

**О. М. ШУШУРА**

доктор технічних наук, доцент,  
професор кафедри цифрових технологій в енергетиці  
Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
ORCID: 0000-0003-3200-720X

**В. В. ПРИСЯЖНЮК**

магістр кафедри цифрових технологій в енергетиці  
Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
ORCID: 0009-0002-2267-3559

## ІНФОРМАЦІЙНА СИСТЕМА КЛАСИФІКАЦІЇ ТА МАРКУВАННЯ ЗОБРАЖЕНЬ ДЛЯ НАВЧАННЯ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ

Стаття присвячена розробці інформаційної системи для класифікації та маркування зображень з метою навчання моделей штучного інтелекту. Покращення швидкості та точності класифікації і маркування зображень шляхом надання їм певних міток або категорій відкриває нові можливості для використання машинного навчання у різних сферах, включаючи комп'ютерний зір, медичну діагностику, розпізнавання образів. Проведений аналіз наявних систем анотування зображень показав, що слабкими сторонами цих технологій є неповнота та незручність реалізованих інструментів, недостатньо висока швидкість виконання операцій. У роботі запропоновано для класифікації та маркування зображень використати технології штучних нейронних мереж. З метою автоматизації класифікації зображень обрано мережу ResNet, яка навчається в процесі виконання роботи в межах одного датасету, що дозволяє скоротити витрати часу на проведення операції. Для задач маркування зображень застосовано мережу SAM, яка дає змогу узагальнювати незнайомі об'єкти та зображення без необхідності додаткового навчання. Дослідження використання цих технологій на контрольній вибірці даних показало достатньо високу точність їх роботи. Сформовано вимоги до інформаційної системи автоматизації класифікації та маркування зображень, які формалізовані у вигляді діаграми прецедентів UML, спроектована її структура та обрані засоби розробки, створене програмне забезпечення на мові Python та проведено його тестування. В якості системи управління базами даних обрано MongoDB через її безкоштовність та продуктивність. Результати досліджень можуть бути використані розробниками інформаційних технологій, що працюють в області навчання моделей штучного інтелекту.

**Ключові слова:** класифікація зображень, маркування зображень, штучні нейронні мережі, модель штучного інтелекту, інформаційна технологія, автоматизація.

**O. M. SHUSHURA**

Doctor of Technical Sciences, Associate Professor,  
Professor at the Department of Digital Technologies in Energy  
National Technical University of Ukraine  
“Igor Sikorsky Kyiv Polytechnic Institute”  
ORCID: 0000-0003-3200-720X

**V. V. PRYSYAZHNUK**

Master at the Department of Digital Technologies in Energy  
National Technical University of Ukraine  
“Igor Sikorsky Kyiv Polytechnic Institute”  
ORCID: 0009-0002-2267-3559

## INFORMATION SYSTEM FOR IMAGE CLASSIFICATION AND LABELING FOR TRAINING ARTIFICIAL INTELLIGENCE MODELS

The paper is devoted to the development of an information technology for image classification and labeling for the purpose of training artificial intelligence models. Improving the speed and accuracy of image classification and labeling by assigning them specific tags or categories opens up new possibilities for the use of machine learning in various fields, including computer vision, medical diagnostics, and image recognition. The analysis of existing image annotation systems has shown that their weaknesses lie in the incompleteness and inconvenience of implemented tools, as well as insufficiently high execution speed. The proposed approach for image classification and labeling involves the use of

*artificial neural network technologies. For the automation of image classification, the ResNet network was selected, which is trained within the framework of a single dataset, thus reducing the time required for the operation. For image labeling tasks, the SAM network was applied, which allows for generalizing unfamiliar objects and images without the need for additional training. Research on the use of these technologies on a test dataset has demonstrated their sufficiently high accuracy. Requirements for an information system for automating image classification and labeling have been formulated, which are formalized in the form of a UML use case diagram. The system's structure has been designed, and development tools have been chosen. The software has been created using the Python programming language and subjected to testing. MongoDB has been selected as the database management system due to its free-of-charge availability and productivity. The research results can be used by information technology developers working in the field of artificial intelligence model training.*

