

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Міністерство освіти і науки України

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису
УДК 004.8, 004.93

ОХРІМЕНКО АНТОН ОЛЕКСАНДРОВИЧ

ДИСЕРТАЦІЯ

**МОДЕЛІ ГЛИБИННОГО НАВЧАННЯ НА СКЛАДНИХ ДЛЯ
РОЗПІЗНАВАННЯ НАБОРАХ ДАНИХ**

113 – Прикладна математика

11 – Математика та статистика

Подається на здобуття наукового ступеня доктора філософії.

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

_____ Охріменко А. О.

Науковий керівник

Куcssуль Наталія Миколаївна, доктор технічних наук, професор.

АНОТАЦІЯ

Охріменко А.О. Моделі глибинного навчання на складних для розпізнавання наборах даних. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 113 «Прикладна математика» (галузь знань 11 «Математика та Статистика»). – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, 2024.

Останнім часом широко використовуються методи машинного навчання для вирішення різноманітних практичних завдань. Дані методи найкраще працюють, якщо різні класи рівномірно представлені у навчальній вибірці. На практиці, реальні набори даних часто не відповідають цій умові, класи у них можуть бути незбалансованими та перекриватися за своїми ознаками. Особливо це проявляється у задачі сегментації, де на кожному зображенні з навчальної вибірки може бути присутні одразу декілька класів у різній пропорції. Як наслідок, модель машинного навчання погано розпізнає найменш представлені класи.

Метою роботи є розробка нових ефективних методів навчання моделей для задачі сегментації при наявній проблемі незбалансованості та перекриття класів у навчальному наборі даних.

Існують стандартні засоби, що застосовуються для балансування вибірок, такі як збільшення ваги малих класів у функції втрат, чи збалансоване формування навчального пакету. Науковці розробляють нові методи, такі як особлива обробка даних в областях перекриття класів, генерація штучних екземплярів даних та динамічні вагові коефіцієнти. Дані підходи не є універсальними та не враховують особливості самих даних. Зокрема, не враховується роздільність представників різних класів та ступінь повноти ознак у наборі даних.

Тому актуальною є розробка нових методів, що враховують властивості даних, а також нових методів балансування навчальних наборів.

Перший розділ дисертаційного дослідження присвячений огляду робіт інших дослідників, що присвячені проблемі незбалансованості, перекриття класів та випадку їх комбінації. Проведено аналіз даних методів, окреслено основні групи поточних досліджень, визначено недоліки кожної групи, сформульовано завдання дисертаційного дослідження. Окрім того, розглянуто підходи, що специфічні для предметних областей даного дослідження.

У *другому* розділі досліджується задача визначення неоднозначних екземплярів даних, які важко коректно віднести до правильного класу через подібність їхніх ознак до представників іншого класу, тобто наявна проблема перекриття класів. Запропоновано метод виявлення неоднозначних екземплярів даних серед навчальної вибірки. Описаний метод апробовано на штучно згенерованому наборі даних.

У *третьому* розділі запропоновано метод вагових масок для покращення розрізнення малих класів у задачі семантичної сегментації на прикладі визначення сільськогосподарських культур на супутникових знімках. Модифіковано функцію втрат до використання вагових масок, що згенеровані відповідно до належності різних пікселів зображення до неоднозначних екземплярів даних. Показано, що у порівнянні з базовою сегментаційною моделлю, використання вагових масок значно покращує метрики малих класів, а також спостерігається незначний приріст загальних метрик.

У *четвертому* розділі досліджується можливість використання генеративних змагальних мереж (GAN) для генерації штучних зображень та балансування навчального набору даних для багатоканальних супутникових знімків. Даний підхід порівнюється з іншими методами генерації штучних зображень, а також перевіряється можливість застосування методу вагових масок з попереднього розділу. Показано, що у порівнянні з базовою сегментаційною моделлю, використання GAN суттєво покращує якість роботи моделі. Окрім того, продемонстровано, що інші методи генерації штучних зображень показують гірші результати у порівнян-

ні з запропонованим методом. Використання методу вагових масок у комбінації з GAN-аугментацією також покращує метрики моделі.

У *п'ятому* розділі розроблений метод вагових масок апробований в іншій предметній області, а саме у задачі семантичної сегментації гістологічних зображень. Запропоновано модифікацію методу виявлення неоднозначних екземплярів даних, спрямовану на врахування деяких просторових властивостей зображення. Це дозволило згенерувати більш досконалі вагові маски. Підтверджено ефективність методу вагових масок, що значно покращує метрики малих класів, у порівнянні з базовою сегментаційною моделлю.

Наукова новизна отриманих результатів:

1. Вперше розроблено метод виявлення неоднозначних даних на основі KNN, що дозволяє ефективно ідентифікувати екземпляри даних, які викликають труднощі через подібність між класами.
2. Вперше запропоновано метод аугментації супутникових зображень з використанням генеративних змагальних мереж (GAN), що забезпечує генерацію нових зразків для компенсації незбалансованості класів.
3. Запропоновано нові метрики для оцінки якості датасетів, які враховують незбалансованість та перекриття класів, що дозволяє точніше аналізувати якість навчальних даних.
4. Удосконалено функції втрат через введення вагових масок, що дозволяє адаптувати процес навчання сегментаційних моделей до умов незбалансованих даних.
5. Отримав подальший розвиток метод вагових масок для компенсації незбалансованості класів, що дозволяє коригувати навчання без втручання в розмір чи структуру датасету.

Практичне значення отриманих результатів. Результати, отримані у даній роботі, застосовані в наступних проектах:

1. Науковий проект програми Horizon Europe «Satellites for Wilderness Inspec-

tion and Forest Threat Tracking» (SWIFTT) (грантова угода 101082732, 2022-2025).

2. Науковий проект НФД України «Методи і моделі глибинного навчання для прикладних задач супутникового моніторингу» (2020.02/0292, 2020-2023).
3. Науково-технічна робота (НТР) «Інформаційні технології геопросторового аналізу розвитку сільських територій і громад» (Договір РН/27-2023 від 25 травня 2023 р.) за рахунок зовнішнього інструменту допомоги Європейського Союзу для виконання зобов'язань України у Рамковій програмі Європейського Союзу з наукових досліджень і інновацій «Горизонт 2020».

Розроблені методи практично застосовані в різних областях машинного навчання, передусім для моделей, що використовують функцію втрат для свого навчання. Окрім того, GAN-аугментація може бути використана для більш широкого класу моделей, а метод визначення частки неоднозначних даних та його майбутні модифікації можуть бути використані для оцінки якості навчальних даних для різних видів задач машинного навчання.

За матеріалами дисертації опубліковано 6 робіт, з яких 2 статті у міжнародних журналах, що індексуються у Scopus (квартилі Q2 та Q3), 1 стаття в журналі, що входить до переліку фахових видань за спеціальністю дисертації, 1 одноосібний розділ у монографії та 2 публікації у матеріалах міжнародних конференцій.

Ключові слова: машинне навчання, глибинне навчання, нейронна мережа, алгоритм k-найближчих сусідів, оцінка якості датасету, перекриття класів, незбалансованість класів, сегментація, генеративна змагальна мережа, генерація навчальних даних, генеративний штучний інтелект, аугментація, функція втрат, міра неконформності.

Список публікацій здобувача

1. L. Shumilo, A. Okhrimenko, N. Kussul, S. Drozd, and O. Shkalikov, “Generative adversarial network augmentation for solving the training data imbalance problem in crop classification,” *Remote Sensing Letters*, vol. 14, pp. 1131–1140, 2023. [Scopus, Q2]
2. A. Okhrimenko and N. Kussul, “Using weight reliability masks on imbalanced datasets for satellite image segmentation,” *Cybernetics and Systems Analysis*, vol. 60, pp. 700–710, 9 2024. [Scopus, Q3]
3. A. Okhrimenko and N. Kussul, “Data mining of machine learning datasets for hard case identification,” *International Scientific Technical Journal «Problems of Control and Informatics»*, vol. 68, pp. 84–95, 5 2023.
4. А. О. Охріменко, “Методи сегментації зображень різної природи для випадку незбалансованого набору даних,” in *Методи комп’ютерного зору і глибинних нейронних мереж для еколого-економічного аналізу*, pp. 143–171, Наукова думка, 2024.
5. A. Okhrimenko and N. Kussul, “Knn-based algorithm of hard case detection in datasets for classification,” in *Proceedings of the 11th International Conference on Applied Innovations in IT*, vol. 11, Anhalt University of Applied Sciences, 3 2023. [Scopus]
6. A. Okhrimenko and N. Kussul, “Unbalanced datasets management for the problem of segmentation of satellite images,” in *2024 IEEE Fourth International Conference on System Analysis Intelligent Computing (SAIC)*, 2024.

ABSTRACT

Okhrimenko A.O. Deep learning models for handling handcase data. – Qualifying scientific work, the manuscript.

PhD thesis in the in speciality 113 Applied Mathematics (field of knowledge 11 “Mathematics and Statistics”). — National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, 2024.

Recently, machine learning methods have been widely used to solve various practical problems. These methods perform best if different classes are evenly represented in the training dataset. However, real-world datasets often do not meet this condition, as classes may be imbalanced and overlap in the feature space. This issue is especially evident in segmentation tasks, where multiple classes may be present in varying proportions within each training image. As a result, machine learning models tend to perform poorly in recognizing underrepresented classes.

The **aim of this study** is to develop novel, effective training methods for segmentation models in the presence of class imbalance and overlapping in the training dataset.

There are standard techniques for balancing datasets, such as increasing the weight of underrepresented classes in the loss function or using balanced batch sampling. Researchers are developing new methods, including specialized data processing in class-overlapping regions, synthetic data generation, and dynamic weighting coefficients. However, these approaches are not universal and do not take into account the specific characteristics of the data. In particular, they often fail to consider the separability of different classes and the completeness of feature representation in the dataset.

Therefore, the development of new methods that take data properties into account, as well as novel approaches for balancing training datasets, remains a relevant research field.

The *first* chapter of this thesis is dedicated to reviewing previous research on

the issues of class imbalance, class overlapping, and their combination. An analysis of existing methods has been conducted, key research directions have been outlined, and the limitations of each approach have been identified. Based on this analysis, the research objectives of the dissertation have been formulated. Additionally, domain-specific approaches relevant to this study have been considered.

The *second* chapter explores the problem of ambiguous data instance detection. Such instances are difficult to correctly assign a class label due to their feature similarity with instances from other classes, i.e., the issue of class overlap. A method for detecting ambiguous data instances within the training dataset is proposed. The described method has been tested on an artificially generated dataset.

In the *third* chapter, a weighted mask method is proposed to improve the distinction of small classes in the semantic segmentation problem using the example of crop identification in satellite images. The loss function is modified to use weight masks, which are generated based on the assignment of different image pixels to ambiguous data instances. The results show that, compared to the baseline segmentation model, the use of weighted masks significantly improves the metrics of small classes while also presenting a slight increase in overall performance metrics.

The *fourth* chapter explores the possibility of using Generative Adversarial Networks (GAN) for generating synthetic images and balancing the training dataset for multichannel satellite images. This approach is compared with other methods of synthetic image generation, and the applicability of the weight mask method from the previous chapter is also tested. The results show that, compared to the baseline segmentation model, the use of GAN significantly improves the model quality. Additionally, it is demonstrated that other synthetic image generation methods yield poorer results compared to the proposed approach. Combining the weight mask method with GAN augmentation also enhances the model metrics.

In the *fifth* chapter, the developed weight mask method is tested in different domains, specifically in the task of semantic segmentation of histological images. A mod-

ification of the ambiguous data instance detection method is proposed, aimed at considering some spatial properties of the image. This modification allowed for the generation of more advanced weight masks. The effectiveness of the weight mask method is confirmed, significantly improving the metrics for small classes compared to the baseline segmentation model.

The **scientific novelty** of the obtained results:

1. For the first time, a method for detecting ambiguous data based on KNN is developed, which effectively identifies data instances that cause difficulties due to class similarity.
2. For the first time, a satellite image augmentation method using Generative Adversarial Networks (GAN) is proposed, enabling the generation of new samples to compensate the class imbalance.
3. New metrics for assessing dataset quality are proposed, which take into account a class imbalance and overlapping, which allows for a more accurate analysis of training data quality.
4. The loss functions have been improved by introducing weight masks, which adapt the training process of segmentation models to unbalanced data conditions.
5. The weight mask method for compensating class imbalance has been further developed, enabling training adjustments without altering the size or structure of the dataset.

The practical significance of the obtained results. The results obtained in this study have been applied in the following projects:

1. Scientific project of the program Horizon Europe «Satellites for Wilderness Inspection and Forest Threat Tracking» (SWIFTT) (grant agreement 101082732, 2022-2025).
2. Scientific project of the NRF of Ukraine «Deep learning methods and models for applied problems of satellite monitoring» (2020.02/0292, 2020-2023).
3. Scientific and technical project «Information technologies for geospatial analy-

sis of rural and community development» (Contract PH/27-2023 dated May 25, 2023) financed by the external aid instrument of the European Union for the fulfillment of Ukraine's obligations in the European Union Framework Program for Scientific Research and Innovation «Horizon 2020».

The developed methods have practical applications in various areas of machine learning, primarily for models that use a loss function for training. Additionally, GAN augmentation can be used with a broader class of models, and the method for determining the proportion of ambiguous data instances and its future modifications can be used to assess the quality of training data for various types of machine learning tasks.

Based on the materials of the thesis, 6 works have been published, including 2 articles in international journals indexed in Scopus (Q2 and Q3 quartiles), 1 article in a journal included in the list of professional publications for the thesis's specialty, 1 individual chapter in a monograph, and 2 publications in the proceedings of international conferences.

Key words: machine learning, deep learning, neural network, k-nearest neighbors algorithm, dataset quality assessment, class overlapping, class imbalance, segmentation, generative adversarial network, training data generation, generative artificial intelligence, augmentation, loss function, measure of nonconformity.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	13
Вступ	14
1 Огляд літератури за тематикою дисертаційного дослідження	20
1.1 Проблема дисбалансу та перекриття класів у задачі класифікації. . .	21
1.2 Вагові коефіцієнти	23
1.3 Огляд предметних областей	25
Висновки до розділу 1	30
2 Метод виявлення складних для розпізнавання зразків в наборах даних для задач класифікації у машинному навчанні	33
2.1 Алгоритм визначення неоднозначних екземплярів даних у вибірках для задач машинного навчання.	35
2.2 Приклад роботи алгоритму на штучно згенерованих даних	41
2.3 Можливості практичного застосування.	44
Висновки до розділу 2	47
3 Використання вагових масок для покращення якості моделей сегментації	49
3.1 Метод вагових масок.	50
3.2 Опис датасету для задачі сегментації супутникових знімків	52
3.3 Тренування сегментаційної моделі	61
3.4 Результати та обговорення	63
Висновки до розділу 3	65
4 Застосування GAN-аугментації у області дистанційного спостереження .	67
4.1 Генеративна змагальна мережа.	68
4.2 Генерація штучного датасету.	71
4.3 Тренування сегментаційної моделі з GAN-аугментацією	74
4.4 Результати порівняння різних методів аугментації та обговорення . .	75
4.5 Використання вагових масок для покращення GAN-аугментації . . .	78

4.6	Результати використання методу вагових масок з GAN аугментацією та обговорення	80
	Висновки до розділу 4	83
5	Семантична сегментація гістологічних зображень	85
5.1	Використані дані та їхня попередня обробка	86
5.2	Покращення методу вагових масок.	89
5.3	Чисельні експерименти з сегментації гістологічних зображень. . . .	93
5.4	Результати та обговорення	97
5.5	Післяопераційна обробка сегментаційних масок	99
	Висновки до розділу 5	104
	Висновки	105
	Список використаних джерел	108

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AAC середня точність (Average Accuracy).

BCE бінарна крос-ентропія (Binary Cross Entropy).

CNN згорткова нейронна мережа (Convolutional Neural Network).

Conv згортковий шар нейронної мережі.

GAN генеративна змагальна мережа (Generative Adversarial Network).

GT істинні значення, ідеальний очікуваний результат (Ground Truth).

IoU перетин через об'єднання (Intersection Over Union).

KNN метод k-найближчих сусідів (K-Nearest Neighbors).

MSE середньоквадратична похибка (Mean Squared Error).

OA загальна точність (Overall Accuracy).

P влучність (Precision).

R повнота (Recall).

ВСТУП

Актуальність роботи. Останнім часом широко використовуються методи машинного навчання для вирішення різноманітних практичних завдань. Особливе місце серед них посідають методи глибинного навчання, що зазвичай показують кращі результати за рахунок більшої складності моделей, але натомість потребують більше даних для навчання. Окрім того, важлива якість навчальних даних. Умовами оптимальної роботи даних методів є рівномірний розподіл різних класів у навчальній вибірці, а також достатньо роздільні ознаки екземплярів різних класів. Однак у багатьох практичних задачах, реальні набори даних часто не відповідають цим умовам, класи у них можуть бути незбалансованими та перекриватися за своїми ознаками.

Незбалансованістю класів у навчальній вибірці є різна частка екземплярів різних класів у ній. Нехай N – розмір навчальної вибірки, кількість екземплярів даних, а C – кількість класів у ній. Ідеально збалансована навчальна вибірка міститиме приблизно $\frac{N}{C}$ представників кожного класу. Значні відхилення від цього значення хоча б для частини класів є ознакою незбалансованості навчальної вибірки. Якщо кількість представників деякого класу є суттєво меншим за $\frac{N}{C}$, то такий клас є недостатньо представленим у вибірці, його також можна назвати малим класом або менш представленим класом. Якщо ж ситуація протилежна і число представників іншого класу суттєво перевищує $\frac{N}{C}$, то такий клас можна назвати великим класом або більш представленим класом. Незбалансована навчальна вибірка містить в собі як і менш представлені класи, так і більш представлені класи.

Перекриттям класів є близькість представників різних класів у просторі ознак, тобто такі екземпляри даних мають подібні ознаки або ж навіть майже ідентичні. Такі екземпляри даних можна назвати складними для розпізнавання (hard-case), неоднозначними або ж ненадійним, оскільки модель машинного навчання,

якою б складною вона не була, ймовірно помилиться при класифікації таких екземплярів даних та не зможе коректно розділити у просторі ознак такі екземпляри.

У випадку незбалансованості навчальної вибірки, більш представлені класи більше впливають на параметри моделі, що навчається, а менш представлені, відповідно, меншою мірою. Як наслідок, навчена модель краще визначає великі класи, та показує гірші результати для малих класів. При наявності проблеми перекриття класів, частина даних визначається некоректно. Поєднання цих двох проблем призводить до систематичної помилки при визначенні малих класів або ж навіть повної нездатності моделі розпізнати найменш представлені класи.

Особливо це проявляється у задачах сегментації, де на кожному зображенні з навчальної вибірки може бути присутні одразу декілька класів у різній пропорції. Отримання ж нових даних, які мають інший розподіл класів та/або більшу кількість ознак, може бути значно ускладненим, економічно недоцільним або ж неможливим. У якості ознак в такому випадку виступають значення інтенсивності сигналу індивідуальних пікселів у відповідному каналі зображення, а їх загальне число є рівним числу каналів у зображенні.

Проблеми незбалансованості та перекриття класів часто зустрічаються у багатьох реальних наборах даних, тому актуальною є розробка нових моделей глибокого навчання, що враховують наявність вищеописаних проблем та не потребують додаткового розширення навчального набору даних.

Мета та задачі дослідження. Метою роботи є розробка нових ефективних методів навчання моделей для задачі сегментації при наявній проблемі незбалансованості та перекриття класів у навчальному наборі даних.

Для виконання поставленої мети необхідно вирішити наступні задачі:

1. Визначити ефективні методи виявлення неоднозначних екземплярів даних та дослідити їх вплив на якість моделей машинного навчання в задачах сегментації.
2. Розробити алгоритм виявлення неоднозначних екземплярів даних на основі

методу k-найближчих сусідів (KNN), щоб зменшити кількість помилкових класифікацій у випадках перекриття класів.

3. Запропонувати метод аугментації супутникових зображень з використанням генеративних змагальних мереж (GAN), щоб ефективно вирішити проблему незбалансованості класів.
4. Розробити нові метрики для оцінки якості датасетів, щоб забезпечити врахування дисбалансу та перекриття класів у задачах сегментації.
5. Модифікувати функцію втрат з використанням вагових масок, щоб покращити навчання моделей сегментації в умовах незбалансованих даних.
6. Розробити метод вагових масок для компенсації незбалансованості класів, щоб підвищити точність класифікації малопоширених класів без розширення датасету.

Об'єктом дослідження є проблеми незбалансованості та перекриття класів у навчальній вибірці.

Предметом дослідження є моделі та методи, що покращують якість розпізнавання різних класів на незбалансованому та складному для розпізнавання навчальному наборі даних.

Методи дослідження: аналіз, моделювання, експеримент, статистичний аналіз, обробка зображень, синтез зображень, морфологічні перетворення, методи машинного та глибинного навчання.

