification of weapons and military equipment], *Ozbroiennia ta viiskova tekhnika*. No 2(34). Pp. 99—107. [https://doi.org/1034169/2414-0651.2022.2\(34\).99-107](https://doi.org/1034169/2414-0651.2022.2(34).99-107).
 4. Vadym Slyusar, Mykhailo Protsenko & Mariia Bondarchuk. (2022). Methodology for Armaments Identification Using a Neural Network. *Problems of Infocommunications. Science and Technology: 2022 Intern. Scient.-Pract. Conf. PIC S&T*. 5 p. <https://doi.org/10.1109/PICST57299.2022.10238568>.
 5. Slyusar, V.I. & Sliusar, I.I. (2021). Lions of Neural Networks Zoo, Neyromerezhni tehnologii ta yih zastosuvannya NMTIZ-2021: zb. nauk. pr. XX Mizhn. Nauk. konf. «Neiromerezhny tehnologii ta yih zastosuvannya NMTIZ-2021». Kramatorsk: DDMA. Pp. 129—133. <https://doi.org/10.13140/RG.2.2.17187.58405>.
 6. Qassim, H., Verma, A. & Feinzimer, D. (2018). Compressed residual-VGG16 CNN model for big data places image recognition. *Computing and Communication Workshop and Conf. (CCWC) 2018 IEEE 8th Annual*. Pp. 169—175.
 7. M. Tan & Q. Le. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks; in *Proc. of Intern. Conf. on Machine Learning*. Pp. 6105—6114.
 8. Sandler, M., Howard, A. & Zhu, M. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, 18-23 June 2018*. Pp. 4510—4520. <https://doi.org/10.1109/CVPR.2018.00474>.
 9. Howard, A., Sandler, M. & Chu, G. (2019) Searching for Mobilenetv3. *Proc. of the IEEE Intern. Conf. on Computer Vision, Seoul, 27 October – November 2019*. Pp. 1314—1324. <https://doi.org/10.1109/ICCV.2019.00140>.
 10. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, & Li Fei-Fei. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conf. on computer vision and pattern recognition*. IEEE. Pp. 248—255.
 11. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. Pp. 1800—1807. <https://doi.org/10.1109/CVPR.2017.195>.
 12. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. Pp. 2818—2826. <https://doi.org/10.1109/CVPR.2016.308>.
 13. GPT-4. Technical Report by OpenAI, 27 March 2023. Available at: <https://arxiv.org/pdf/2303.08774v3.pdf>.
 14. Liu, H., Li, C., Wu, Q. & Lee, Y.J. (2023). Visual Instruction Tuning. 19 p. Available at: <https://arxiv.org/abs/2304.08485>.

REFERENCES

1. Sliusar, V.I. (2022). “Architekturno-matematychni osnovy udoskonalennia neironnykh merezh z klasyfikatsii zobrazhen” [Architectural and mathematical foundations of improving neural networks for image classification], *Shtuchnyi intelekt*. No 1. Pp. 127—138. <https://doi.org/10.15407/2022.01.127>.
2. Slyusar, V., Protsenko, M., Chernukha, A., Gornostal, S., Rudakov, S., Shevchenko, S., Chernikov, O., Kolpa-