**Key words:** *image classification, image labeling, artificial neural networks, artificial intelligence model, information technology, automation.*

### Постановка проблеми

Класифікація та маркування зображень використовуються при навчанні моделей штучного інтелекту, які застосовуються у різних сферах, включаючи комп'ютерний зір, медичну діагностику, розпізнавання образів, та ін. Сьогодні значна кількість компаній використовує штучний інтелект для прискорення виконання багатьох завдань. Необхідність обробки величезних обсягів інформації для навчання моделей штучного інтелекту, яка здебільшого виконується персоналом компаній, визначає актуальність розробки інформаційних технологій з метою автоматизації цього процесу. Очікується, що до 2026 року ринок маркування даних зросте до 5,5 мільярдів доларів США [1].

Автоматизація класифікації та маркування зображень (анотування) для навчання моделей штучного інтелекту є процесом використання різноманітних технологій, методів та алгоритмів для автоматизованого аналізу зображень та надання їм певних міток або категорій. Це допомагає в підготовці даних для навчання моделей машинного навчання, що забезпечує більш точну класифікацію зображень у майбутньому.

### Аналіз останніх досліджень і публікацій

Анотовані дані є джерелом інформації для моделей машинного навчання, оскільки продуктивність і точність таких моделей залежать від якості та кількості анотованих даних. Моделі машинного навчання мають широкий спектр критично важливих застосувань, в тому числі ті (наприклад, охорона здоров'я), де їх помилки можуть бути небезпечними. Наприклад, застосування методів машинного навчання в патології значно поліпшує виявлення метастазів у лімфовузлах, оцінку показника Ki67 у раку грудей, рейтинг Глісона у раку простати та оцінку лімфоцитів, що проникають у пухлину (TIL) у меланомі [2, 3].

Для анотування даних застосовуються спеціалізовані інформаційні системи Computer Vision Annotation Tool (CVAT) та Label Studio (LS). Система CVAT пропонує ряд інструментів для створення анотацій, які допомагають у процесі маркування для подальшого навчання штучного інтелекту [4]. Програмне забезпечення LS пропонує різні типи анотацій, включаючи обмежувальні рамки, полігони, полілінії, ключові точки і маски семантичної сегментації [5]. Основним недоліком вказаних платформ є обмеженість інструментів для тих типів задач, де потрібно об'єднати класифікацію та маркування зображень.

Для класифікації та аналізу зображень застосовуються глибокі нейронні мережі [6], в тому числі з використанням переднього навчання (pre-training) на великих наборах даних для досягнення навчання за допомогою методів transfer learning на даних з обмеженим обсягом [7]. В процесах маркування зображень використовується мультимодальне активне навчання з глибоким навчанням та підкріпленням для виділення цільових ознак [8]. Широке розповсюдження отримало застосування нейронних мереж ConvNet, ResNet та DenseNet [9].

Вибір між DenseNet, ConvNet та ResNet залежить від конкретної задачі, набору даних та обчислювальних ресурсів. Щільні зв'язки DenseNet вимагають, щоб кожен шар отримував прямі вхідні дані від усіх попередніх шарів, що може призвести до збільшення використання пам'яті порівняно з традиційними архітектурами, такими як ResNet. Мережі ConvNets обробляють локальні особливості за допомогою згорткових фільтрів, які є ефективними для захоплення локальних шаблонів. Однак вони можуть не мати повного розуміння глобального контексту, особливо в задачах, де розуміння ширшого контексту має вирішальне значення. В даному дослідженні було обрано мережу ResNet через пом'якшення проблеми зникаючого градієнту в дуже глибоких мережах та використання меншої кількості пам'яті та обчислювальної потужності для навчання моделі [10].