Наукова новизна отриманих результатів:

1. Вперше розроблено метод виявлення неоднозначних даних на основі KNN, що дозволяє ефективно ідентифікувати екземпляри даних, які викликають труднощі через подібність між класами.
2. Вперше запропоновано метод аугментації супутникових зображень з використанням генеративних змагальних мереж (GAN), що забезпечує генерацію нових зразків для компенсації незбалансованості класів.
3. Запропоновано нові метрики для оцінки якості датасетів, які враховують

незбалансованість та перекриття класів, що дозволяє точніше аналізувати якість навчальних даних.

4. Удосконалено функції втрат через введення вагових масок, що дозволяє адаптувати процес навчання сегментаційних моделей до умов незбалансованих даних.
5. Отримав подальший розвиток метод вагових масок для компенсації незбалансованості класів, що дозволяє коригувати навчання без втручання в розмір чи структуру датасету.

Практичне значення отриманих результатів. Результати, отримані у даній роботі, застосовані в наступних проектах:

1. Науковий проект програми Horizon Europe «Satellites for Wilderness Inspection and Forest Threat Tracking» (SWIFTT) (грантова угода 101082732, 2022-2025).
2. Науковий проект НФД України «Методи і моделі глибинного навчання для прикладних задач супутникового моніторингу» (2020.02/0292, 2020-2023).
3. Науково-технічна робота (НТР) «Інформаційні технології геопросторового аналізу розвитку сільських територій і громад» (Договір РН/27-2023 від 25 травня 2023 р.) за рахунок зовнішнього інструменту допомоги Європейського Союзу для виконання зобов'язань України у Рамковій програмі Європейського Союзу з наукових досліджень і інновацій «Горизонт 2020».

Розроблені методи можуть бути застосовані в різних областях машинного навчання, передусім для моделей, що використовують функцію втрат для свого навчання. Окрім того, GAN-аугментація може бути використана для більш широкого класу моделей, а метод визначення частки неоднозначних даних та його майбутні модифікації можуть бути використані для оцінки якості навчальних даних для різних видів задач машинного навчання.

Особистий внесок здобувача. Усі основні результати дисертаційного дослідження, представлені до захисту, одержані автором самостійно.

У публікаціях у співавторстві, здобувачеві належать такі результати:

У роботі [1] здобувачем виконано практичні експерименти та аналіз отриманих результатів. Реалізовано метод GAN-аугментації та інші підходи до генерації штучних зображень, навчено сегментаційні моделі та обчислено їхні метрики. Експериментально доведено перевагу методу GAN-аугментації над базовою моделлю іншими підходами до генерації штучних зображень.

У роботі [3] здобувачем проведено аналіз робіт попередніх дослідників, сформульовано та описано метод виявлення неоднозначних екземплярів даних, та проаналізовано можливості практичного застосування. Роботу методу продемонстровано на датасеті багатоканальних супутникових знімків.

У роботі [2] здобувачем проведено аналіз робіт попередніх дослідників, сформульовано та описано метод вагових масок, продемонстровано його роботу як для задачі сегментації, так і для задачі генерації штучних зображень по заданій масці. Показано, що використання методу дає приріст у метрик, особливо для малих класів.

Апробація матеріалів дисертації. Результати та основні положення роботи подавалися та обговорювалися на:

1. A. Okhrimenko and N. Kussul, “Knn-based algorithm of hard case detection in datasets for classification,” in *Proceedings of the 11th International Conference on Applied Innovations in IT*, vol. 11, Anhalt University of Applied Sciences, 3 2023
2. A. Okhrimenko and N. Kussul, “Unbalanced datasets management for the problem of segmentation of satellite images,” in *2024 IEEE Fourth International Conference on System Analysis Intelligent Computing (SAIC)*, 2024

Публікації. За результатами досліджень опубліковано 4 наукові роботи. Серед них 2 статті у науковому фаховому виданні України [2, 3], одне з них включено до списку міжнародної наукометричної бази Scopus з квантилем Q3 [2]. Окрім того, 1 статтю опубліковано у іноземному виданні, що включено до списку між-

народної наукометричної бази Scopus з квантилем Q2 [1]. Опубліковано 1 одноосібний розділ у монографії [4].

Структура та обсяг дисертації. Дисертаційна робота складається із вступу, чотирьох розділів, висновків, списку використаних джерел із 75 найменувань. Робота містить 118 сторінок, у тому числі: 94 сторінок основного тексту, 27 рисунків, 12 таблиць, список використаних джерел на 11 сторінках.

1 ОГЛЯД ЛІТЕРАТУРИ ЗА ТЕМАТИКОЮ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ

Незбалансованість навчальних вибірок – це фундаментальна проблема при навчанні будь-якої моделі машинного навчання. Окрім очевидного схильності модель добре вивчати лише ті дані, які є найбільш представленими у датасеті, постає проблема оцінки якості її роботи на тестовому наборі даних, який так само може бути незбалансованим. Іншою перешкодою є перекриття класів – подібність екземплярів різних класів між собою аж до неможливості розрізнення. У випадку, коли ці дві проблеми наявні одночасно, створення ефективної моделі машинного навчання значно ускладнене.

Розширення датасету за рахунок найменш представлених класів зведе до нуля проблему незбалансованості класів. В свою чергу, збільшення кількості ознак, які характеризують кожний екземпляр даних, може вирішити проблему перекриття класів. Для задачі сегментації зображень, це еквівалентно збільшенню кількості каналів у зображенні. У будь-якому випадку, постає необхідність отримати нові дані для розширення вже наявного датасету. Це не завжди можливо, через фізичну неможливість, через обмеження бюджету або ж з інших причин.

Задачу сегментації зображення у першому наближенні можна розглянути як задачу по-піксельної класифікації. У такому випадку, кожен піксель розглядається ізольовано один від одного, і на результат впливають лише значення каналів у даному конкретному пікселі. Точність такого наближення невелика, особливо для складних моделей, які враховують велике число сусідніх пікселів. З огляду на це, можна проаналізувати задачі класифікації у випадку незбалансованих класів та/або класів, що перекриваються, узагальнити отримані результати та перейти до задачі сегментації, як до більш складної.

1.1 Проблема дисбалансу та перекриття класів у задачі класифікації

На тему проблеми перекриття класів у датасетах проведено численні дослідження. Більшість із них розглядають етап навчання моделі, та пропонують модифікації стратегії вибору навчальних даних для згладжування впливу недосконалості датасету [7]. Для удосконалення і створення збалансованого набору даних часто використовують методи аугментації (штучного доповнення датасету) [8]. Проте таке доповнення навчального набору даних є прийнятним та ефективним не в усіх задачах і предметних областях. Загалом, проблему незбалансованості класів класів можна пом'якшити, повністю ігноруючи дані в неоднозначних регіонах, або ігноруючи найбільш численний клас в таких регіонах, або створюючи окремі правила для довірених і ненадійних даних [9]. Це демонструє важливість надійного методу виявлення підпросторів, де відбувається перекриття класів та тих екземплярів даних, що входять до них.

Одним із відомих підходів для аналізу наборів навчальних даних є використання методу KNN [10]. Проводилися дослідження продуктивності алгоритму KNN на наборах даних, де були присутні проблеми незбалансованості та перекриття класів [11]. Під час експериментів використовувалися штучно згенеровані датасети. Основними гіперпараметрами для його генерації були відсоток зон перекриття, а також співвідношення класів у них.

Робота [12] описує метод вирішення обох проблем: як перекриття класів, так і незбалансованості наборів даних. Для цього використовується модифікація навчального процесу, спрямована на балансування класів, які модель отримує на вхід під час навчання. Комбінація одразу двох проблем у датасеті вкрай негативно впливає на якість роботи моделі, яка була на ньому навчена. Така модель матиме тенденцію класифікувати всі екземпляри даних у суперечливих зонах як найбільш представлений у датасеті клас, ігноруючи класи з меншою кількістю екземплярів.

Автори пропонують за допомогою підходу, заснованому на KNN, визначати, які екземплярів даних найпоширеніших класів знаходяться поблизу представників найменш поширених класів. Пізніше ці точки видаляються з набору навчальних даних, зменшуючи частку найбільш представленого класу і таким чином балансуючи класи. Як наслідок, модель буде навчатися приділяти практично однакову увагу всім представленим класам.

Автори [13] також досліджують випадок одночасного перекривання класів і проблему незбалансованості класів. Подібно до попередньої роботи вони пропонують видалити найбільш представлений клас у зоні перекриття класів. Наступним етапом є генерація нових датасетів для ансамблю моделей шляхом випадкового відбору збалансованого набору даних з базового незбалансованого датасету. Ансамбль робить прогноз на основі низки простих класифікаторів, кожен з яких навчається на власному унікальному наборі даних.

Інший спосіб вирішення проблеми дисбалансу класів – це робити окремі класифікації для різних підпросторів у просторі ознак. У [14] дослідники пропонують тренувати дві класифікаційні моделі: перша приймає бінарне рішення щодо приналежності до підпростору, де наявне перекриття класів; друга модель, на основі SVM, приймає рішення про класифікацію у випадку приналежності екземпляру даних до надійного підпростору. Кілька досліджень також вводять дві моделі, першу для підпросторів з проблемою перекриття класів, і другу модель для решти простору.

Подібний підхід також можна використовувати для динамічного вибору моделі з ансамблю на етапі визначення остаточного рішення ансамблю [15]. Крім підпросторів, де класи перекриваються, існують «локальні несправедливі» зони, де більшість моделей ансамблю не можуть зробити правильне передбачення, і як наслідок, їх ансамбль також зробить помилку. Однак менша частина моделей можуть зробити правильне передбачення. Автори [15] запропонували рішення для виявлення таких моделей і вибору підмножини моделей, які здатні робити коре-

ктне передбачення у даній конкретній зоні. Таким чином, остаточна модель матиме кращі метрики для екземплярів даних у всьому просторі ознак.

Розглянуті методи боротьби з перекриттям класів сильно залежать від правильності визначення проблемних зон та роблять деяке спотворення оригінального датасету, видаляючи небажані дані, що може бути некоректним. Це ще раз доводить важливість точного визначення неоднозначних екземплярів даних для подальших маніпуляцій. Також важливо використовувати дану інформацію не лише під час навчання моделі, але під час інших етапах розробки.

1.2 Вагові коефіцієнти

Для боротьби з проблемною незбалансованістю класів у датасеті, використовуються різноманітні підходи та алгоритми [16, 17]. Одним з них є корекція процесу підбору навчального пакету, зазвичай з метою зрівняти пропорцію класів у нього чи виключити аномальні екземпляри даних [18]. Вирівнювання пропорцій класів за може відбуватися з рахунок малих класів, а саме шляхом штучного дублювання даних найменш представлених класів, генерації штучних екземплярів таких класів і включення їх до оригінального датасету, зміни класів спеціально вибраних екземплярів даних або ж довільних комбінацій перелічених методів. Також вирівнювання може відбуватися з рахунок найбільш представлених класів, шляхом видалення частини його екземплярів з датасету [19]. У першому підході дані можуть спотворюватися, а другий підхід може призвести до втрати інформації.

Іншим методами класифікації незбалансованих даних є адаптація моделей машинного навчання, використання різноманітних технік аугментації даних (рис. 1.1), а також введення спеціальних метрик та/або функцій втрат [16, 17]. Створення ансамблів пропонується у роботах [15, 20, 21], також пропонується мати різні

моделі для різних підпросторів у просторі ознак, та загальну класифікаційну модель, яка буде визначати приналежність до того чи іншого простору [14].

Спеціально розроблені функції втрат можуть значно покращити результативність моделей на незбалансованих даних, однак подібний результат може бути досягнутий використанням модифікацій стандартних функцій за допомогою вагових коефіцієнтів [22, 23]. Більш представленим класам відповідають менші значення вагових коефіцієнтів, а класам з відносно невеликою кількістю екземплярів – більші вагові коефіцієнти. Це спонукає модель вивчати насамперед малі класи. До переваг такого підходу належить збереження датасету у оригінальному вигляді без спотворень та відсутність необхідності зміни процесу навчання моделі та її архітектури.

У [24] пропонується використовувати різні значення вагових коефіцієнтів для різних класів та динамічно змінювати їх в процесі навчання відповідно до поточних метрик. Завдяки цьому зникає необхідність підбору вагових коефіцієнтів, які б були актуальними впродовж усього час навчання моделі. Динамічні вагові коефіцієнти, що призначаються кожному окремому екземпляру даних, також присутні в алгоритмі AdaBoost [25].

Робота [26] присвячена проблемі незбалансованості класів у задачі сегментації супутникових знімків і пропонує використовувати згенеровані зображення для генерації мало представлених у датасеті класів. Також доцільно згадати генеративні мережі, що виконують умовну генерацію [27]. Такі моделі здатні згенерувати багатоканальне зображення по заданій сегментаційній масці, що дозволяє створювати зображення з бажаною конфігурацією полів та інших об'єктів.

У таблиці 1.1 згруповано більшість розглянутих методів по категоріям, перелічено приклади та недоліки кожної групи.

Численні дослідження проблеми незбалансованих класів у навчальних наборах даних демонструють її актуальність, а обмеження поточних рішень обумовлюють необхідність розробки нових методів та підходів.

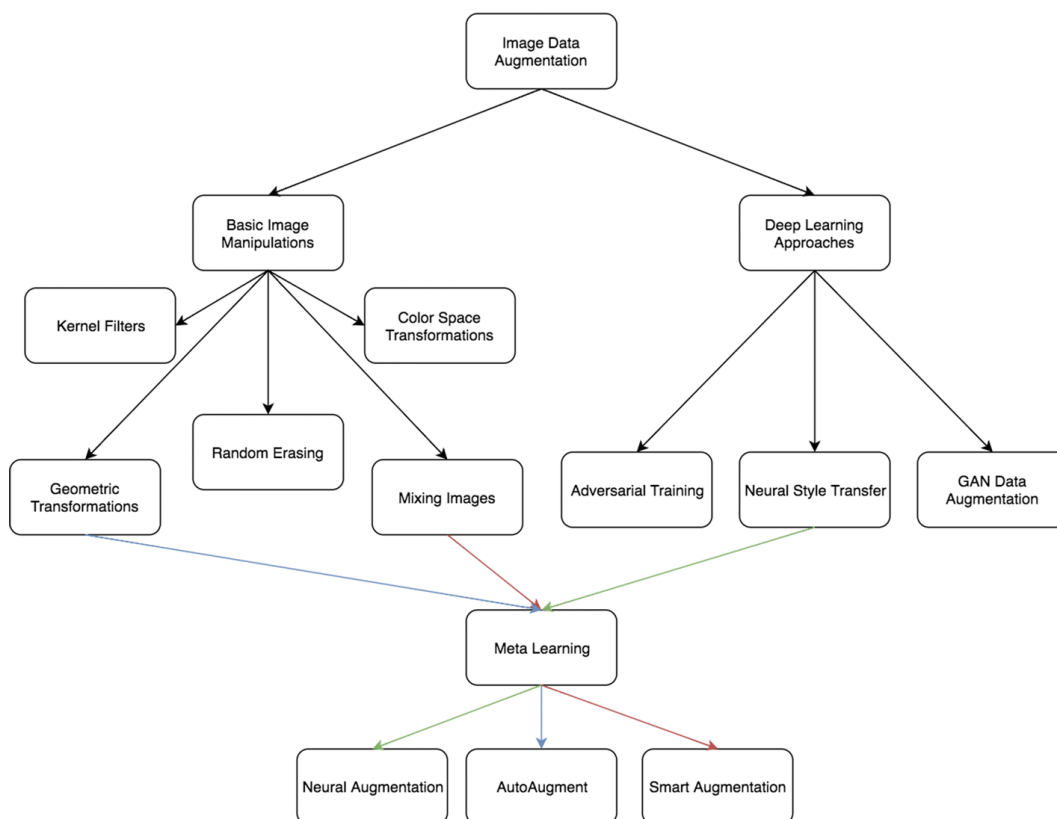


Рисунок 1.1 — Класифікація різних методів аугментації зображень з роботи [8]

1.3 Огляд предметних областей

У даній роботі досліджується задача семантичної сегментації на прикладі визначення сільськогосподарських культур на багатоканальних супутникових знімках та виявленні різних типів біологічних клітин на гістологічних зображеннях. Обидві предметні області характеризуються значним дисбалансом класів, що не може бути усунений шляхом збору додаткових даних, оскільки вони так само будуть незбалансованими. Окрім того, отримання нових даних та їх розмітка є коштовним та трудомістким процесом. Так само економічно недоцільно збільшувати число ознак для боротьби з проблемою перекриття класів, оскільки для цього необхідно збільшити число каналів у зображеннях.

Наступні підрозділи присвячено короткому огляду існуючих проблем, підходів та важливих результатів інших дослідників для даних предметних областей

Таблиця 1.1 — Групи методів, спрямовані на боротьбу з проблемами незбалансованості та перекриття класів, приклади та недоліки

Група методів	Приклади	Недоліки
Особливі стратегії навчання	корекція підбору навчального пакету [7, 18]; ігнорування усіх даних або тільки великих класів у області перекриття класів [12, 13];	Втрачається частина даних, спотворюється навчальна вибірка.
Ансамблі	ансамблі [15, 20, 21]; AdaBoost [25]; різні моделі для різних підпросторів [9, 14];	Необхідно навчати декілька моделей.
Введення вагових коефіцієнтів	статичні вагові коефіцієнти на кожний клас [22, 23]; динамічні вагові коефіцієнти [24]; AdaBoost [25];	Збільшується фокус моделі на всі екземпляри найменш представлених класів, незалежно від їх роздільності та важкості розрізнення.
Аугментація (штучне доповнення датасету)	генерація представників малих класів [8, 13]; використання GAN [28];	Штучно згенеровані екземпляри даних можуть мати відмінні властивості від оригінальних.
Розширення датасету реальними даними	збільшення кількості екземплярів даних та кількості ознак	Нові дані також можуть бути незбалансованими; їх отримання може бути ускладнене або економічно недоцільне

з урахуванням тематики даної роботи.

1.3.1 Сегментація сільськогосподарських культур на багатоканальних супутникових знімках

Карти класифікації типів сільськогосподарських культур мають важливе значення для сільського господарства, моніторингу довкілля та аналізу. Точна класифікація сприяє вирішенню проблем виробництва сільськогосподарських культур та оцінці урожайності [29], оцінці ризиків та аналізу умов навколишнього середовища [30]. Ці карти є необхідними для оцінки земельного покриття та земельного

використання, дозволяючи вивчати соціальні та економічні аспекти у дослідницьких областях. Вони також допомагають оцінити важливі екологічні та сільськогосподарські показники для оцінки процесів опустелювання, стійкого використання землі та досягнення цілей сталого розвитку [30]. Крім того, інвентаризація сільськогосподарських культур за допомогою класифікації супутникових даних стала необхідною для прийняття рішень на рівні країни в багатьох регіонах світу. Послуги, такі як CropScape Сполучених Штатів [31] і система Sen-4-CAP Європейського Союзу [32], підтримують сільськогосподарську політику в Північній Америці і Європі.

Стрімкий ріст індустрії та доступності даних дистанційного спостереження призвів до нових досліджень щодо можливостей впровадження штучного інтелекту для спостереження за земною поверхнею. Сучасні методи глибокого навчання пропонують потенціал покращення якості супутникових продуктів для точної оцінки сільського господарства та економічних аспектів [33]. Глибоке навчання, зокрема згорткові нейронні мережі, стало перспективним підходом до технологій картографування відповідно до потреб у класифікації [34]. Однак глибокі моделі машинного навчання стикаються з двома критичними викликами, що ускладнюють масштабованість, універсальність та застосування в реальному світі: необхідність великої кількості розмічених супутникових даних і наявність дисбалансу даних через розподіл покриття земної поверхні та типів сільськогосподарських культур, що призводить до існування менш представлених і більш представлених класів. Крім того, складність взаємозв'язків між багатоспектральними та текстурними характеристиками обмежує можливості методів аугментації.

Зазначені вище питання не можуть бути вирішені за допомогою традиційних технік аугментації комп'ютерного зору, таких як обертання, випадкове видалення та змішування [8]. Замість цього, найбільш ефективним рішенням є використання передових технологій: генеративних змагальних нейронних мереж (GAN), запропонованих у роботі [35]. Незважаючи на популярність GAN в різних науко-

вих галузях, їх застосування в дистанційному спостереженні залишається недостатньо розвиненим і в основному обмежується обробкою зображень і завданнями супер-роздільності. Дослідження [36] представляє GAN-підходи для підвищення роздільності та масштабування супутникових зображень і підтверджує високу якість вивчення текстурних характеристик реального світу такими GAN. Дослідження [37, 38] показали, що модель GAN може генерувати реалістичні супутникові зображення за текстурними та багатоспектральними характеристиками. Автори [39] у своєму дослідженні показали, що моделі GAN здатні вивчати закономірності часових змін багатоспектральних характеристик супутникових даних для прогнозу наступного зображення в послідовності та таким чином заповнювати прогалини в часових рядах. Схожий висновок був зроблений для складних багатоспектральних характеристик різних типів сільськогосподарських культур у дослідженні [40]. Крім того, методи, такі як “pixel-to-pixel” (pix2pix), вже можуть створювати реалістичні супутникові зображення за допомогою існуючих карт покриття землі [41].

Розширення датасету за допомогою методів аугментації збільшує в тому числі і кількість екземплярів малих класів. Одним з новітніх підходів у цій області є використання генеративних змагальних мереж для генерації штучних екземплярів даних [8]. Генеративні мережі здатні створювати високоякісні багатоканальні зображення, а їх спектральні характеристики відповідають реальним знімкам [37].

1.3.2 Сегментація різних типів біологічних клітин на гістологічних зображеннях

Гістологічні зображення тканин широко використовуються при виявленні патологій, зокрема злоякісних пухлин. Використання методів комп’ютерного зо-

ру дозволяє частково автоматизувати обробку даних, що може використовуватися як допоміжний інструмент під час роботи лікарів при встановленні клінічного діагнозу [42]. У порівнянні з традиційними алгоритмами машинного зору, де ознаки виділяються вручну, глибокі нейронні мережі показують кращі результати у задачах сегментації тканин, окремих клітин чи виділенні границь[43].

Подальше покращення результатів може відбуватися за рахунок зміни архітектури моделі або її окремих шарів. У роботі [44] розглядається модифікації згорткового шару, де вхідні дані ущільнюються безпосередньо перед операцією згортки. Це дозволяє виділяти більше ознак на кожному послідовному шарі нейронної мережі, зберігаючи при цьому просторові властивості. Як наслідок, загальна кількість шарів зменшується, а також зменшується час обробки зображення.

Для боротьби з незбалансованістю навчальних вибірок, присутньою у більшості публічно доступних наборах даних, авторами [45] запропоновано використання зваженої функції втрат відповідно до частоти класів у вибірці. Менш представленим класам відповідають більші вагові коефіцієнти, а більш представленим класам – відповідно менші. Таким чином, вплив кожного класу на зміни параметрів моделі стає приблизно рівним. Також у даній роботі представлено алгоритм післяопераційної обробки результатів роботи моделі, який враховує впевненість моделі та передбачену відстань до границі клітини для кожного пікселя. Таким чином отримується повна оцінка форми клітини та фільтруються хибні передбачення пікселів фону.

Подібний підхід продемонстровано у роботі [46] присвяченій сегментації уражень шкіри, де в ролі функції втрат використовується зважена контурна крос-ентропія (WCCE). Вона змушує модель приділяти більше уваги контурам областей інтересу, що загалом покращує якість сегментації та точного розпізнавання границь уражених зон.

У роботі [47] для подолання проблеми незбалансованості класів використовується узагальнена функція втрат Дайса (Dice Loss). Вклад кожного класу в оста-

точно значення функції регулюються ваговими коефіцієнтами, які обернено пропорційні площі, що даний клас займає на конкретному зображенні. Окрім того, наведено кроки попередньої обробки даних, такі як зміна розмірів та розбиття зображень, нормалізація та аугментація.

Більшість публічно доступних набрів даних складаються з невеликої кількості зображень, а деякі класи зустрічаються набагато частіше за інші [45]. Можливо розширити навчальну вибірку за рахунок штучно згенерованих зображень. Генеративні змагальні мережі здатні створювати реалістичні гістологічні зображення за заданою сегментаційною маскою [28], що дозволяє згенерувати багато зображень з найменш представленими класами та вирівняти розподіл класів у об'єднаному датасеті з реальний та штучних даних.

Для покращення розпізнавання границь окремих клітин, автори [48] вводять додатковий клас, який відповідає цій межі, а його пікселі знаходяться поруч зі границею на невеликій відстані. Загальне число пікселів такого класу дуже мале, що посилює незбалансованість класів у датасеті. Тому при навчанні використовується зважена крос-ентропія, де нововведеному класу відповідає більший ваговий коефіцієнт у порівнянні з рештою класів. Також більші вагові коефіцієнти отримали пікселі, як клітин так і фону, які лежать у безпосередній близькості від нового класу, тобто поруч з границею – чим далі від границі, тим менший коефіцієнт.

Висновки до розділу 1

Значна кількість розглянутих методів, особливо з категорій «Особливі стратегії навчання» та «Ансамблі» (табл. 1.1), потребують визначення множини даних, які знаходяться у області перекриття класів. Слід також взяти до уваги можливість спотворення частини екземплярів даних, які можуть виникнути від час

збору та реєстрації датасету. З цього випливає необхідність мати метод визначення таких неоднозначних екземплярів даних, що дасть змогу врахувати особливості датасету під час навчання моделі. Застосування методів цих категорій у задачі сегментації досить обмежене, оскільки на кожному зображенні міститься одночасно декілька класів, а окремі пікселі не можуть бути видалені з нього. Підходи, що базуються на ансамблях, важко застосувати у задачах комп'ютерного зору через велику кількість параметрів навіть у найпростіших моделей.

Генеративні змагальні мережі здатні створювати реалістичні зображення з тими ж самими властивостями що і реальні дані. З огляду на це, доцільно розглянути їх застосування для розширення датасету за рахунок згенерованих зображень. На перший погляд, це має однозначно покращити результати роботи моделі, навченої на розширеному датасеті. З іншого ж боку, кількість реальних даних не змінилася, а згенеровані дані є лише похідним від реальних. Також постає питання, чи буде ефективною навчена на незбалансованих даних GAN-модель. Тому немає гарантії, що розширення датасету цим методом аугментації покращить результат, і лише експеримент може дати остаточну відповідь на це запитання.

Існують численні варіанти модифікації функцій втрат, адаптоване до використання вагових коефіцієнтів, проте серед них мало підходів, які враховують ступінь надійності кожного екземпляру даних незалежно від приналежності до малого чи великого класу. Навіть серед малих класів може бути підмножина даних, які у просторі ознак впевнено розділяються від представників інших класів. У задачі сегментації, кожне зображення розглядається як один екземпляр, хоч воно і складається з величезної кількості пікселів різних класів. На одному і тому ж зображенні можуть бути присутні пікселі з різною надійністю, належні до класів з різним розподілом тощо. Окреме зображення може одночасно мати області, які добре визначаються моделлю, так і області, які погано визначаються. У таких умовах, присвоєння кожному зображенню чи класу тільки одного вагового коефіцієнту є великим спрощенням, що може негативно впливати на навчання моделі.

Таким чином, недослідженими залишаються питання ефективності GAN-аугментації, використання індивідуальних вагових коефіцієнтів з огляду на роздільність різних екземплярів даних між собою та розробки ефективного методу для визначення екземплярів даних, які складно класифікувати їхніми ознаками.

2 МЕТОД ВИЯВЛЕННЯ СКЛАДНИХ ДЛЯ РОЗПІЗНАВАННЯ ЗРАЗКІВ В НАБОРАХ ДАНИХ ДЛЯ ЗАДАЧ КЛАСИФІКАЦІЇ У МАШИННОМУ НАВЧАННІ

Якість та продуктивність моделей машинного навчання значною мірою залежать від розміру вибірки та якості навчальних даних. У загальному випадку, чим більше даних, тим кращою буде модель. Однак якість навчальних даних також важлива, ми не можемо просто дублювати дані чи постійно відбирати їх із того самого джерела чи об'єкта. Дані мають бути різноманітними, щоб охопити якомога більший об'єм у просторі ознак. Розглядаючи задачу класифікації, набір даних має бути роздільним, наприклад повинні існувати граничні поверхні, які чітко відокремлюють точки даних, які належать до різних класів. В ідеальному випадку такі поверхні мають бути достатньо гладкими, щоб уникнути перенавчання моделі. В такому випадку дані утворюють певні кластери, кожен з яких містить екземпляри даних лише одного класу.

Проблема перекриття класів унеможлиблює побудову чітких та однозначних граничних поверхонь [49]. Як наслідок, дані не можуть бути розділені на окремі кластери, а значить, частину даних не можна чітко розрізнити між собою у просторі ознак. Крім того, в просторі ознак існують деякі підпростори, які містять суміш точок різних класів без будь-якої логіки чи структури. При наявності вказаної проблеми, дослідник стикається із наступними питаннями: чи підходить даний набір даних для даного завдання класифікації? Чи треба удосконалити процес збору даних? як досягти найкращого можливого результату, використовуючи даний набір даних, якщо проблему з накладанням класів не вдалося виправити?

Для глибшого розуміння проблеми розглянемо причини виникнення проблеми «перекриття» класів в просторі ознак. Ця ситуація може бути спричинена похибками в процесі збору даних та/або їх розмітки, або недостатньою інформативністю ознак (відсутністю складності даних). В останньому випадку додавання

додаткових ознак, що еквівалентно додаванню додаткових вимірів до простору ознак, може значно покращити придатність даних для даного завдання. Водночас, збільшення розмірності простору ознак, в свою чергу, може призводити до перенавчання моделі. Як наслідок, дослідникам потрібен алгоритм для визначення недоліків датасетів, тобто виявлення частки неоднозначних даних і підпросторів із такими даними у навчальному та тестовому наборі даних. Результати роботи алгоритму можна використати для прийняти рішення про модифікацію процесу збору даних, додавання нових ознак або збільшення кількості і точності екземплярів даних у «сумнівних» підпросторах і навколо них.

Найбільш простим методом дослідження датасету є його візуалізація у двовимірному просторі. Масштабні датасети з великим числом ознак неможливо візуалізувати без додаткових перетворень. У цьому випадку для відображень даних з простору з великою розмірністю у простір з меншою розмірністю можна використовувати такі алгоритми, як PCA [50] та tSNE [51], та нанести отриманий результат на двовимірний графік. До недоліків цих методів слід віднести залежність від людської суб'єктивності та неможливість ефективно візуалізувати й аналізувати дані великої розмірності через великі втрати інформації під час перетворення.

Однак у більшості випадків дослідник не може вплинути на процес збору набору даних і змушений працювати з даними, в яких наявна проблема перекриття класів. У цьому випадку йому також потрібен алгоритм, який оцінить потенційну точність класифікації на заданому наборі даних і визначить неоднозначні екземпляри даних для корекції навчання моделі.

Визначені на попередніх кроках «сумнівні» підпростори в просторі ознак можуть бути використані для корекції результатів (передбачень) моделі. Для прикладу, для екземплярів даних, що потрапляють у неоднозначну частину простору ознак можна використовувати інші правила та навіть інші моделі.

Для задачі класифікації зображень визначення неоднозначних екземплярів даних стає особливо важливим. Зазвичай згорткова нейронна мережа (CNN) [52]

робить відображення з простору зображень у простір ознак, таким чином вхідне зображення перетворюється на вектор ознак. Остаточна класифікація виконується на базі цього одновимірному вектору. Виявлення неоднозначних екземплярів даних у просторі ознак є важливою науковою проблемою для покращення якості розпізнавання зображень за допомогою CNN. Варто зауважити, що вищевказане перетворення у простір ознак не детерміноване, а роздільність класів залежить не лише від якості даних, але й від якості роботи згорткових шарів нейронної мережі, що транслюють зображення у вектор ознак (embedding).

У цьому дослідженні запропоновано новий метод для виявлення неоднозначних екземплярів даних на основі метода K найближчих сусідів (KNN) [53], продемонстровано його роботу на супутникових даних (багатоспектральних оптичних знімках) в задачі класифікації сільськогосподарських культур та розглянуто можливі варіанти використання даного алгоритму під час розробки моделей машинного навчання. Вперше ідею цього описаного в цьому розділі методу представлено на конференції ICAIT 2023, де його роботу було проілюстровано на прикладі штучно згенерованого набору даних в двовимірному просторі [5].

2.1 Алгоритм визначення неоднозначних екземплярів даних у вибірках для задач машинного навчання

Нехай \hat{X} – множина екземплярів даних, а $\vec{x}_i \in \hat{X}$ – i -й екземпляр даних з цієї множини. Довжина даного вектору є рівною числу ознак. Аналогічно, \hat{Y} – множина міток класів, до яких можуть належати екземпляри даних \vec{x}_i , $y_i \in \hat{Y}$ – істинний клас для екземпляра даних \vec{x}_i . Загальне число екземплярів даних визначимо як N . Разом вони утворюють датасет:

$$(\vec{x}_i, y_i), i \in [1, 2, \dots, N], \vec{x}_i \in \hat{X}, y_i \in \hat{Y}$$

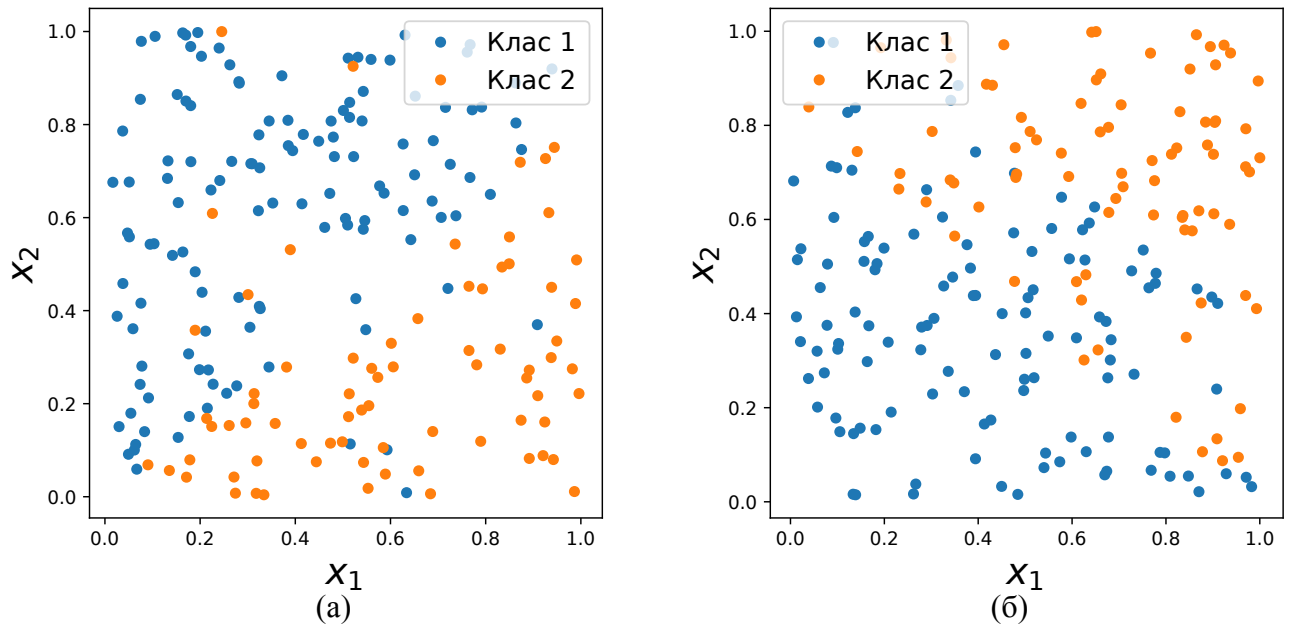


Рисунок 2.1 — Проблема викидів (а) та проблема перекриття класів (б)

де $N = |\hat{X}|$ – розмір датасету, а $|\hat{Y}|$ – число класів у ньому.

Неоднозначним екземпляром даних будемо вважати такий екземпляр (\vec{x}_i, y_i) , який за своїми ознаками є подібний до інших екземплярів (\vec{x}_j, y_j) , $j \neq i$, аж до ідентичності, в той же час інші екземпляри (\vec{x}_j, y_j) належать до інших класів $y_i \neq y_j$. Розглядаючи екземпляр даних як вектор довжиною рівною кількості ознак, подібність ознак двох екземплярів можна виразити як норму різниці векторів, або ж $L2$ відстань між екземплярами даних у просторі ознак. Таким чином, математичне визначення неоднозначного екземпляру даних (\vec{x}_i, y_i) можна представити у вигляді:

$$\forall \vec{x}_q \in \hat{X}, \vec{x}_q \neq \vec{x}_i, \exists \vec{x}_j \in \hat{X} : \|\vec{x}_i - \vec{x}_j\|_{L2} \sim \|\vec{x}_i - \vec{x}_q\|_{L2}$$

Розглянемо дві можливі ситуації, які можуть виникати у задачі класифікації: викиди та перекриття класів (рисунок 2.1). Згідно з визначенням неоднозначного екземпляру даних, будь-який викид з усіх боків оточений представниками іншого класу, тож може вважатися неоднозначним екземпляром даних. При пе-

рекриті, утворюється область, де знаходяться представники обох класів, тож для будь-якого екземпляру даних поруч знаходяться також представників інших класів. Однак на відміну від викидів, поруч знаходься і представники того ж самого класу, тож постає проблема визначання граничного співвідношення між класами, за якого даний екземпляр даних слід вважати неоднозначним.

Потрібно відповісти на питання: чи можливо деякий екземпляр даних \vec{x}_i правильно класифікувати як клас y_i . Для цього буде використано ансамбль класифікаторів KNN з різними номерами сусідів $n = [0, 1, \dots, N], n \in \mathbb{N}$. Для кожного $\vec{x}_i \in \hat{X}$ ми отримаємо вектор \vec{m}_i , де елемент m_i^j – це результат класифікації \vec{x}_i за допомогою класифікатора KNN з параметром числа сусідів рівним j , який було навчено за допомогою набору даних $\hat{X} \setminus \vec{x}_i$. У якості функції відстані між двома екземплярами даних використовується $L2$ норма.

$$\vec{m}_i : m_i^j = KNN(\vec{x}_i, j, \hat{X} \setminus \vec{x}_i)$$

Таким чином, кожен екземпляр даних \vec{x}_i матиме відповідний йому вектор \vec{m}_i та з цих векторів можна побудувати матрицю M для оптимізації практичної реалізації методу та доступності паралельних обчислень:

$$M : M_{ij} = \{m_i^j\}$$

Тепер ми можемо порівняти кожен вектор \vec{m}_i з істинним класом y_i . Розглянемо декілька можливих випадків:

- 1) більшість елементів \vec{m}_i відповідають справжньому класу y_i ;
- 2) перші елементи \vec{m}_i відповідають справжньому класу y_i , решта ні;
- 3) передбачений клас m_i^j постійно змінюється в залежності від j , відбуваються “стрибки між класами”.

Для того щоб вважати екземпляр даних надійним та однозначним, означимо перші два випадки як обов’язкові. Вони не є повністю незалежними, оскільки

друга умова завжди істинна, коли справджується перша. Однак виконання другої умови при невиконанні першої може бути ознакою належності даного екземпляру класу до “граничної” зони у просторі ознак, де починаються представники інших класів.

Екземпляр даних не може бути надійним, якщо справджується третій випадок, це означає, що даний екземпляр даних у просторі ознак оточений іншими екземплярами з іншими мітками класу та, ймовірно, належить до зони перекриття класів.

На рисунку 2.2 схематично зображені вищеописані випадки.

Таким чином, перша умова C_1 : з елементів вектора \vec{m}_i хоча б r має дорівнювати істинному класу y_i .

$$C_1 : \sum_{j=1}^n \mathbb{1} \{m_i^j = y_i\} \geq r$$

r – гіперпараметр, невелике ціле число. У загальному випадку найкращі значення даного параметру залежать від щільності набору даних у просторі ознак.

Друга умова C_2 : з k перших елементів вектора \vec{m}_i , більшість має дорівнювати істинному класу y_i .

$$C_2 : \sum_{j=1}^k \mathbb{1} \{m_i^j = y_i\} > \sum_{j=1}^k \mathbb{1} \{m_i^j \neq y_i\}$$

Остання умова C_3 : якщо присутня часта зміна класів, екземпляр даних є неоднозначним. Математично це можна розглядати як одновимірну згортку вздовж вектора \vec{m}_i з ядром $K = [-1, 1]$. Якщо два сусідні елементи однакові, результат згортки буде нульовим. Для кожного випадку зміни класів результат згортки буде ненульовим. Сума модулів результатів згортки не повинна перевищувати деяке ціле число q , яке повинно бути достатньо низьким.