### Формулювання мети дослідження

Метою даної роботи є розробка інформаційної технології класифікації та маркування зображень для навчання моделей штучного інтелекту. Для прискорення процесу класифікації і маркування, зменшення частки ручної праці під час обробки масивів навчальних даних слід розглянути можливість застосування в інформаційній технології штучних нейронних мереж. Необхідно визначити загальні та функціональні вимоги до інформаційної системи автоматизації класифікації та маркування зображень, сформувати її структуру та обрати засоби розробки, розробити програмне забезпечення та провести його тестування.

### Викладення основного матеріалу дослідження

Анотування є це важливою задачею, яка вимагає надійної, швидкої та ефективної платформи для роботи з даними в різній кількості та з використанням різних інструментів. Класифікація вихідного масиву зображень необхідна для його сортування по навчальним наборам.

Для автоматизації класифікації та підвищення продуктивності роботи персоналу, що здійснює обробку зображень, запропоновано використовувати штучну нейронну мережу Residual neural network (ResNet), яка є глибокою нейронною мережею, що має особливу архітектуру, яка дозволяє досягати кращих результатів у задачах класифікації зображень за рахунок можливості навчання в процесі виконання роботи в межах одного датасету. Архітектура ResNet передбачає використання блоків, що названі "residual blocks". Ці блоки містять звичайні шари нейронної мережі, але з деяким додатковим з'єднанням, яке дозволяє зберігати вхідну інформацію та додавати її до виходу блоку. Обрана модель залишкової нейронної мережі з тришаровими пропусками, що містять пакетну нормалізацію та нелінійність між ними [10]. Для навчання мережі було використано 50000 фото. Перед проведенням навчання цей масив було розділено на підмасиви Train, Test, та Validation, у відношенні 10%, 20% та 70% відповідно. Результати навчання моделі показано на рисунку 1.

```
Epoch 1/6127
43/43 [=====] - 17s 367ms/step - loss: 0.6303 - val_loss: 0.6133
Epoch 2/6127
43/43 [=====] - 17s 301ms/step - loss: 0.6235 - val_loss: 0.6055
Epoch 3/6127
43/43 [=====] - 17s 317ms/step - loss: 0.6171 - val_loss: 0.6011
Epoch 6125/6127
43/43 [=====] - 17s 322ms/step - loss: 0.0575 - val_loss: 0.0742
Epoch 6126/6127
43/43 [=====] - 17s 325ms/step - loss: 0.0506 - val_loss: 0.0732
Epoch 6127/6127
43/43 [=====] - 17s 331ms/step - loss: 0.0503 - val_loss: 0.0727
```

Рис. 1. Результат навчання моделі ResNet для автоматизації класифікації зображень

Як видно з рисунку 1, точність класифікації складає приблизно 93%, що свідчить про можливість практичного використання моделі. З метою подальшого покращення якості роботи мережі можна збільшити датасет та кількість епох навчання моделі.

Для автоматизації маркування даних запропоновано використовувати нейронну мережу Segment Anything Model (SAM), що є швидкою системою сегментації, яка дозволяє узагальнювати незнайомі об'єкти (в тому числі зображення) без додаткового навчання [11].

На рисунку 2 представлено приклад роботи моделі SAM, розділеної на кодер одноразових зображень і декодер масок, який може працювати в браузері всього за кілька мілісекунд на запит. Архітектурою, що використовується в моделі для сегментації об'єктів, є Convolutional Neural Network (CNN), які складаються з багатьох шарів, що вивчають ієрархічні представлення вихідних даних. Вони складаються зі шарів для вилучення ознак даних і шарів підвищеної вибірки, призначених для генерації сегментованого результату. Для навчання моделі сегментації потрібен достатньо великий набір зображень, позначених анотаціями на рівні пікселів, де кожному пікселю присвоюється мітка класу, що вказує на об'єкт, до якого він належить. Під час навчання модель вчиться зіставляти вхідні зображення з масками сегментації на рівні пікселів.

Процес навчання включає оптимізацію параметрів моделі для мінімізації різниці між прогнозованими масками сегментації та істинними анотаціями. Після того, як модель навчена, її можна використовувати для сегментації об'єктів на нових зображеннях, що не були використані при навчанні. Під час виведення вхідне зображення пропускається через навчену модель, яка створює карту ймовірностей, що вказує на ймовірність належності кожного пікселя до певного класу. В деяких випадках результат сегментації може піддаватися обробці методами постобробки з метою уточнення результатів. Це можуть бути такі методи, як морфологічні операції (наприклад, розширення) для видалення шуму або згладжування для покращення меж сегментації.

Ефективність моделі "Segment Anything Model" залежить від різних факторів, зокрема від якості та різноманітності навчального набору даних, а також від застосованих методів постобробки.

Для автоматизації класифікації та маркування зображень на основі зазначених штучних нейронних мереж необхідно розробити інформаційну систему. На рисунку 3 представлено основні функціональні можливості системи у вигляді діаграми прецедентів UML.

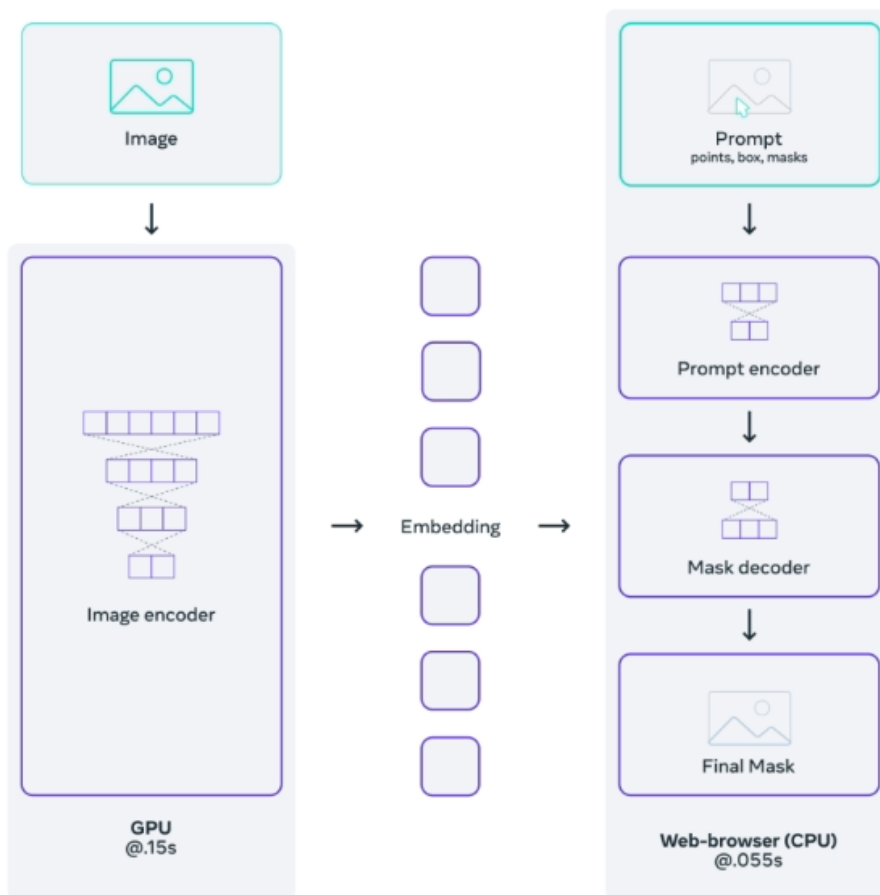


Рис. 2. Приклад роботи моделі SAM

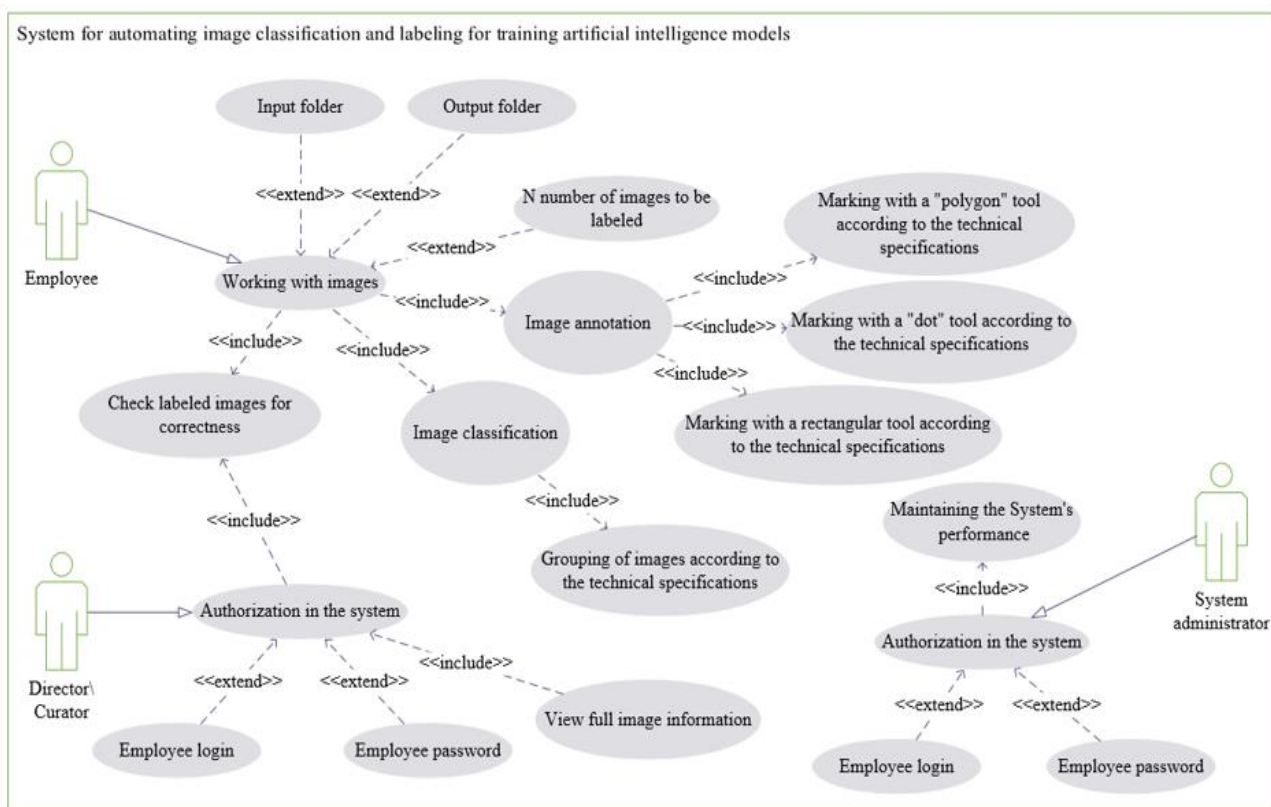


Рис. 3. UML діаграма прецедентів інформаційної системи класифікації та маркування зображень

Як видно з рисунку 3, у системи виділено 3 ролі користувачів, а саме «працівник», який має доступ до маркування та класифікації зображень, а також до їх перевірки на коректність, «системний адміністратор», функцію якого є підтримка працездатності системи, та «директор/куратор», що має доступ до перевірки даних на коректність та перегляду повної інформації зображень.

До структури інформаційної системи класифікації та маркування зображень входить модуль візуальної компоненти програми, який складається з декількох вікон для взаємодії з користувачем, та бази даних зображень, які містять вичерпну інформацію про фотографії та відповідні їм мітки. В якості системи управління базами даних було обрано MongoDB через її безкоштовність та продуктивність, що є важливим показником для компаній, які не мають великого бюджету для розгортання подібних систем. Програмне забезпечення системи розроблене на мові Python, яка має необхідні бібліотеки.