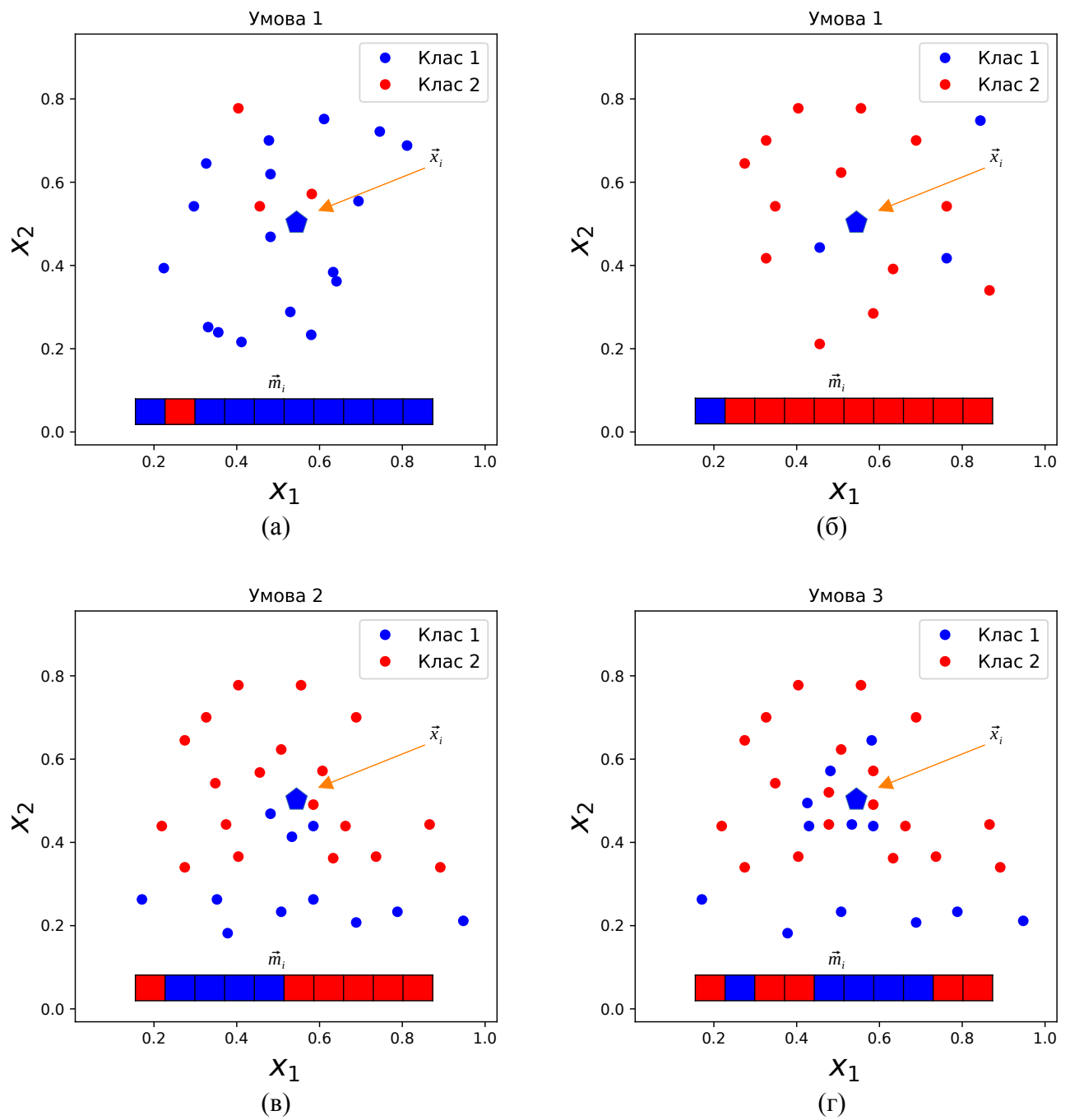


Рисунок 2.2 — Схематичні приклади різних умов методу виявлення неоднозначних екземплярів даних

$$C_3 : \sum_{j=1}^{n-1} \mathbb{1} \{ [\vec{m}_i \otimes K]_j \neq 0 \} < q, K = [-1, 1]$$

Таким чином, отримаємо остаточне правило для класифікації екземпляру даних як нормального:

$$(C_1 \vee C_2) \wedge \neg C_3$$

Та правило для класифікації екземпляру даних як неоднозначного:

$$(\neg C_1 \wedge \neg C_2) \vee C_3$$

Величини n, r, k, q є гіперпараметрами даного алгоритму та підбираються експериментально. Параметр n – визначає довжину вектора \vec{m}_i та не повинен бути надто великим чи надто малим. У першому випадку описаний метод потребуватиме надто багато обчислювальних ресурсів, у другому – результати роботи алгоритму будуть малоінформативними. Змінюючи значення r, k, q , можемо зробити одну умову важливішою за іншу.

Алгоритм визначення неоднозначних екземплярів даних в датасеті мовою псевдокоду представлено на схемі:

Data: dataset $(\vec{x}_i, y_i), i \in [1, 2, \dots, N], \vec{x}_i \in \hat{X}, y_i \in \hat{Y}$
Result: Is given data sample is hardcase: status vector \vec{s}
for $i \leftarrow 0$ **to** N **do**
 \vec{x}_i, y_i ;
 $\vec{m}_i : m_i^j = 0, |\vec{m}_i| = n$;
 $s[i]$;
 for $j \leftarrow 0$ **to** N **do**
 $m_i^j \leftarrow KNN(\vec{x}_i, j, \hat{X} \setminus \vec{x}_i)$;
 end
 $C_1 \leftarrow \sum_{j=1}^n \mathbb{1} \{m_i^j = y_i\} \geq r$;
 $C_2 \leftarrow \sum_{j=1}^k \mathbb{1} \{m_i^j = y_i\} > \sum_{j=1}^k \mathbb{1} \{m_i^j \neq y_i\}$;
 $C_3 \leftarrow \sum_{j=1}^{n-1} \mathbb{1} \{[\vec{m}_i \otimes K]_j \neq 0\} < q, K = [-1, 1]$;
 $s_i \leftarrow [(\neg C_1 \wedge \neg C_2) \vee C_3]$;
end

Розглядаючи подібні алгоритми, слід згадати роботу [54], де вводиться наступна міра неконформності для класифікації методом найближчого сусіда:

$$\alpha_i^y = \frac{\sum_{j=1}^k D_{ij}^y}{\sum_{j=1}^k D_{ij}^{-y}}$$

де α_i^y – міра неконформності для екземпляру даних x_i , D_i^y та D_i^{-y} – відсортовані вектори найближчих сусідів, що належать або не належать до класу y_i відповідно. Таким чином, екземпляри даних, що знаходяться у просторі ознак поруч з представниками інших класів, мають велику міру неконформності. Натомість, якщо деякий екземпляр даних знаходиться далеко від представників інших класів, то міра його неконформності є низькою. Це дозволяє виявляти неоднозначні екземпляри даних, використовуючи лише один гіперпараметр. Розглядаючи рисунок 2.1, такий метод здатний коректно визначити неоднозначні екземпляри у випадку (а), та лише частково – у випадку (б). Причиною є те, що даний підхід не враховує геометричну конфігурацію даних у просторі ознак та можливість побудови достатньо гладкої розділювальної площини.

2.2 Приклад роботи алгоритму на штучно згенерованих даних

З метою демонстрації роботи вищенаведеного алгоритму, згенеровано штучний набір даних, у якому наявна проблема перекриття класів, а також декілька викидів. Екземпляри даних мають дві ознаки, що робить його зручним для візуалізації на двохвимірній площині. Для генерації використано три двовимірні нормальні випадкові розподіли $\mathcal{N}_2(\bar{\mu}, \bar{\sigma})$ з такими параметрами:

- Клас 1: $\mu_1 = 1.0, \mu_2 = 1.0, \sigma_1 = 0.5, \sigma_2 = 0.2$;
- Клас 2: $\mu_1 = 2.0, \mu_2 = 1.0, \sigma_1 = 0.3, \sigma_2 = 0.3$;
- Клас 3: $\mu_1 = 1.0, \mu_2 = 2.0, \sigma_1 = 0.2, \sigma_2 = 0.5$;

Дані, які виходили за межі інтервалу $[0.0, 2.5]$, проігноровано.

Отриманий датасет зображено на рисунку 2.3а. Незважаючи на малу розмірність даного датасету, запропонований алгоритм працює з даним будь-якої роз-

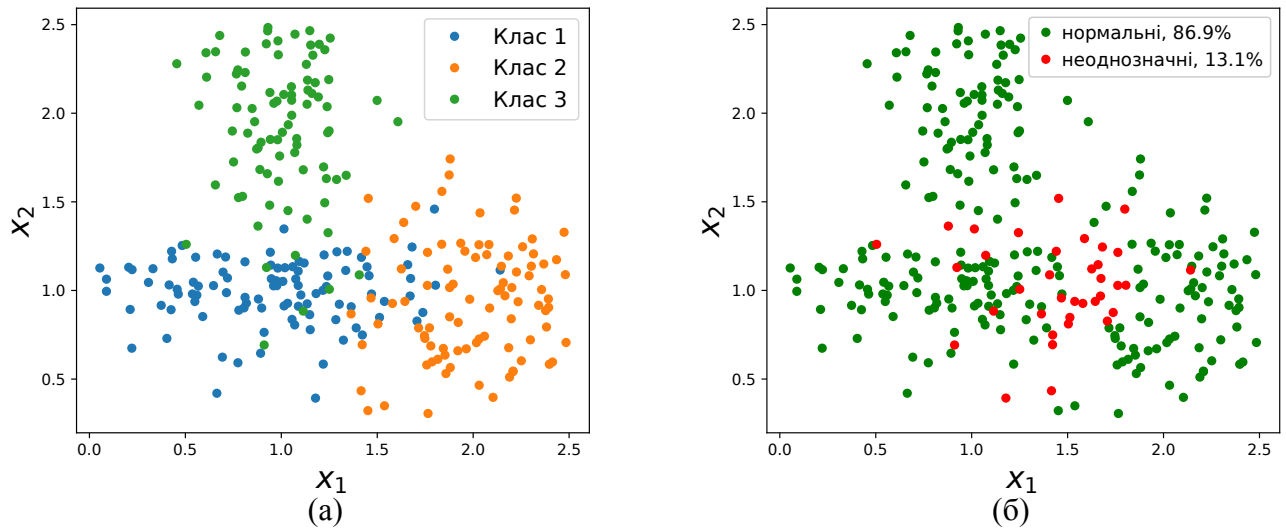


Рисунок 2.3 — Згенерований дасатет, з наявним перекриттям класів (а) та визначені неоднозначні екземпляри даних, позначені червоним кольором (б)

мірності.

Для практичної реалізації використовувати наведений алгоритм у оригінальній формі неоптимально, оскільки необхідно тренувати $n \cdot |\hat{X}|$ KNN-класифікаторів. Рішенням є використання К-D дерева [55], що дозволить знаходити довільне число сусідів з логарифмічною складністю [56].

Для початку роботи алгоритму, необхідно побудувати К-D дерево на множині \hat{X} . Далі, для кожного екземпляру даних $\vec{x}_i \in \hat{X}$, робиться запит до К-D дерева з числом сусідів $n + 1$, перший елемент відповіді ігнорується, бо це і є \vec{x}_i як найближчий до самого себе. Таким чином, отримано n найближчих сусідів до \vec{x}_i та істинні мітки їхніх класів. Після цього стає можливим заповнювати вектор $\vec{m}_i = [m_i^1, \dots, m_i^n]$, беручи різне число сусідів та визначаючи найпоширеніший клас серед них.

Приклад роботи алгоритму показано на рисунку 2.3б. Коректно визначено абсолютну більшість неоднозначних точок, що знаходяться у зоні перекриття класів, однак наявні деякі неточності що до крайових точок.

Гіперпараметри алгоритму приймали наступні значення: $n = 10, r = 5, k =$

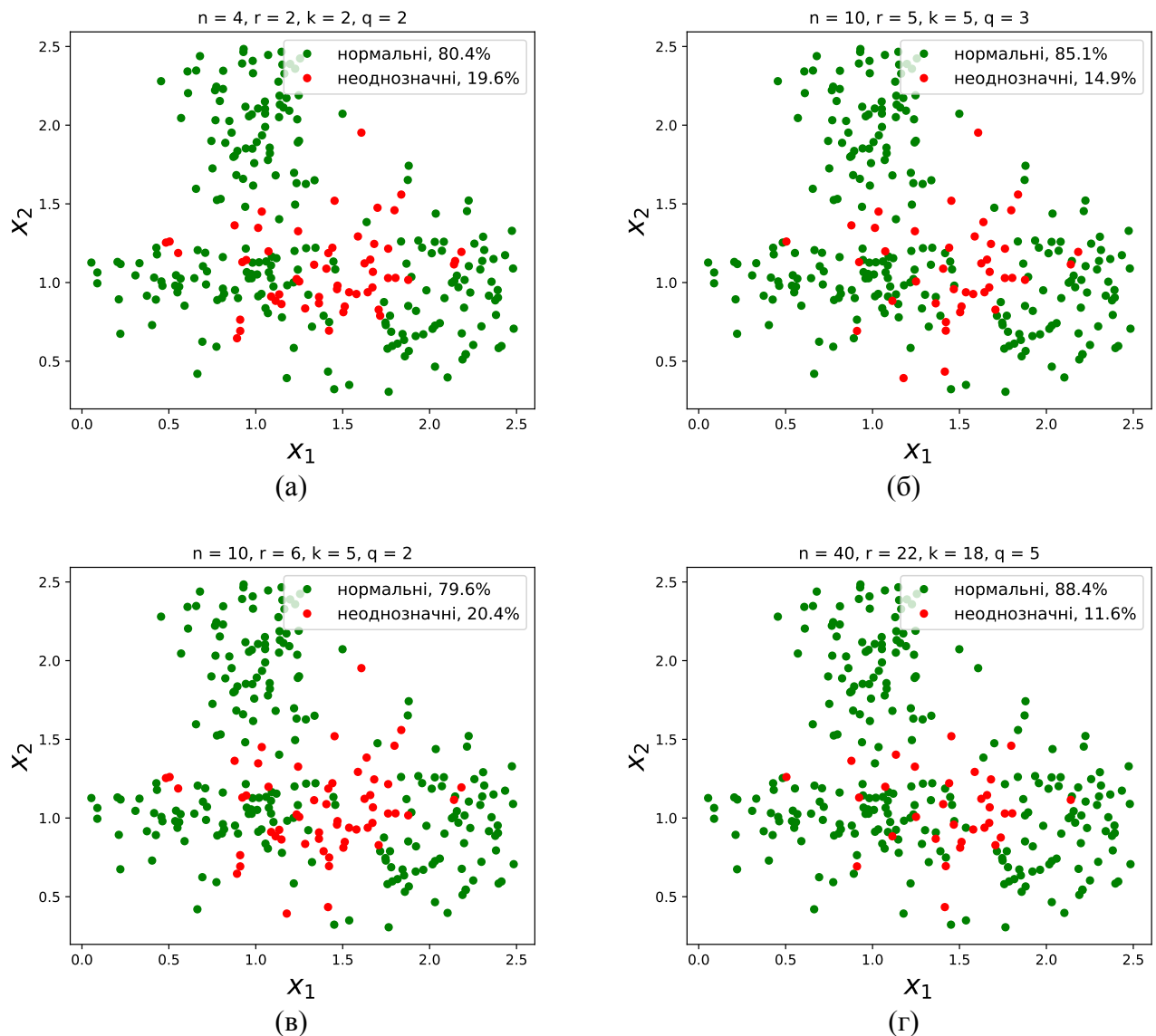


Рисунок 2.4 — Визначені неоднозначні екземпляри даних з іншими гіперпараметрами

$6, q = 4$. $n = 10$ було обрано виходячи з невеликого розміру датасету, $r = 5, k = 6$ було як близькі числа до половини від n , а $q = 4$ допускає лише три зміни класу у векторі \vec{m}_i .

На рисунку 2.4 зображено приклади роботи алгоритму для інших значень n, r, k, q . Видно, що при малих значеннях гіперпараметрів до неоднозначних даних зараховується надто багато точок, а при великих – помітно менше. При $n = 40$ кількість необхідних обчислень значно збільшилася. На даному малому датасеті час виконання алгоритму збільшився незначно, однак на великих датасетах з

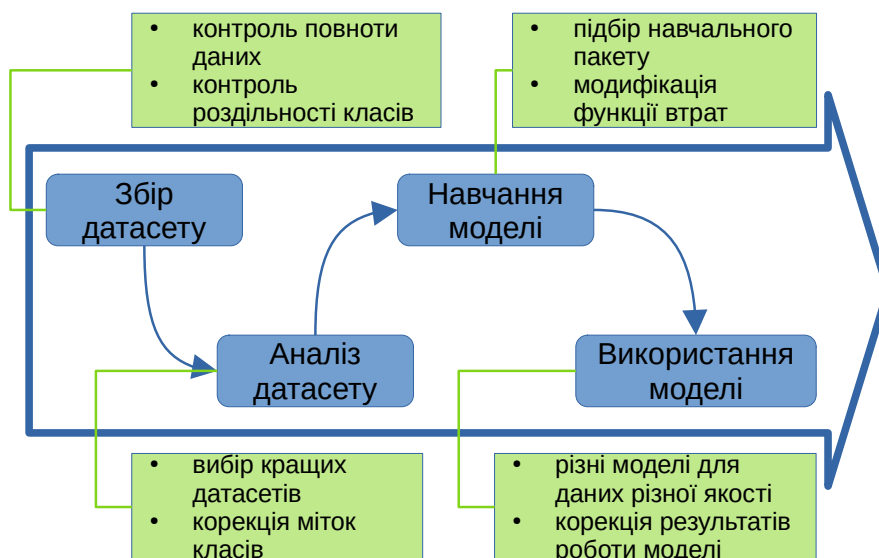


Рисунок 2.5 — Схема можливих застосувань алгоритму

великим числом екземплярів слід уникати надто великих значень n .

2.3 Можливості практичного застосування

Алгоритм визначення неоднозначних екземплярів даних здатний стати ще одним інструментом у розпорядженні дослідника і може бути використаний протягом усього циклу розробки моделі машинного навчання, від збору даних до розгортання моделі для використання у реальних умовах (Рисунок 2.5).

У наступних підрозділах детально розглядаються можливі варіанти використання запропонованого алгоритму.

2.3.1 Оцінка якості датасету

Представлений алгоритм може бути корисним як під час збору даних, так і під час аналізу наявних датасетів. Більшість сучасних задач машинного навчання вимагають великих датасетів, з великою розмірністю простору ознак та численними екземплярів даних. Досить часто не існує набору даних, який відповідає поточній задачі. Тому багато дослідників вимушені конструювати свій власний датасет, наприклад, через аутсорсинг або краудсорсинг [57, 58]. Подібна ситуація виникає, наприклад, коли в якості навчальних даних в державних системах агромоніторингу використовується інформація про посіви, надана фермерами. Пілотний проект, проведений в Україні з залученням даних від респондентів Державного департаменту статистики показав, що недостовірна інформація в даних респондентів може складати до 30% [59]. У цій ситуації надзвичайно важливо мати інструмент для контролю якості отриманих даних. Таким чином можливо виявити проблему на ранніх стадіях і дослідники матиме змогу скорегувати процес збору даних.

Можлива і протилежна ситуація, коли для даної проблеми існує декілька датасетів. Зазвичай вибираються найкращі з них або їх комбінації. Маючи можливість визначити відсоток неоднозначних даних, дослідник може вибрати дані найкращої якості. Також можливо побудувати новий об'єднаний датасет з кількох найкращих, з найменшою часткою неоднозначних даних у них.

2.3.2 Корекція формування навчального пакету

Здатність визначити ненадійні та неоднозначні екземпляри даних дає можливість покращити процес навчання моделі. Один з можливих методів – це змінити стратегію формування пакету даних при навчанні.

Під час навчання моделі можна використовувати нормальні дані для навчання частіше, ніж неоднозначні. Змінюючи частоту потрапляння неоднозначних даних у навчальний пакет, можливо надати моделі бажані властивості, аж до повного видалення неоднозначних екземплярів даних. Отримана модель прийматиме рішення більшою мірою (суттєво) на основі надійних зразків даних, а не на неоднозначних. Це дозволяє зробити правильний прогноз у підпросторах, де немає перекриття класів, і приділяти менше уваги підпросторам, де класи перекриваються, оскільки у таких областях неможливо правильно визначити належність до класу.

2.3.3 Особлива стратегія ансамблювання

Можливість чітко розрізняти різні типи екземплярів даних за їх якістю дає можливість навчити кілька моделей, кожна з яких працює лише у певних зонах простору ознак. Таким чином, правила класифікації для підпростору з звичайними екземплярами даних будуть відрізнятися від неоднозначних. Остаточний ансамбль моделей складатиметься з двох підмножин: перша – з моделей, призначених для надійних даних, а друга – для даних, які знаходяться біля неоднозначних екземплярів даних у просторі ознак. Подібний підхід до побудови динамічних ансамблів моделей у випадку незбалансованих вибірок запропоновано в [15].

Варто зауважити, що неоднозначний екземпляр даних може або лежати в зонах, що перекриваються, або бути викидом. Зрозуміло, що останньому випадку ми не можемо використовувати другу підмножину класифікаторів, і нам потрібно фільтрувати такі випадки, тож потрібне подальше вдосконалення даного підходу.

2.3.4 Модифікація датасету

Багато методів роботи з незбалансованими датасетами і проблемами перекриття класів пропонують видалити з датасету екземпляри даних, які відносяться до найбільш представленого класу та знаходяться у зонах перекриття класів [60]. Натомість, можна внести зміни до їх істинного класу без зайвого видалення даних.

Нам потрібно замінити мітки класів у зонах, де присутнє перекриття класів. Цільова мітка істинного класу, на яку необхідно поміняти мітки екземплярів даних у таких зонах, сильно залежить від поточних цілей. Для найкращих метрик має сенс змінити всі мітки класів на мітку найбільш представленого класу. Щоб боротися з проблемою дисбалансу класів, усі мітки класів у сумнівній зоні можна змінити на найменш представлений клас.

Як і в попередньому підрозділі, нам потрібно відфільтрувати викиди, для них мітки класів можна змінити на мітки екземпляр даних, які оточують їх у просторі ознак.

Висновки до розділу 2

Цей розділ присвячений дослідженню проблеми перекриття класів разом з іншими можливими важкими випадками, такими як викиди та незбалансованість класів. Існує багато причин виникнення описаних проблеми, найважливішими з яких є недостатня точність збору даних при формування датасету та замала кількість ознак, з чого випливає надто мала розмірність простору ознак. Незалежно від причин, деякі екземпляри даних просто неможливо класифікувати правильно, оскільки вони мають схожі ознаки з іншими екземплярами, які мають іншу мітку класу. Як приклад можна розглянути задачу сегментації супутникових знімків для

визначення землекористування [61]. Через схожість спектральних характеристик таких культур, як пшениця та ячмінь, їх неможливо розділити у просторі ознак. Такі екземпляри даних заважають процесу навчання моделі та можуть стати причиною неправильного передбачення на реальних даних.

У цьому розділі представлено новітній алгоритм, який дає змогу будь-якому досліднику мати більш чітке уявлення про якість датасетів, та не залежить від розмірності простору ознак. На відміну від методів візуальної оцінки, заснованих на декомпозиції, описаний алгоритм видає детерміновані числові показники якості даних, такі як відсоток надійних даних. Крім того, він дозволяє чітко відрізнити нормальні екземпляри даних від неоднозначних, відкриваючи можливість коригування.

Розглянуто задачу по-піксельної класифікації штучно згенерованого двовимірного датасету. Визначено неоднозначні точки, які важко класифікувати правильно, та обчислено відсоток неоднозначних даних.

Запропонований алгоритм може бути використаний протягом усього циклу розробки моделі машинного навчання. Сюди входить контроль та корекція процесу збору навчальних даних, а також вибір найкращих датасетів та їх змішування. Виявивши неоднозначні екземпляри даних, можна застосувати різні правила для них і для надійних даних, як під час навчання, так і під час роботи на реальних даних. Алгоритм також відкриває простір для модифікації набору даних, мітки класів неоднозначних зразків даних змінюються відповідно до поточних задач, таких як високі метрики або балансування класів.

3 ВИКОРИСТАННЯ ВАГОВИХ МАСОК ДЛЯ ПОКРАЩЕННЯ ЯКОСТІ МОДЕЛЕЙ СЕГМЕНТАЦІЇ

Існує багато факторів які впливають на якість моделі машинного навчання. Серед головних з них можна виділити тип та архітектуру моделі, алгоритм навчання, а також кількість та якість даних з навчальної вибірки. До деякої межі недоліки одного компоненту можна компенсувати за рахунок покращення іншого компоненту, наприклад збільшивши датасет, чи використати більш складну модель [62]. Частим недоліком навчального датасету у задачі класифікації чи сегментації є його незбалансованість: існує певна підмножина класів, що містять в собі мало екземплярів даних в порівнянні з іншими класами [16]. При навчанні моделі на таких даних, вона має тенденцію добре розпізнавати найбільш представлені класи та мати нижчі метрики на малих класах. Оскільки найменш представлені класи зазвичай складають малу частку датасету, це мало впливає на загальні метрики і помітити дану проблему можна шляхом обчислення метрик окремо для кожного класу.

У багатьох задачах отримання нових даних надзвичайно ускладнене, обмежене бюджетом або ж зовсім неможливе, що виключає вирішення наведеної ситуації шляхом розширення датасету екземплярами малих класів. Особливо гостро дана проблема стоїть у області розпізнавання супутникових знімків, де кількість зображень та їх якість обмежена режимом руху супутників, часом доби, погодними умовами, артефактами роботи камер тощо.

У даному розділі розглядається задача сегментації супутникових зображень для розпізнавання полів з різними сільськогосподарським культурами. При незбалансованому датасеті, модель, що навчена на ньому, може систематично помилятися на малих полях з малопоширеними культурами. До всіх попередніх обмежень збору датасету додається незбалансована структура посівів: представників однієї культури може висіватися набагато менше ніж інших, тому незбалансова-

ність навчальних даних є цілком природнім та невідворотнім явищем.

Математичне формулювання задачі сегментації зображень наступне: нехай \hat{X} – зображення розмірністю $[W \times H \times Ch]$ та числом каналів Ch , а \hat{Y} – істинна маска сегментації розмірністю $[W \times H]$ з числом класів C . Необхідно знайти таку модель $F : \mathbb{R}^{W \times H \times Ch} \rightarrow \mathbb{R}^{W \times H \times C}$, що:

$$\hat{Y}' = F(\hat{X}), \sum_{i=1}^W \sum_{j=1}^H \mathbb{1} \left\{ \arg \max_k (Y'_{ijk}) = Y_{ij} \right\} \rightarrow W \cdot H$$

Пропонується метод, який дозволяє зменшити упередження моделі, без збору додаткових тренувальних даних. Його мета полягає у спробі компенсувати недостатню якісь датасету більш досконалим алгоритмом навчання моделі, розглянуто використання вагових масок з модифікованими функціями втрат.

3.1 Метод вагових масок

Для генерації вагових масок необхідним алгоритм, за допомогою якого можна оцінити важливість та/або надійність пікселів та їх вплив на загальну точність сегментації. Оскільки екземпляри даних, що належать до різних класів матимуть майже ідентичні ознаки, то модель з великою ймовірністю не зможе визначити до якого класу віднести такі дані, бо подібні дані спостерігаються в декількох класах одночасно. Такі неоднозначні екземпляри можуть суттєво впливати на процес навчання, оскільки модель буде намагатися навчитися класифікувати точки різних класів, які майже не відрізняються між собою. Створення вагових масок допоможе врахувати ці особливості при навчанні моделі машинного навчання.

У подальших експериментах використано алгоритм виявлення неоднозначних екземплярів даних, описаний у попередньому розділі, однак варто зауважити що це не єдиний доступний варіант. Для даного етапу може бути використаний будь-який алгоритм, який здатний присвоїти кожному екземпляру даних з датасе-

ту деяке число, яке прямо корелює з надійністю чи довірою до цього екземпляру даних.

У даному розділі пропонується використовувати просторові вагові коефіцієнти при обчисленні функції втрат. Для кожного фрагменту зображення на основі значень його пікселів генерується відповідна йому вагова маска.

Отримані маски використовуються для модифікації процесу тренування моделей глибокого навчання, шляхом зміни функції втрат. Звичайну функцію втрат можна описати наступною залежністю:

$$Loss = F(\hat{Y}, \hat{Y}')$$

де \hat{Y} , \hat{Y}' – істинна маска сегментації та маска, яку вивела модель, відповідно.

Запропонований метод ускладнює цю функцію, додаючи залежність від маски, яка виконує роль вагових коефіцієнтів для кожного пікселю:

$$Loss = F(\hat{Y}, \hat{Y}', \hat{M})$$

де \hat{M} – вагова маска такої ж розмірності як і \hat{Y} .

Слід зазначити відмінності даного способу від інших методів введення вагових коефіцієнтів у функцію втрат. Відмінності від застосування вагових коефіцієнтів для кожного окремого класу у даному випадку очевидні, оскільки значення вагової маски у запропонованому підході не залежить прямо від істинного класу пікселів на фрагментів. Інший же підхід, який характеризується застосуванням вагових коефіцієнтів для кожного фрагменту з набору даних, є більш вузьким у порівнянні з представленим. Він має лише один коефіцієнт на індивідуальний фрагмент, який входить до набору даних, а запропонований метод має по одному коефіцієнту для кожного пікселю у кожному фрагменті. Відповідно, сукупність таких коефіцієнтів, впорядкована за координатами відповідних пікселів, може розглядатися як окреме зображення або ж маска, звідки і пішла їх назва.

Таким чином, використання запропонованого підходу дозволяє контролювати вплив кожного пікселю на навчання моделі. При цьому важливими є не належність пікселів до малого чи великого класів, а лише можливість розділення даних пікселів у просторі ознак. Для задачі сегментації зображень, за ознаки пікселів приймаються їх кольори, а саме інтенсивності сигналів у відповідних каналах зображення.

3.2 Опис датасету для задачі сегментації супутникових знімків

3.2.1 Опис даних

В якості набору даних, який ми будемо досліджувати, розглянемо навчальну вибірку для задачі класифікації земного покрову, використану для побудови нейромережевої задачі класифікації в [63]. Вона включає дані оптичного супутника Sentinel-2 з просторовим розрізненням 10 метрів, які покривають Київську область. Для уникнення проблеми хмарності, в даному дослідженні використовується композит, отриманий з супутникових знімків Київської області, знятих супутником Sentinel-2 впродовж липня 2021 року. Даний композит має 4 канали: синій (490 нм), зелений (560 нм), червоний (665 нм) та інфрачервоний (842 нм). Фрагмент 3-канального зображення композиту для Київської області в палітрі true color, яка включає канали 490, 560 та 665 нм, показано на рисунку 3.1а.

Істинна карта сегментації (GT) взята з роботи [64]. Вона була отримана на інших композитах супутникових зображень з використанням згорткової нейромережевої моделі, розробленої Інститутом космічних досліджень НАНУ-ДКАУ та описаної в [64]. Засобами бібліотеки GDAL, дана карта сегментації зіставлена з власним, вищеописаним композитом. Приклад фрагменту карти GT показаний на рисунку 3.1б.

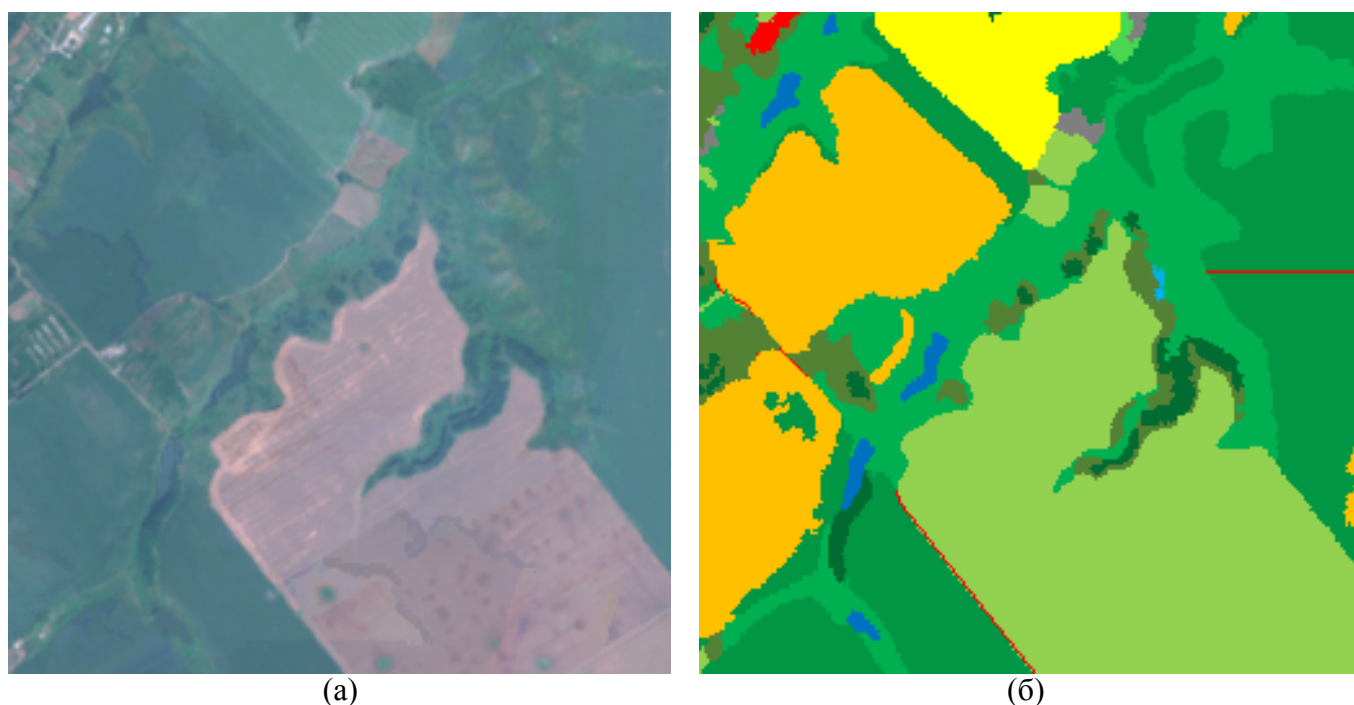


Рисунок 3.1 — Фрагмент триканального супутникового зображення (а) та відповідної карти класифікації земного покриття (б)

3.2.2 Аналіз та візуалізація датасету

Розглянемо задачу по-піксельній класифікації, що зводить дану проблему до задачі класифікації на декілька класів за чотирма ознаками, що дорівнює числу каналів. Для формування навчального датасету виберемо 8 класів сільськогосподарських культур, інші культури віднесено до окремого класу “Інші культури”. Класи, які не відносяться до сільськогосподарських культур, були відкинуті, наприклад забудівлі чи водні ресурси.

На рисунку 3.2 зображена гістограми розподілу значень пікселів в залежності від номеру каналу, а також індивідуальні гістограми для перших п’ятих класів. Значення в каналах які відповідають видимому світлу, лежать в межах $[0.0, 0.2]$. Натомість останній канал, який відповідає інфрачервоному випромінюванню, має більш широкий діапазон значень, а саме $[0.0, 0.55]$. Діаграми розподілу по класам показують, що представники різних класів дуже схожі між собою за спектральни-

ми характеристиками, зокрема пшениця та рапс. У другому та третьому каналів гістограми усіх класів дуже подібні між собою, за винятком кукурудзи. Розподіли значень різних класів у першому і четвертому каналі мають різні середні значення, хоч і в деякій мірі перекриваються. Можна зробити висновок, що саме ці канали є найбільш важливими для коректної класифікації.

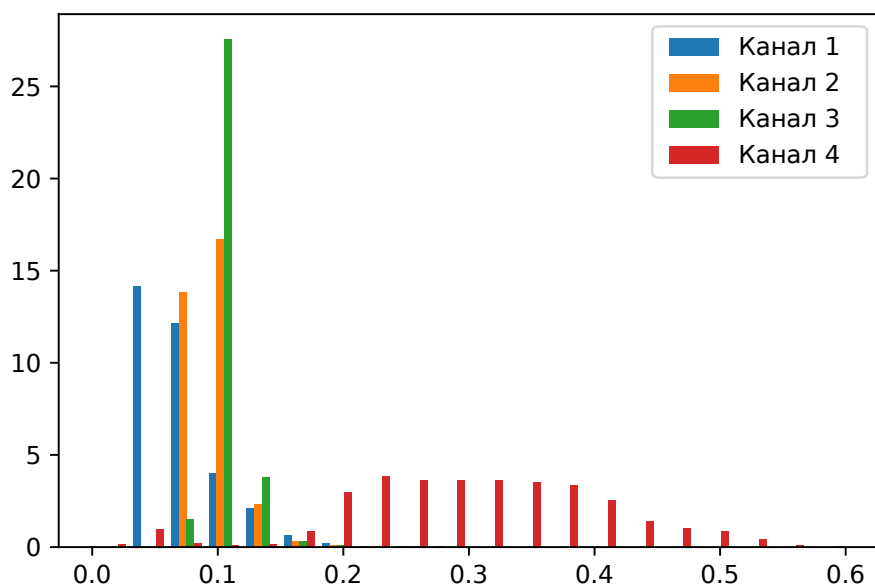
Для зменшення обсягу досліджуваного набору даних з обраних 8 класів випадковим чином обрано 25 000 відповідних йому пікселів на супутниковому композиті. Якщо певний клас включає меншу кількість пікселів, для експерименту візьмемо їх всі.

На рисунку 3.3 представлено візуалізації отриманого датасету. Через неможливість візуалізації чотиривимірних даних на площині, розмірність даних була понижена до двовимірної за допомогою алгоритмів PCA та tSNE.

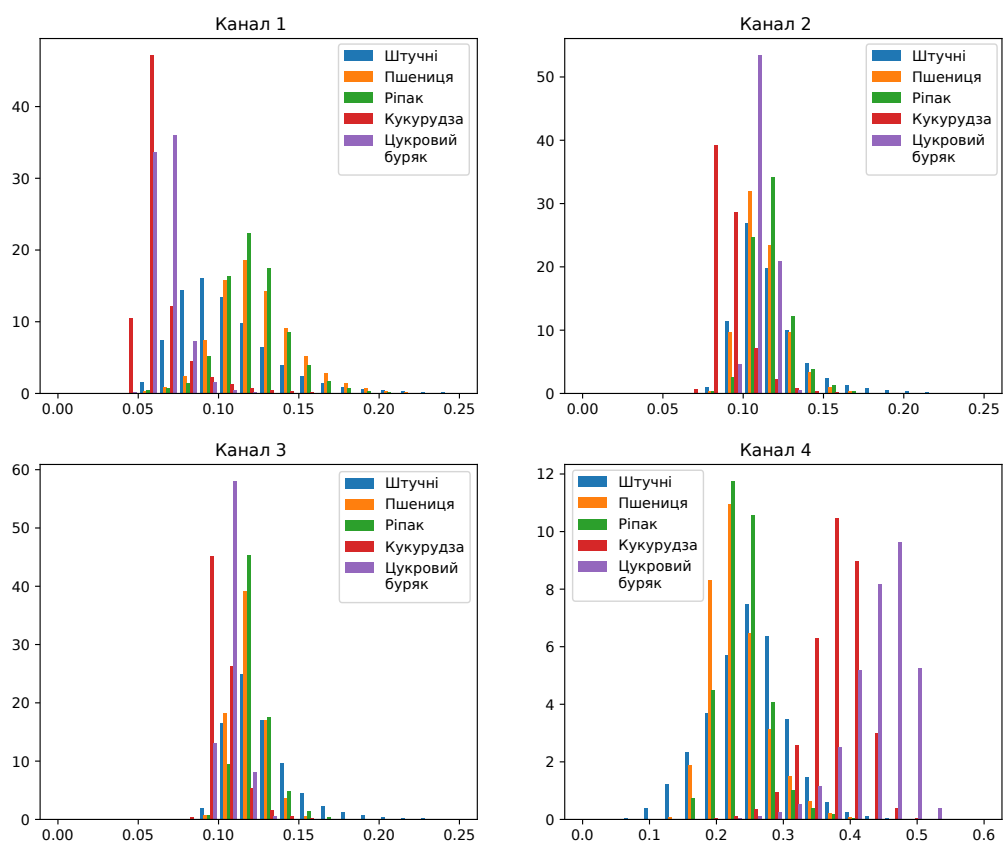
Отриманий датасет містить 9 класів та має проблему перекриття класів. Для візуалізації ми використовуємо редуковані двовимірні дані. Відзначимо, що всі обчислення запропонованого алгоритму виконуються на повних чотиривимірних даних, а сам алгоритм здатний працювати з даними будь-якої розмірності.

У методі визначення неоднозначних екземплярів даних обчислюється $L2$ відстань між різними екземплярами даних у просторі ознак. Для пікселів на зображенні, такими ознаками виступають їх кольори, а саме значення інтенсивності сигналу у різних каналах. Кількість каналів у зображенні рівне 4, тож кожен піксель також матиме 4 ознаки. $L2$ відстань у просторі ознак між двома пікселями $p^{(1)}, p^{(2)}$ обчислюватиметься наступним чином:

$$\|p^{(1)} - p^{(2)}\|_{L2} = \sqrt{(p_{490}^{(1)} - p_{490}^{(2)})^2 + (p_{560}^{(1)} - p_{560}^{(2)})^2 + (p_{665}^{(1)} - p_{665}^{(2)})^2 + (p_{842}^{(1)} - p_{842}^{(2)})^2}$$



(а)



(б)

Рисунок 3.2 — Розподіл значень пікселів по каналах (а) та в залежності від істинного класу (б) для датасету супутникових зображень

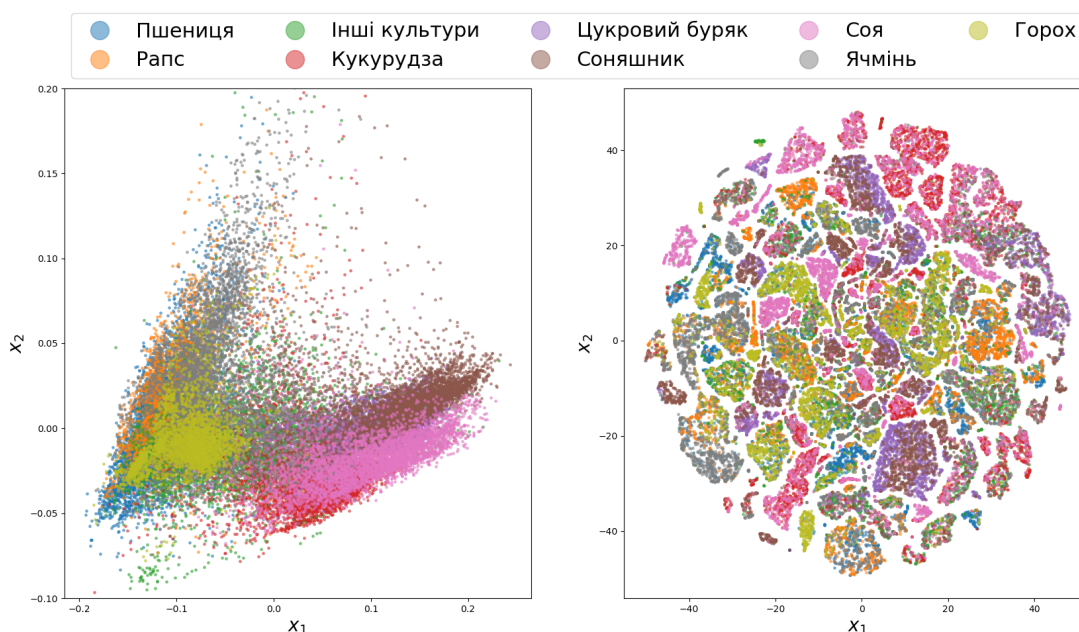


Рисунок 3.3 — Візуалізація датасету за допомогою PCA (а) та tSNE (б)

де 490,560,665,842 – довжини хвиль відповідного каналу пікселя.