### Висновки

В даній роботі представлено інформаційну технологію класифікації та маркування зображень для навчання моделей штучного інтелекту. За рахунок використання наведених в роботі штучних нейронних мереж для класифікації та сегментації зображень скорочується час обробки вихідних масивів даних при підготовці навчальних наборів зображень. Об'єднання цих інструментів в єдину інформаційну систему дозволить компаніям, які спеціалізуються на ануванні даних, підвищити продуктивність виробничих процесів. Результати досліджень можуть бути використані розробниками інформаційних технологій, що працюють в області навчання моделей штучного інтелекту.

### Список використаної літератури

1. Data Labeling: How to Choose a Data Labeling Partner in 2023 [Електронний ресурс]. Режим доступу до ресурсу: <https://research.aimultiple.com/data-labeling/#why-is-it-important-now>
2. Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. *J Intern Med*. 2020. Vol. 288. P. 62–81.
3. Gulbahar Karatas. Data Annotation in 2023: Why it matters & Top 8 Best Practices. *AIMultiple* [Електронний ресурс]. Режим доступу до ресурсу: <https://research.aimultiple.com/data-annotation/>.
4. Документація Computer Vision Annotation Tool (CVAT) [Електронний ресурс]. Режим доступу до ресурсу: <https://github.com/opencv/cvat>.
5. Документація Label studio(LS) [Електронний ресурс]. Режим доступу до ресурсу: <https://labelstud.io/guide/>.
6. Abdou, M.A. Literature review: efficient deep neural networks techniques for medical image analysis. *Neural Comput & Applic*. 2022. Vol. 34. P. 5791–5812. <https://doi.org/10.1007/s00521-022-06960-9>.
7. Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. *Sustainability*. 2023. Vol. 15, 1906. <https://doi.org/10.3390/su15031906>.
8. Dhiman, G., Kumar, A.V., Nirmalan, R. et al. Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications. *Multimed Tools Appl* Vol. 82. P. 5343–5367. <https://doi.org/10.1007/s11042-022-12178-7>.
9. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger. Densely Connected Convolutional Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>.
10. Огляд ResNet та його варіантів [Електронний ресурс] – Режим доступу до ресурсу: <https://datascience.eu/machine-learning/an-overview-of-resnet-and-its-variants/>.
11. Офіційний Сайт Segment Anything Model (SAM) [Електронний ресурс]. Режим доступу до ресурсу: <https://segment-anything.com/>.

### References

1. Data Labeling: How to Choose a Data Labeling Partner in 2023. Retrieved from <https://research.aimultiple.com/data-labeling/#why-is-it-important-now>.
2. Acs B, Rantalainen M, Hartman J. (2020). Artificial intelligence as the next step towards precision pathology. *Journal of Internal Medicine*, 288, 62-81.
3. Gulbahar Karatas. (2023). Data Annotation in 2023: Why it matters & Top 8 Best Practices. *AIMultiple*. Retrieved from <https://research.aimultiple.com/data-annotation/>.
4. Computer Vision Annotation Tool (CVAT) Documentation. Retrieved from <https://github.com/opencv/cvat>.
5. Label studio (LS) Documentation. Retrieved from <https://labelstud.io/guide/>.
6. Abdou, M.A. (2022). Literature review: efficient deep neural networks techniques for medical image analysis. *Neural Computing & Applications*, 34, 5791-5812. <https://doi.org/10.1007/s00521-022-06960-9>.

7. Gulzar, Y. (2023). Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. *Sustainability*, 15, 1906. <https://doi.org/10.3390/su15031906>.
8. Dhiman, G., Kumar, A.V., Nirmalan, R. et al. (2022). Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications. *Multimedia Tools and Applications*, 82, 5343-5367. <https://doi.org/10.1007/s11042-022-12178-7>.
9. G. Huang, Z. Liu, L. Van Der Maaten, & K. Q. Weinberger. (2017). Densely Connected Convolutional Networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, pp. 2261-2269. <https://doi.org/10.1109/CVPR.2017.243>.
10. Overview of ResNet and Its Variants. Retrieved from <https://datascience.eu/machine-learning/an-overview-of-resnet-and-its-variants/>.
11. Official Website of Segment Anything Model (SAM). Retrieved from <https://segment-anything.com/>.