За результатами роботи алгоритму на супутниковому композиті визначено неоднозначні пікселі для задачі класифікації за чотирма ознаками. Візуалізацію надійних та неоднозначних точок представлено на рисунку 3.4.

Також обраховано відсоток неоднозначних точок по кожному класу та загалом, дані наведено у таблиці 3.1. Використовувалися такі значення гіперпараметрів: $n = 10$, $r = 5$, $k = 6$, $q = 4$.

Як видно із таблиці 3.1, більшість точок з датасету потрапила до неоднозначних даних, що свідчить про неможливість точної по-піксельної класифікації обраних 8 видів сільськогосподарських культур на основі єдиного знімку. Це узгоджується з емпіричними результатами сучасних досліджень в сфері дистанційного зондування Землі, які свідчать про те, що різні типи сільськогосподарських культур неможливо відрізнити на основі 1 знімку. Тому в сучасних роботах для класифікації сільськогосподарських культур використовуються часові ряди супутникових даних, отриманих протягом всього вегетаційного сезону [65, 66]. Для подолання цієї проблеми необхідно розширити датасет різночасови-

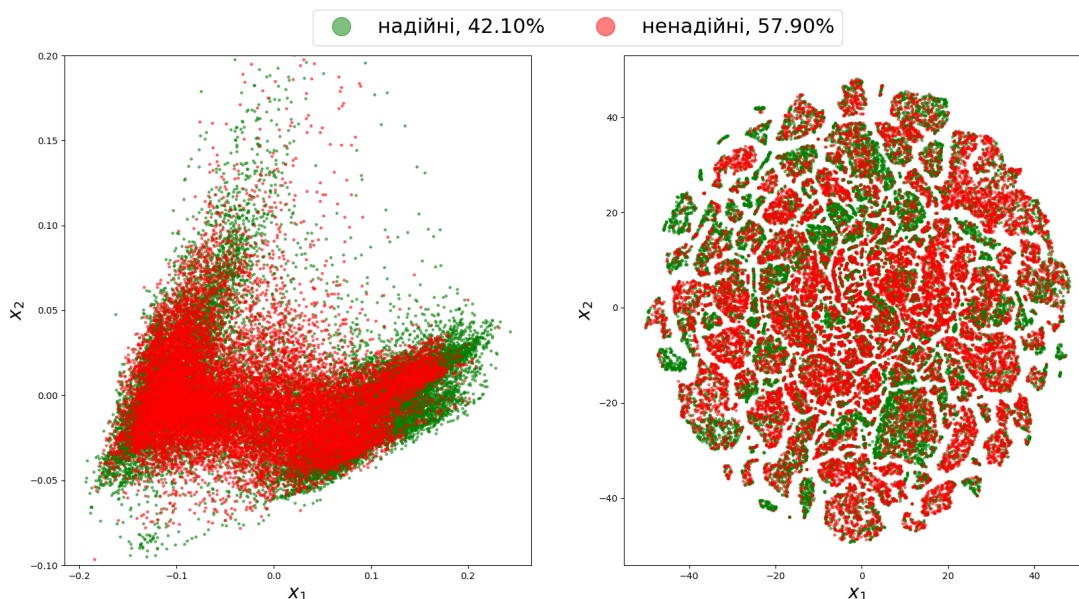


Рисунок 3.4 — Візуалізація визначених неоднозначних точок датасету за допомогою PCA (a) та tSNE (b). Зеленим кольором позначено надійні точки червоним – неоднозначні

ми знімками з додатковими каналами та/або використовувати просторові властивості знімків, наприклад за допомогою згорткових нейронних мереж.

3.2.3 Обробка датасету для задачі сегментації

Для проведення експериментів у задачі сегментації, супутниковий композит поділено на фрагменти фіксованого розміру, а саме 256×256 пікселів кожний. З урахуванням наявності чотирьох каналів, кожен фрагмент можна розглядати як масив розмірністю $[256 \times 256 \times 4]$.

У якості істинних значень класів використовувалося карта класифікації, розроблена Інститутом космічних досліджень НАНУ та ДКАУ [64]. Оригінальна карта налічує 19 класів, однак у цьому експерименті число класів зменшено до 16 шляхом об'єднання деяких культур з класом “Інші культури”.

З усього набору фрагментів вибрано 4 250 екземплярів, які містили в собі

Таблиця 3.1 — Загальне число пікселів та частка неоднозначних даних по кожному класу

Назва культури	Кількість пікселів, тис.	Відсоток неоднозначних даних, %
Пшениця	25	51.36
Ріпак	25	63.17
Кукурудза	25	68
Цукровий буряк	25	78.66
Соняшник	25	61.99
Соя	25	54.88
Ячмінь	25	47.44
Горох	18	60.55
Інші культури	25	35.8
Разом	218	57.90

значну частку пікселів, що відповідали сільськогосподарським культурам. Отримана множина була поділена на тренувальну і тестову вибірку, розміром 2 125 фрагментів кожна. Приклад візуалізації перших трьох каналів фрагменту та відповідної йому маски сегментації зображено на рисунку 3.1б.

3.2.4 Генерація вагових масок

Тестова множина датасету, який використовувався у експерименті, містить у собі 2 125 фрагментів, розміром 256×256 пікселів кожний. Загальна кількість пікселів становить майже 140 мільйонів, з них приблизно 41.89% пікселів сільськогосподарських культур, за таких умов застосування оригінального алгоритму виявлення неоднозначних екземплярів даних не є раціональним рішенням. Час на обробку кожного пікселю можливо значно скоротити, використовуючи K-D дерево для швидкого отримання N найближчих екземплярів даних. Незважаючи на цю оптимізацію, час виконання залишається великим. Окрім того, велика густин-

на даних у просторі ознак може ускладнити підбір параметрів та/або спотворити результати роботи алгоритму.

З огляду на вищевказане, визначення неоднозначних даних було проведено на скороченому датасеті, який було сформовано шляхом випадкового вибору 25 000 пікселів для кожної культури або ж усіх пікселів даної культури, якщо їх було менше за 25 000. У результаті, скорочений датасет містив 218 000 пікселів, які не були пов'язані між собою просторовими зв'язками. За допомогою алгоритму визначення неоднозначних екземплярів даних та K-D дерева, визначено приналежність кожного елементу до надійних або ж неоднозначних даних.

Наступним кроком було визначення неоднозначних даних на повному наборі даних. На скороченому датасеті було побудовано K-D дерево, з якого було отримані найближчі точки скороченого датасету для кожного пікселю повного датасету. Якщо найближча точка була класифікована як неоднозначна на попередньому етапі, то і відповідний екземпляр даних повного датасету також буде вважатися як неоднозначним, і навпаки. Іншим пікселям, які не відносяться до сільськогосподарських культур, присвоюється статус надійних, задля запобігання їх впливу на процес навчання і фінальну точність класифікації культур. Не виключено використання більш складних методик на даному етапі, які можуть присвоювати статус надійного чи неоднозначного екземпляру даних у інший спосіб.

Маючи статус надійності для кожного пікселю сільськогосподарської культури, стає можливим побудувати маску для кожного фрагменту супутникового композиту. Числові значення маскам можна присвоювати двома протилежними підходами. Один з них: більш надійні пікселі отримають більше значення відповідного їм в масці коефіцієнту, менш надійні – менші значення коефіцієнту. Такі маски при обчисленні функцію втрат спонукатимуть модель звертати більше уваги на надійні пікселі та теоретично можуть підвищити стабільність результатів. Інший підхід полягає у наданні неоднозначним пікселям більших коефіцієнтів,

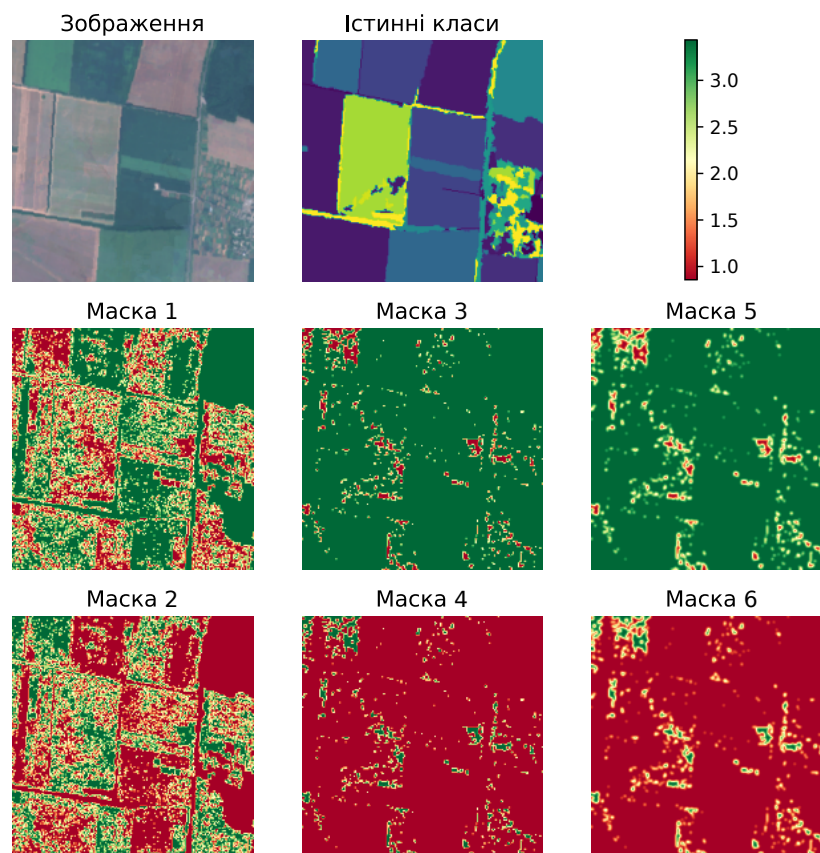


Рисунок 3.5 — Приклади супутникового зображення, його істинної маски сегментації (істинні класи), вагових масок з різних наборів (маска N), цифра відповідає номеру набору N

це теоретично може спонукати модель більш якісно використовувати просторову інформацію з сусідніх пікселів. Приклади таких масок зображено на рисунку 3.5.

Отримані маски було модифіковано шляхом застосування до них гаусового розмиття та морфологічних перетворення: операція звуження (Erosion) та розтягування (Dilation). Такі перетворення мали на меті прибрати поодинокі пікселі та згладити саму маску, щоб зробити переходи між ваговими коефіцієнтами більш плавними. Розміри ядер 3×3 для морфологічних перетворень та 5×5 для гаусового розмиття обрано експериментально. Вони не повинні бути надто малими, оскільки тоді отримані набори масок будуть мало відрізнятися від наборів, в яких

дані перетворення відсутні. А також розміри ядер не повинні бути надто великими, оскільки тоді маски стануть майже однорідними. Після усіх перетворень маска нормалізується таким чином, щоб середнє значення її коефіцієнтів дорівнювала одиниці.

Всього було згенеровано 6 наборів масок, параметри кожного набору представлені в таблиці 3.2, приклади масок представлені на рисунку 3.5.

Таблиця 3.2 — Список перетворень, застосованих до різних наборів вагових масок

№	Назва набору	Фокус вагових коефіцієнтів	Морфологічні перетворення	Гауссове розмиття
1	Normal focus	нормальні пікселі (4:1)	-	-
2	Hardcase focus	неоднозначні пікселі (4:1)	-	-
3	Normal focus Dilation	нормальні пікселі (4:1)	Dilation (3×3)	-
4	Hardcase focus Erosion	неоднозначні пікселі (4:1)	Erosion (3×3)	-
5	Normal focus Dilation Gauss	нормальні пікселі (4:1)	Dilation (3×3)	5×5
6	Hardcase focus Erosion Gauss	неоднозначні пікселі (4:1)	Erosion (3×3)	5×5

3.3 Тренування сегментаційної моделі

Наявна для кожного фрагменту вагова маска використовується при обчисленні функції втрат. У даному випадку використовується сегментаційна модель, яка базується на архітектурі U-Net [67], докладно її структура зображена на рисунку 3.6. Вона приймає на вхід порцію даних розмірністю $N \times 256 \times 256 \times 4$, де N – кількість фрагментів у порції даних, 256×256 – розміри фрагменту, а 4 – кількість каналів. Дана модель має єдиний вихід, на якому розмірність вихідних даних становлять $N \times 256 \times 256 \times 16$, де 16 – кількість класів.

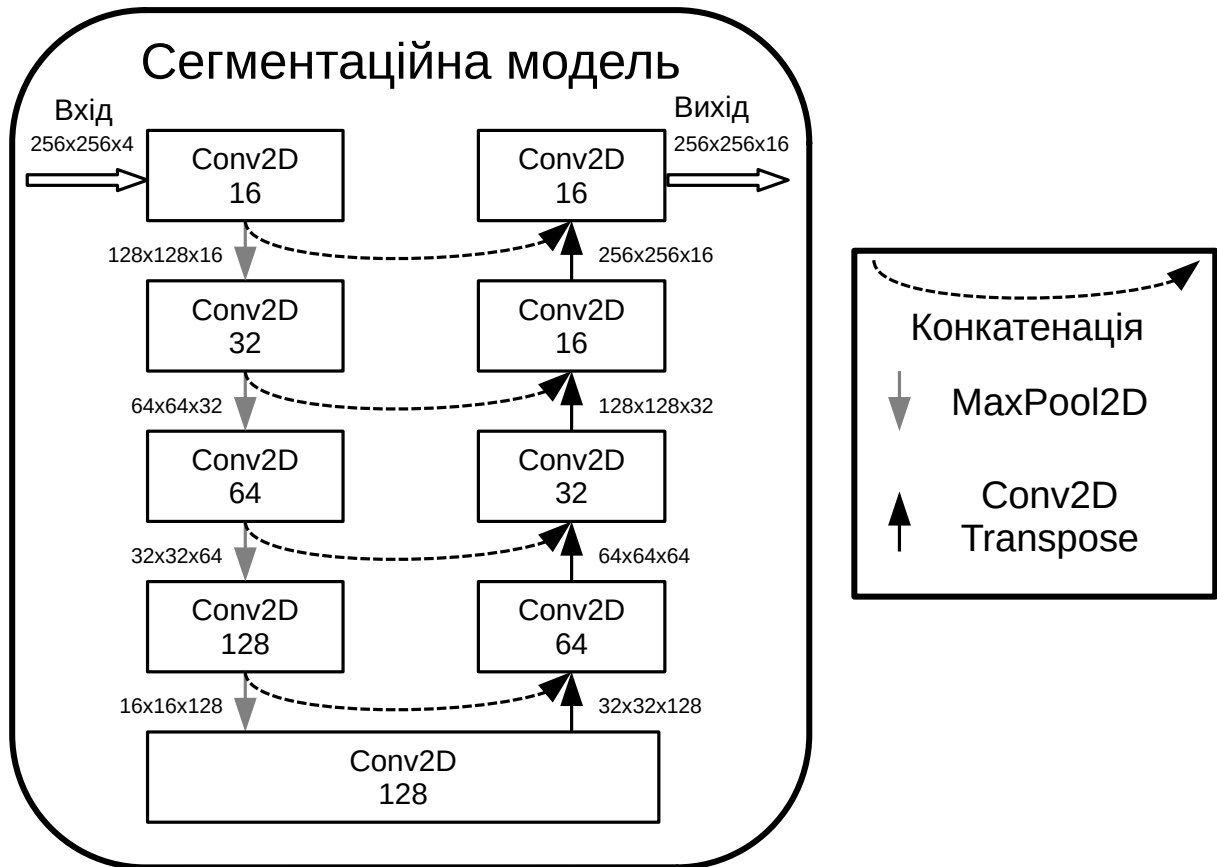


Рисунок 3.6 — Архітектура сегментаційної моделі

Для виконання одного кроку тренування без використання сегментаційних масок, необхідно сформувати порцію даних розмірністю $N \times 256 \times 256 \times 4$ та подати на вхід моделі. У результаті буде отримано дані розмірністю $N \times 256 \times 256 \times 16$, які необхідно порівняти з істинними даними розмірністю $N \times 256 \times 256$ за допомогою функції втрат. Для задачі сегментації у якості функції втрат використовується крос-ентропія [68]:

$$BCE(\hat{Y}, \hat{Y}') = -\frac{1}{n} \sum_{i,j} \sum_{k=1}^C \tilde{Y}_{ijk} \log Y'_{ijk}, \tilde{Y}_{ijk} = \mathbb{1} \{Y_{ij} = k\}$$

де \hat{Y}, \hat{Y}' – одиничні сегментаційні маски, істинна і передбачена відповідно, \tilde{Y} – розріджене представлення істинної маски сегментації \hat{Y} , n – кількість пікселів в одному фрагменті, C – кількість класів.

Модифікація цього процесу для використання просторових вагових масок виглядає наступним чином. При формування чергової порції даних, в наявності є не лише власне дані та істинні класи але й вагові маски розмірністю $N \times 256 \times 256$. Перша частина цього алгоритму залишається без змін і завершується отриманнями вихідних даних розмірністю $N \times 256 \times 256 \times 16$. Для обчислення функції втрат використано значення масок, які слугують ваговими коефіцієнтами:

$$BCE(\hat{Y}, \hat{Y}', \hat{M}) = -\frac{1}{n} \sum_{i,j} M_{ij} \sum_{k=1}^C \tilde{Y}_{ijk} \log Y'_{ijk}$$

У результаті, регулюється вплив кожного пікселя на значення функції втрат, відповідно, значення градієнтів параметрів моделі, що навчається. Пікселі, яким відповідають більші значення пікселів вагової маски, матимуть більший вплив на навчання моделі та її ваги.

З метою визначення оптимальних параметрів генерації просторових вагових масок, було проведено декілька експериментів, кожний з них використовував різні вагові маски; а також один експеримент, який використовував виключно реальні дані.

3.4 Результати та обговорення

Усього було натреновано 7 сегментаційних моделей, до числа яких входить базова сегментаційна модель, яка була натренована без використання вагових масок, а також 6 моделей, під час навчання яких було використано відповідний набір вагових масок з таблиці 3.2.

Для моделей, що використовували лише реальні дані, обраховано метрики, такі як точність, влучність (P), повнота (R) та перетин через об'єднання (IoU). Лише частина моделей продемонструвала покращення метрик у порівнянні з базовою моделлю, дві найкращі з них наведені у таблиці 3.3. Вагові маски цих мо-

делей мали більше значення коефіцієнтів для нормальних пікселів, ніж для неоднозначних, а також до масок застосована операція розтягування з ядром 3×3 , що зменшило кількість крайових неоднозначних пікселів. До набору масок останньої моделі також було застосовано гаусове розмиття з ядром 5×5 . Для більшості класів ці моделі показують кращий результат ніж базова модель, особливо для культур які мало представлені у датасеті, наприклад для цукрового буряку та сої. Загальні метрики виросли приблизно з 77.3% до 78.7% та з 64.1% до 66.3% для точності та IoU відповідно.

Таблиця 3.3 — Порівняння метрик моделей класифікації с/г культур (найвищі метрики виділено жирним шрифтом)

Назва с/г культури	Базова модель			Модель Normal focus Dilation			Модель Normal focus Dilation Gauss		
	Метрики								
	R (%)	P (%)	IoU	R (%)	P (%)	IoU	R (%)	P (%)	IoU
Пшениця	83.5	71.5	62.7	86.6	71.8	64.6	84.2	72.4	63.8
Рапс	29.3	37.7	19.7	3.1	42.9	3.0	17.8	49.5	15.1
Кукурудза	89.9	83.8	76.6	89.5	87.6	79.5	89.2	87.8	79.4
Цукровий буряк	0.0	0.0	0.0	52.0	54.5	36.3	31.7	62.1	26.6
Соняшник	92.7	80.5	75.7	89.7	86.3	78.5	92.8	82.3	77.4
Соя	54.9	80.8	48.6	62.4	81.3	54.6	66.0	78.4	55.8
Інші культури	3.2	26.6	2.9	7.4	33.2	6.5	7.4	35.8	6.5
Ячмінь	22.4	45.8	17.7	22.9	50.2	18.7	26.6	45.9	20.3
Горох	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ААС (%)	41.8	47.4	33.8	46.0	56.4	38.0	46.2	57.1	38.3
ОА (%)	77.3		64.1	78.7		66.3	78.6		66.2

Для тренування сегментаційних моделей кращий результат дають маски, де нормальним пікселями присвоєно більші вагові коефіцієнти. Важливо зменшити кількість неоднозначних пікселів, які межують з надійними, шляхом застосування морфологічних операцій. Гаусове розмиття маски є опціональним, воно може як покращити результат, так і не вплинути на нього.

Також було навчено модель на 12-канальному супутниковому композиті,

який було створення конкатенацією трьох 4-канальних композитів за три різні місяці (квітень, червень та липень). Сегментаційна модель, навчена на цьому датасеті, продемонструвала загальну точність на тестовому наборі даних у 81.5%. Збільшення кількості каналів, а отже і ознак, у три рази, збільшило загальну точність на 4.2%. При цьому датасет також збільшився у три рази, як і витрати на його отримання, що може бути проблемою для комерційних супутникових даних. Використання методу вагових масок дозволяє отримати приріст точності +1.4% абсолютно безкоштовно, а також кращі метрики для мало представлених класів.

Висновки до розділу 3

У розділі розглянуто проблему незбалансованості класів у навчальній вибірці для задачі сегментації та вперше запропоновано використання вагових масок для модифікації процесу навчання моделі. Даний підхід не потребує збирання і розмітки додаткових даних. Проведено порівняння розробленого методу зі стандартною процедурою навчання моделі на тих самих даних і тією самою архітектурою моделі. Це дає змогу стверджувати, що зміна метрик відбулася саме за рахунок використаного нового методу, а не через інші дані чи іншу архітектуру моделі. За результатами порівняння, використання розробленого методу збільшує точність розпізнавання для мало представлених класів, таких як цукровий буряк та соя, і незначно збільшує загальні метрики, а саме дає приріст на 1.4% для точності та 2.2% для IoU. Для порівняння, використання композиту з втричі більшим числом каналів, збільшує загальну точність на 4.2%.

Для даної задачі найкращі результати давали маски, де більш надійним пікселям присвоювалися більші значення вагових коефіцієнтів. Виділення крайових неоднозначних пікселів за допомогою морфологічних перетворень та присвоєння їм статусу надійних покращувало результат.

Не знайдено зв'язку між присвоєнням статусу неоднозначного екземпляру даних для пікселів які відповідають артефактам на зображеннях. Явне виявлення таких артефактів можна включити до майбутніх модифікацій даного підходу.

4 ЗАСТОСУВАННЯ GAN-АУГМЕНТАЦІЇ У ОБЛАСТІ ДИСТАНЦІЙНОГО СПОСТЕРЕЖЕННЯ

Незбалансованість навчального набору даних також може бути вирішена шляхом додавання штучно створених зображень до датасету (табл. 1.1). Згенеровані зображення мають успадковувати такі самі властивості, в тому числі текстурні, як і реальні зображення.

З урахуванням потенціалу моделей GAN в дистанційному спостереженні, у цьому розділі досліджено новий метод застосування GAN для вирішення питань дисбалансу даних у класифікації покриття землі та типів сільськогосподарських культур. Представлено першу методологію аугментації супутникових зображень на основі GAN, використовуючи підхід ріх2ріх [8], тим самим вперше вводячи цей метод аугментації в галузь дистанційного спостереження.

Математичне формулювання задачі генерації штучних зображень з підходом ріх2ріх GAN наступне: нехай \hat{X} , \tilde{X} – реальне та штучне зображення відповідно, розмірністю $[W \times H \times Ch]$ та числом каналів Ch , а \tilde{Y} – бажана маска сегментації розмірністю $[W \times H]$ з числом класів C , яка є умовою для генерації. Необхідно знайти таку модель генератора $G : \mathbb{R}^{W \times H \times C} \rightarrow \mathbb{R}^{W \times H \times Ch}$, що:

$$\hat{X}' = G(\tilde{Y}), \sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^{Ch} |X_{ijk} - X'_{ijk}| \rightarrow \min$$

Та знайти таку модель дискримінатора $D : \mathbb{R}^{W \times H \times Ch} \rightarrow \mathbb{R}$, що:

$$D(\hat{X}) = 1, D(\hat{X}') = 0$$

Варто зауважити, що GAN-аугментація – не єдиний метод генерації штучних зображень по заданій сегментаційній масці \tilde{Y} , З огляду на це, у цьому розділі також виконується порівняння з іншими підходами аугментації, що базуються

на відмінних методах генерації штучних зображень, такі як статистичний метод та метод випадкового відбору пікселів.

4.1 Генеративна змагальна мережа

Ефективна архітектура GAN дозволяє генерувати реалістичні синтетичні супутникові зображення для аугментації навчальних даних. Взаємозв'язок між супутниковими зображеннями та сегментаційними масками ґрунтується на текстурних та багатоспектральних ознаках у межах зображень. Модель, здатна генерувати реалістичні супутникові зображення для будь-якої штучно створеної сегментаційної маски, може вирівняти розподіл класів у датасеті. Цей метод дозволяє створювати синтетичні пари супутникових зображень і масок, генерувати розподіли сільськогосподарських культур, які не зустрічаються у природному середовищі, та надає можливість керування балансом класів в наборі даних [69]. У цьому дослідженні використано сегментаційні маски розміром 256×256 пікселів для створення синтетичних супутникових зображень з 4 каналами.

Для створення моделі генератора використано техніку трансляції зображень в зображення [70], яка спрямована на перетворення зображень з одного домену в інший із відмінною розмірністю простору ознак. Основною ідеєю, яка лежить в основі GAN аугментації, є навчання моделі у дві стадії. Спочатку pix-2-pix GAN модель навчається генерувати реалістичні супутникові зображення на основі довільної сегментаційної карти. Потім модифікуються реальні сегментаційні карти, щоб створити штучні маски з більшою часткою менш представлених класів. Після цього початкові супутникові дані і згенеровані об'єднуються в єдину навчальну колекцію даних, яка далі використовується для навчання сегментаційної моделі з вищою точністю розділення менш представлених класів. Схема аугментаційного алгоритму показано на рисунку 4.1.

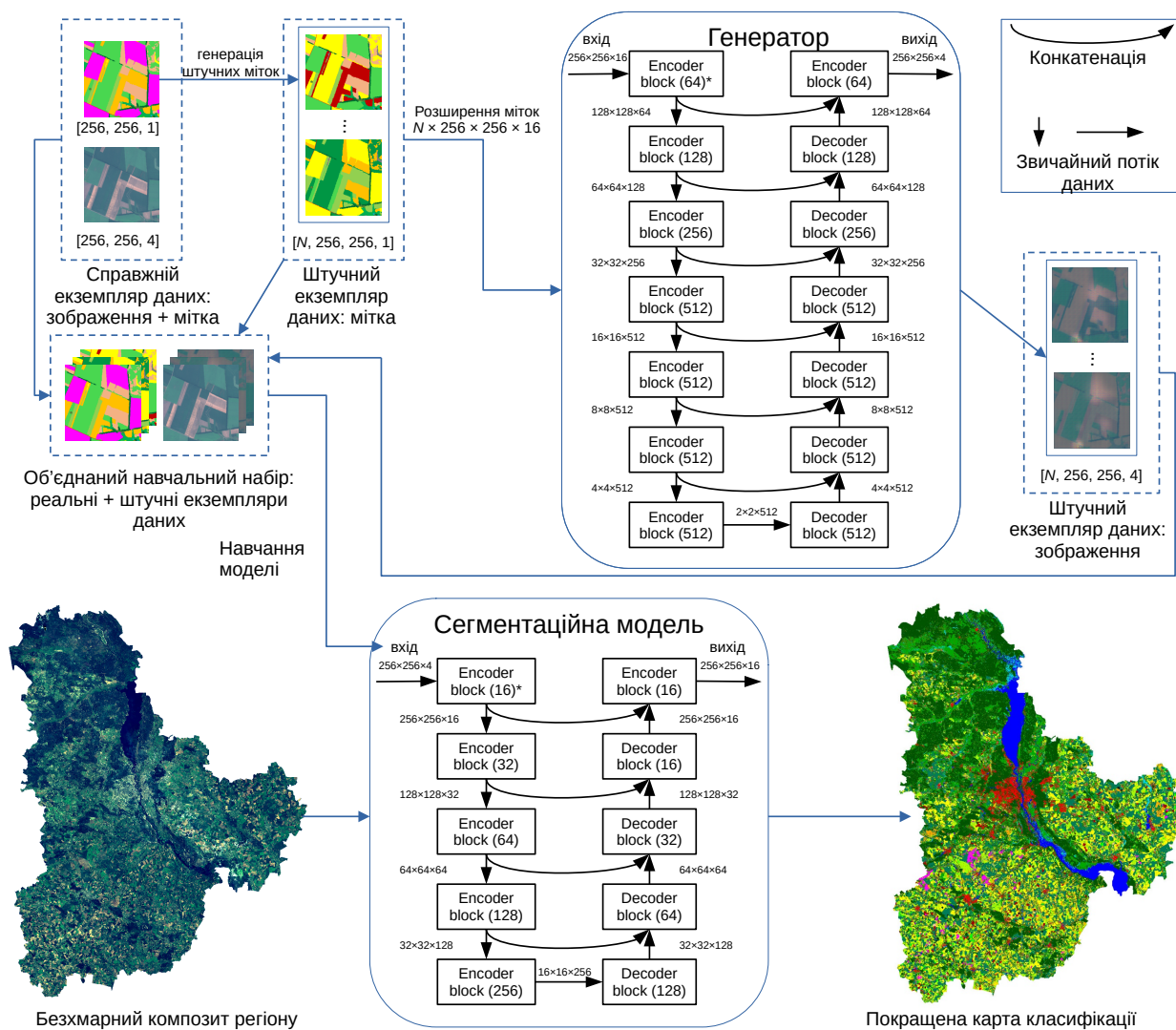


Рисунок 4.1 — Схема GAN-аугментації

Генеративна змагальна мережа (GAN) для генерації супутникових даних складається з двох моделей нейронних мереж: генератора, який створює зображення на основі сегментаційних масок, і дискримінатора, який відрізняє реальні та створені зображення (рисунок 4.2). Генератор вчиться створювати реалістичні фрагменти супутникових зображень, які не може розрізнити дискримінатор. Тим часом, дискримінатор вчиться розрізняти реальні та штучні зображення. Обидві моделі працюють з вхідними даними розміром 256×256 пікселів.

Для генератора використано модель pix2pix на основі U-Net із 7 блоками

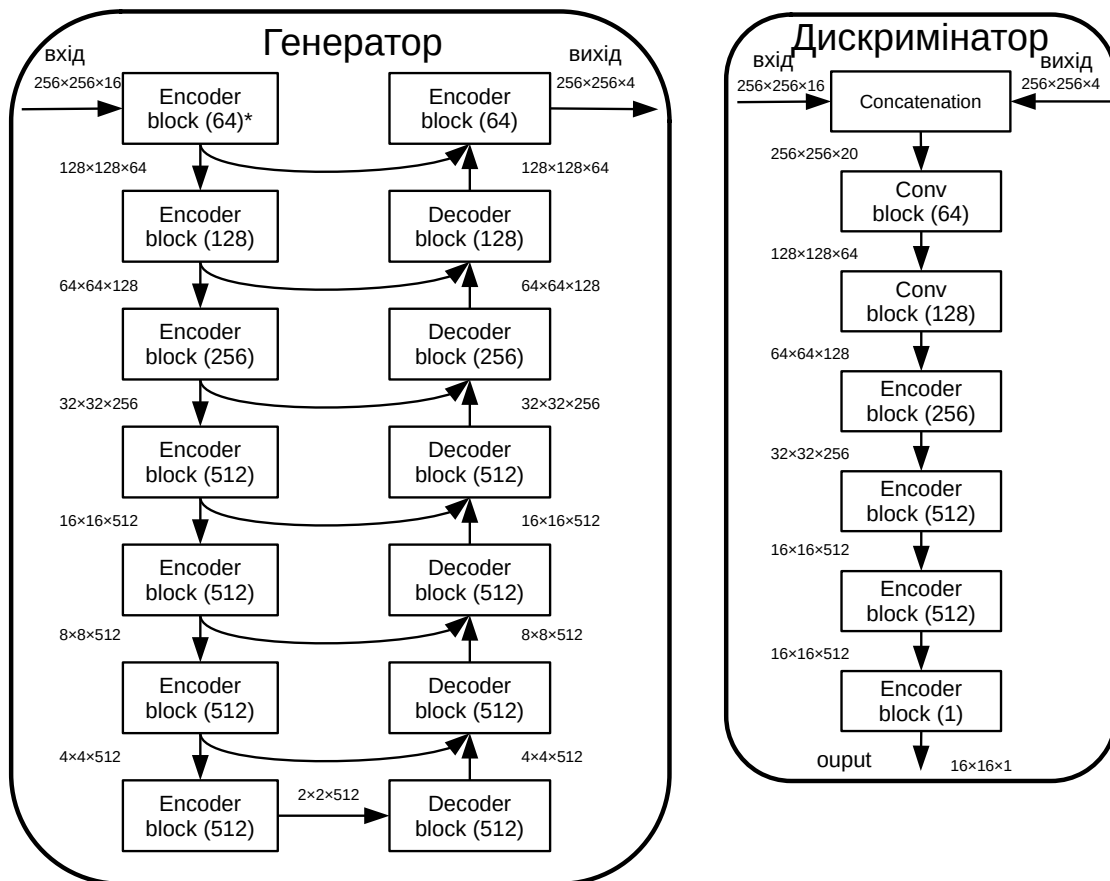


Рисунок 4.2 — Архітектура сегментаційної моделі

енкодера і декодера, в яких присутні 6 обхідних з'єднань [71] та гіперболічний тангенс як активаційна функція на вихідному шарі. Для дискримінатора використано глибоку згорткову нейронну мережу з 5 шарами 2D-згорток, включаючи нормалізацію пакету та активаційні функції Leaky ReLU, а також вихідний згортковий шар із сигмоїдною активаційною функцією (рисунок 3.6). Функцію втрат дискримінатора оптимізовано за допомогою бінарної крос-ентропії:

$$BCE = \frac{1}{n} \sum_{i=1}^n [y_i \times \log_2(\hat{y}_i) + (1 - y_i) \times \log_2(1 - \hat{y}_i)]$$

де n – кількість екземплярів у навчальному пакеті, y_i – мітка екземпляру i та \hat{y}_i – вихід дискримінатора для даного фрагменту i . Функція втрати генератора комбінує бінарну крос-ентропію дискримінатора та середньоквадратичну помил-

ку між згенерованим (g_i) та реальним (r_i) зображеннями за формулою:

$$Loss_{gen} = MSE(r_i, g_i) + 0.01 \times BCE(r_i, g_i)$$

Для оптимізації параметрів під час навчання використано алгоритм Adam [72] з швидкістю навчання рівною 2×10^{-4} та коефіцієнтом β рівним 0.5.

4.2 Генерація штучного датасету

Для використання штучних зображень, що згенеровані за допомогою GAN, у процесі навчання сегментаційної моделі, необхідно отримати штучні сегментаційні маски, модифікуючи початкові маски із навчального набору даних. Цей підхід зберігає просторові шаблони сільськогосподарських полів подібно до таких у реальному середовищі. Однак, навіть маючи навчену $pix2pix$ GAN-модель, для генерації принципово нових супутникових знімків з іншим розподілом класів необхідно створити нові сегментаційні карти. Використання вже наявних карт не несе в собі ніякої користі – по-перше розподіл класів так і залишиться незмінним, по-друге – GAN-модель вчилася на тих же даних, тому з високою ймовірності згенерує зображення, подібні до вже наявних у датасеті.

4.2.1 Створення нових сегментаційних масок

Об'єкти та території на супутникових знімках мають певну форму та/або структуру. Наприклад, поля, засіяні сільськогосподарськими культурами зазвичай мають певну типову форму, яка значно відрізняється від типової форми для стежок, доріг чи водних об'єктів. Генерація нових сегментаційних масок має врахувати ці особливості, а також в жодному разі така маска не може бути просто випадковим набором пікселів. У даній ситуації, найбільш логічним виглядає рі-

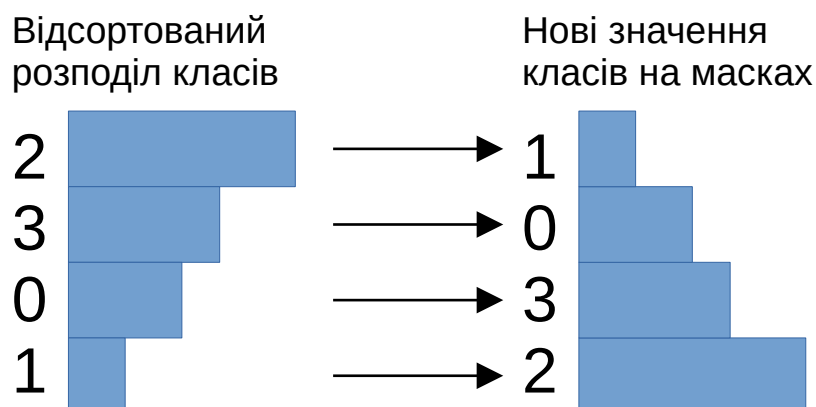


Рисунок 4.3 — Схема балансування класів шляхом генерації нових сегментаційних масок

шення поміняти місцями різні класи сільськогосподарських культур, без жодних змін для інших класів. Таким чином, нові сегментаційні маски будуть відрізнятися від справжніх лише тим, що в полях наявний інший розподіл культур.

Спочатку необхідно зібрати статистику датасету, а саме підрахувати загальне число пікселів для кожного класу. Результати слід відсортувати за частотою та створити словник заміन. Він замінює найбільш поширений клас на найменш поширений, другий найбільш поширений на другий найменш поширений і так далі. Нарешті, замінюються всі пікселі класу сільськогосподарських культур на нові згідно з словником. Число пікселів, які займають найменш представлені класи збільшується, і стають одного порядку з числом пікселів найбільш представлених класів. Схема даного перетворення показана на рисунку 4.3. Якщо ж у об'єднаному датасеті зі згенерованих та реальних даних все одно наявний дисбаланс класів, то слід повторити дану процедуру. Таким чином, вирішується проблема дисбалансу навчальних даних.

Для датасету супутникових знімків Київської області даний алгоритм мав дві ітерації, створено 2 384 нових масок, та відповідно, стільки ж нових знімків.

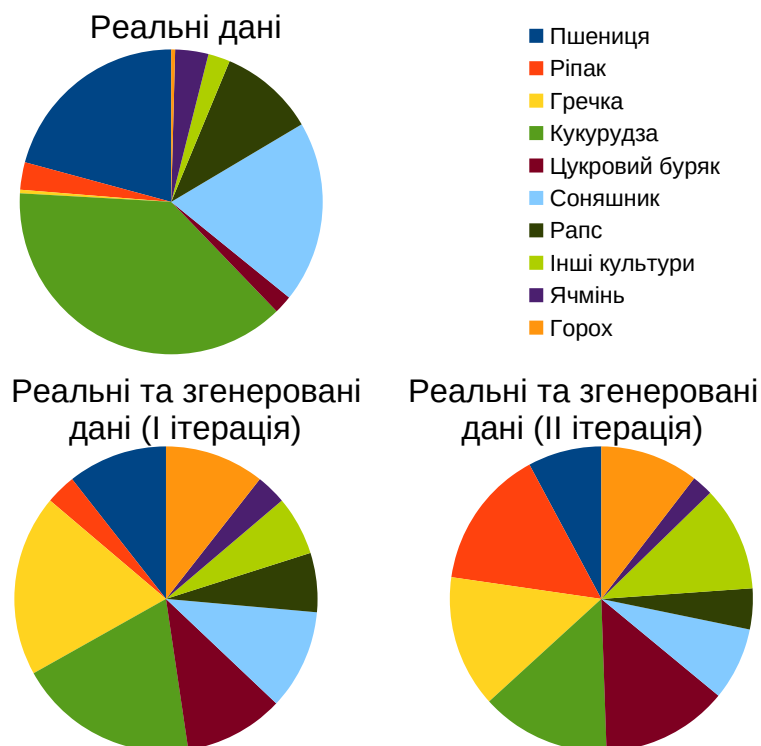


Рисунок 4.4 — Розподіл класів у датасеті з реальними знімками, та у датасеті з реальними та згенерованими даними після першої та другої ітерацій алгоритму

Це число менше за 4 250, дві ітерації мали б потроїти датасет. Це пов'язано з відсутністю будь-яких сільськогосподарських культур на частині зображень з датасету. Такі маски залишалися незмінними та не брали участь у подальшій генерації нових зображень. Розподіл класів датасету до, та після генерації нових зображень, показано на рисунку 4.4.

4.2.2 Класичний підхід до аугментації

Для тих же самих сегментаційних масок застосовані класичні методи створення штучних зображень [8] для порівняння з вищеописаним підходом. Перший

метод використовує метод Монте-Карло для оцінки статистичних характеристик кожного класу та генерації синтетичних сегментаційних масок. Синтетичні супутникові зображення заповнюються, по черзі вибираючи пікселі на основі ймовірного розподілу вибраного класу сільськогосподарських культур у багатоспектральному просторі ознак. Цей підхід, який загалом використовується в завданнях глибокого навчання, допомагає контролювати розподіл даних та запобігає перенавчанню, надаючи набір сегментаційних масок та синтетичних супутникових зображень для навчання поряд із реальними даними.

Другий метод, запропонований [73], включає в себе вибір пікселів безпосередньо з реальних зображень за допомогою тієї ж штучної сегментаційної маски, що і попередніх методах. Для заповнення нового зображення, випадковим чином вибираються декілька екземплярів реальних даних, які містять в собі бажані класи. На обраних зображеннях має бути достатня кількість відповідних пікселів сільськогосподарських культур. Потім пікселі з реальних зображень випадково переносяться на штучне зображення, у ті місця, які відповідають їхньому класу. Цей підхід дозволяє зберігати стабільність значень пікселів шляхом вибору з одного класу в малій області, враховуючи незначні відмінності в властивостях пікселів на різних регіонах.

У підсумку, для кожного підходу було згенеровано по 2 384 нових зображень, що дає у результаті три об'єднаних датасети по 4 509 зображень в кожному.

4.3 Тренування сегментаційної моделі з GAN-аугментацією

Як базову модель для нашого завдання семантичної сегментації, використано класичну модель U-Net із входним розміром $256 \times 256 \times 4$ пікселів. U-Net – це широко використовувана архітектура глибокого навчання в завданнях моніторингу земель [67]. Її ключовим інноваційним рішенням є включення додаткових з'єд-

нань від шарів, які зменшують розмір простору ознак до відповідних шарів, які розширюють його. Цей дизайн дозволяє виділяти високорівневі ознаки на кожному кроці, зберігаючи при цьому просторову інформацію в зображенні. Крім того, це допомагає вирішувати проблему зникнення градієнтів у певних сценаріях.

У класичній моделі U-Net, запропонованій [67], кожен блок енкодера або декодера містить два 3×3 згорткових шари, за якими слідує активація ReLU. Водночас кількість каналів подвоюється на кожному блоку енкодера (крім першого, де є 64 канали), і на кожному блоку декодера вона зменшується вдвічі. Max-pooling із кроком 2 в блоках енкодера використовується для зменшення розміру карт ознак, тоді як декодер використовує зворотні згортки для масштабування. Вихідний шар мережі – це згортка 1×1 із функцією активації сигмоїду.

Для навчання моделі сегментації U-Net для класифікації сільськогосподарських культур використовувався алгоритм оптимізації Adam із швидкістю навчання 2×10^{-4} та функцію втрати focal loss з параметром гамма рівним 2. Її було обрано, оскільки вона є дуже популярною і, у більшості випадків, найбільш ефективною функцією втрат в завданнях з незбалансованими наборами даних.

4.4 Результати порівняння різних методів аугментації та обговорення

У нашому експерименті створено 2 384 синтетичних супутникових зображень із штучними масками, використовуючи як методи GAN, так і класичні методи. Навчено чотири моделі: (I) із використанням реальних супутникових даних; (II) комбінуванням реальних супутникових даних із синтетичними зображеннями зі статистичного методу; (III) із синтетичними зображеннями використовуючи метод вибірки пікселів; (IV) із синтетичними зображеннями з методу GAN-аугментації. Окрім того, застосовано стандартні аугментації, такі як обертання та відображення. Навчені чотири моделі оцінено на незалежному наборі даних для

тестування, який складається із 2 125 реальних зображень, та жодним чином не був використаний у попередніх етапах. Результати представлені в таблиці 4.1 для порівняння.

Модель (I) досягла 77.3% загальної точності (Overall accuracy, OA) і 64.1% перетину над об'єднанням (Intersection over Union, IoU). Однак середні точності для класів сільськогосподарських культур (Average accuracy, AAC), є дуже низькими через великий дисбаланс представлення класів сільськогосподарських культур у реальних даних. Це справедливо і для інших метрик, таких як точність користувача (User Accuracy, UA), точність виробника (Producer Accuracy, PA) та IoU. Генерація нових зразків із використанням статистичної моделі покращила представленість малих класів у наборі даних і не спричинила перенавчання. В результаті точність для більшості класів моделі (II) не змінилася, але точність малих класів значно зросла, що призвело до підвищення AAC на 8.5% для UA, 5.3% для PA і 9.0% для IoU. Метод вибірки пікселів в цілому показує схожі результати, і AAC для малих класів моделі (III) збільшилося на 6.7% для UA, 6.2% для PA і 8.0% для IoU.

Модель (IV), навчена за допомогою запропонованої методології аугментації GAN, випередила модель (I), модель (II) та модель (III). В порівнянні з моделлю (II) середні значення UA зросли на 2.7%, PA на 1.1% і IoU на 1.2%. Також OA та загальний IoU збільшилися на 1.4% та 1.6% відповідно.

Точність UA та PA для 4 із 8 класів сільськогосподарських культур та IoU для 7 із 8 класів у моделі (IV) є найвищими у експерименті. Є лише 2 винятки, де показники інших моделей значно вищі, ніж у моделі (IV): це кукурудза, де модель (IV) має найкращий результат лише за PA та IoU, та цукровий буряк, де модель (IV) випередила модель (III) лише за UA. Загалом аугментація GAN покращила продуктивність класифікації сільськогосподарських культур на 11.2%, 6.4% і 10.2% для середніх значень UA, PA та IoU відповідно, при цьому OA та загальний IoU покращилися на 1.5% і 2.1%.

Таблиця 4.1 — Порівняння метрик моделей класифікації с/г культур (найвищі метрики виділено жирним шрифтом)

Назва с/г культури	Model (I)			Model (II)			Model (III)			Model (IV)		
	Метрики											
	R,%	P,%	IoU	R,%	P,%	IoU	R,%	P,%	IoU	R,%	P,%	IoU
Пшениця	71.5	83.5	62.7	72.2	83.2	63.1	72.7	82.4	63.0	73.2	85.8	65.3
Ріпак	37.7	29.3	19.7	42.7	32.1	22.4	44.4	31.9	22.8	52.0	37.1	27.6
Кукурудза	83.8	90.0	76.6	83.9	89.4	76.3	82.6	89.6	75.4	87.6	88.2	78.4
Цукровий буряк	0.0	0.0	0.0	61.2	33.8	27.8	50.2	47.9	32.4	65.8	15.5	14.3
Соняшник	80.4	92.7	75.7	83.5	91.0	77.1	82.8	89.6	75.5	82.8	92.9	77.9
Соя	80.8	54.9	48.6	73.4	60.7	49.8	76.7	55.1	47.1	73.0	68.2	54.5
Інші культури	26.6	3.2	2.7	32.7	4.9	41.0	30.1	6.9	34.7	33.2	9.7	43.8
Ячмінь	45.8	22.4	17.7	44.4	22.9	17.8	40.1	22.3	16.7	48.2	30.0	22.7
ААС (%)	53.3	47.0	37.9	61.8	52.3	46.9	59.9	53.2	45.9	64.5	53.4	48.1
ОА (%)	77.3		64.1	77.4		64.6	77.0		64.0	78.8		66.2

На рисунку 4.5 показана візуалізація отриманих карт класифікації на основі моделей (I) – (IV). Можна відзначити, що для основних класів, таких як кукурудза, пшениця та соняшник, які є більш представленими, якість в обох картах сегментації є високою. Однак для менш представленого класу – ріпак (темно-червоний) результат, отриманий моделлю (I), має дефекти та артефакти та набагато меншу якість порівняно з моделлю (IV).

Цей результат можна пояснити тим, що GAN, на відміну від класичних статистичних методів генерації, дозволяє створити штучні приклади, які збережуть подібність не лише в точкових статистичних характеристиках класів, але й у їхніх текстах.

Цікавим виглядає порівняння метрик моделі, що натренована з використанням штучно згенерованих зображень, та моделі, яка вчилася на реальних даних з використанням вагових масок (таблиця 4.2). GAN-модель у деяких випадках краще за модель з масками, у деяких випадках ситуація обернена. Можна стверджувати, що приріст точності у цих моделей зіставний, однак використання вагових

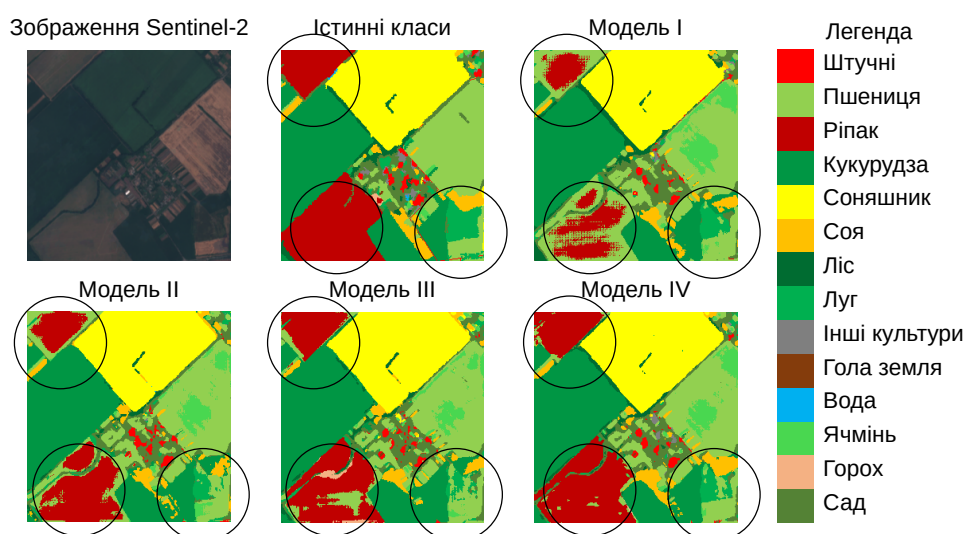


Рисунок 4.5 — Порівняння істинних масок сегментації з результатами роботи моделей

масок не потребує додаткового етапу навчання проміжної моделі та генерації нових даних, тому використання вагових масок є дешевшою альтернативою GAN аугментації.

4.5 Використання вагових масок для покращення GAN-аугментації

Генеративна модель використовується для створення розширеного датасету для тренування сегментаційної моделі, з метою зменшити наявний дисбаланс класів [1]. В свою чергу, в процесі тренування генеративної моделі можливо застосувати вагові маски. На рисунку 4.2 зображена архітектура моделі, заснована на U-Net [67]. Генератор та його вихід розмірністю $256 \times 256 \times 16$ є основним результатом навчання даного комплексу. Двокомпонентна функція втрат GAN складається з середньоквадратичної похибки (MSE) та бінарної крос-ентропії (BCE). Перший компонент відповідає за генерацію зображень максимально схожих на реальні картинки. Другий компонент відповідає за схожість на реальні зображення за

Таблиця 4.2 — Результати порівняння метрик моделі, навченої на реальних та згенерованих даних, та моделі, навченої тільки на реальних даних з використанням вагових масок, з метриками базової моделі (найвищі метрики виділено жирним шрифтом)

Назва с/г культури	Базова модель			Базова модель GAN			Модель Normal focus Dilation		
	Метрики								
	R (%)	P (%)	IoU	R (%)	P (%)	IoU	R (%)	P (%)	IoU
Пшениця	83.5	71.5	62.7	85.8	73.2	65.3	86.6	71.8	64.6
Ріпак	29.3	37.7	19.7	37.1	52.0	27.6	3.1	42.9	3.0
Кукурудза	89.9	83.8	76.6	88.2	87.6	78.4	89.5	87.6	79.5
Цукровий буряк	0.0	0.0	0.0	15.4	65.8	14.3	52.0	54.5	36.3
Соняшник	92.7	80.5	75.7	92.9	82.8	77.9	89.7	86.3	78.5
Соя	54.9	80.8	48.6	68.2	73.1	54.5	62.4	81.3	54.6
Інші культури	3.2	26.6	2.9	9.7	33.2	8.1	7.4	33.2	6.5
ЯчмінЬ	22.4	45.8	17.7	30.0	48.2	22.7	22.9	50.2	18.7
ААС (%)	41.8	47.4	33.8	47.5	57.9	38.8	46.0	56.4	38.0
ОА (%)	77.3		64.1	78.6		66.2	78.7		66.3

своєю структурою і неможливість розрізнення реальних і згенерованих картинок за допомогою дискримінатора. Фінальна функція втрат є зваженою сумою двох компонентів:

$$\begin{aligned}
 Loss &= MSE(\hat{R}, \hat{G}) + \alpha BCE(\hat{Y}, \hat{Y}'), \\
 MSE(\hat{R}, \hat{G}) &= \frac{1}{n} \sum_{i,j,k} (R_{jik} - G_{ijk})^2 \\
 BCE(\hat{Y}, \hat{Y}') &= -\frac{1}{m} \sum_{i,j} \sum_{k=1}^2 \tilde{Y}_{ijk} \log Y'_{ijk}
 \end{aligned}$$

де α – коефіцієнт, що відповідає за відносну вагу двох функцій, у експерименті приймався за 0.01, n – кількість пікселів у зображенні помножене на число

каналів, \hat{R}, \hat{G} – реальне та згенероване зображення, \hat{Y}', \hat{Y} – відповідь дискримінатора та аналогічні за розмірностями істинні значення, m – кількість пікселів у фрагменті зображення.

Доданок, який відповідає за BCE , важко пристосувати до використання вагових масок через різні розмірності по ширині та висоті. Натомість, для доданку, який відповідає за MSE , це зробити набагато легше, тому перепишемо його у наступному вигляді:

$$MSE(\hat{R}, \hat{G}, \hat{M}) = \frac{1}{n} \sum_{i,j} M_{ij} \sum_{k=1}^C (R_{jik} - G_{ijk})^2$$

Таким чином, генеративна модель буде намагатися видавати зображення, де згенеровані пікселі з більшими коефіцієнтами у вагових масках будуть більше схожі до реальних пікселів за своїми абсолютними значеннями кольорів, а пікселі з меншими коефіцієнтами відповідно матимуть менший пріоритет. Важливо зауважити, що це стосується лише генератора, оскільки функція втрати дискримінатора залишилася незмінною, через наявність лише компоненти BCE.

Усього навчено 6 генеративних моделей, кожна з яких використовувала відповідний набір вагових масок, та було згенеровано 6 штучних датасетів, які були використані під час навчання сегментаційних моделей.

4.6 Результати використання методу вагових масок з GAN аугментацією та обговорення

Для моделей, що вчилися на об'єднаному датасеті з реальних і згенерованих даних та використовували вагові маски, лише одна модель продемонструвала незначне покращення метрик у порівнянні з базовою, яка вчилася на об'єднаному датасеті, але не використовувала маски. Вагова маска цієї моделі мала більші значення коефіцієнтів для неоднозначних пікселів, ніж для надійних, а також до

масок була застосована операція звуження з ядром 3×3 та гаусове розмиття з ядром 5×5 . Було проведено додатковий експеримент, при якому вагові маски були використані під час навчання генеративної моделі, а під час навчання сегментаційної моделі жодні маски не використовувались. Метрики цих двох моделей у порівнянні з базовою представлені у таблиці 4.3. Позначка *NW* означає, що для даної моделі не використовувалися вагові маски на етапі навчання сегментаційної моделі. Для найкращої моделі загальні метрики збільшилися на 0.7% та 1.0% для OA та IoU відповідно, а точність розпізнавання малих культур незначно збільшилася.

Таблиця 4.3 — Результати порівняння метрик моделей, навчених на реальних та згенерованих даних з використанням вагових масок та метрик базової моделі (найвищі метрики виділено жирним шрифтом)

Назва с/г культури	Базова модель GAN			Модель Hardcase focus Erosion Gauss			Модель Hardcase focus Erosion Gauss NW		
	Метрики								
	R (%)	P (%)	IoU	R (%)	P (%)	IoU	R (%)	P (%)	IoU
Пшениця	85.8	73.2	65.3	82.9	76.6	65.1	84.6	73.8	64.3
Ріпак	37.1	52.0	27.6	34.4	53.6	25.6	35.0	49.0	28.1
Кукурудза	88.2	87.6	78.4	87.2	90.0	78.6	89.6	86.5	78.7
Цукровий буряк	15.4	65.8	14.3	51.2	59.3	34.5	45.9	58.0	32.6
Соняшник	92.9	82.8	77.9	90.9	86.0	78.2	92.0	83.9	77.6
Соя	68.2	73.1	54.5	67.5	78.0	53.8	64.9	75.9	55.4
Інші культури	9.7	33.2	8.1	5.7	38.1	7.9	9.3	34.7	4.2
ЯчмінЬ	30.0	48.2	22.7	32.5	47.7	21.2	28.6	44.9	16.4
ААС (%)	47.5	57.9	38.8	60.0	63.4	48.4	58.8	62.2	47.4
ОА (%)	78.6		66.2	79.3		67.2	78.7		66.4

Приклади сегментації на фрагменті реального зображення з тестового набору даних для вищенаведених моделей зображені на рисунку 4.6. Моделі які використовували маски мають менше візуальних графічних артефактів. Загалом же можна зробити висновок, що для тренування генеративних моделей найкращий результат дають вагові маски з більшими коефіцієнтами для неоднозначних пі-

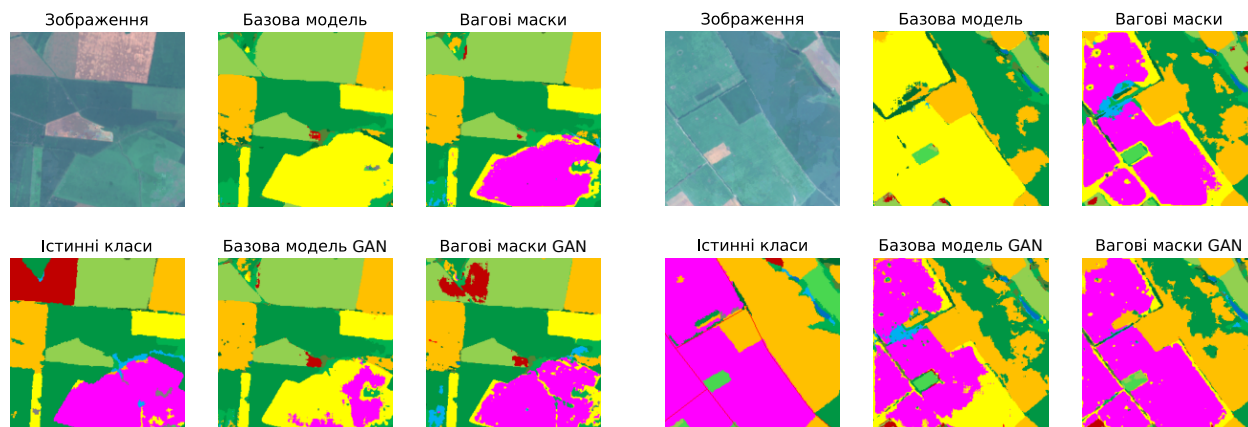


Рисунок 4.6 — Візуалізація супутникових зображень, істинних масок сегментації (істинні класи), результатів роботи базових моделей (базова модель, базова модель GAN), та результати роботи моделей з використанням вагових масок (вагові маски, вагові маски GAN)

кселів. Для тренування сегментаційних моделей ситуація протилежна – кращий результат дають маски, де нормальним пікселями присвоєно більші вагові коефіцієнти. У обох випадках важливо зменшити кількість неоднозначних пікселів, які межують з надійними, шляхом застосування морфологічних операцій.

Використання обох методів: GAN-аугментації та вагових масок, дозволяє збільшити загальну точність на тестовому наборі даних з 77.3% до 79.3%. загальний приріст становить 2.0%. Модель, навчена в попередньому розділі з використанням 12-канального композиту, дає приріст 4.2%, за рахунок використання додаткових даних, а саме за рахунок збільшення розмірності простору ознак втричі. У даному випадку, комбінація обох методів дозволяє отримати половину ефекту від використання нових даних та отримати кращі метрики для малих класів.

Висновки до розділу 4

У цьому розділі запропоновано новий підхід до аугментації даних у завданні класифікації сільськогосподарських культур на основі моделі GAN pix2pix для генерації реалістичних зображень, які відповідають багатоспектральним та текстурним характеристикам. Це дозволяє усунути проблему дисбалансу набору даних для глибокого навчання методами семантичної сегментації. Запропонований метод був порівняний із класичними підходами до генерації зображень на основі статистичних характеристик багатоспектральних ознак класів сільськогосподарських культур і був протестований на основних аугментаціях та функціях втрат, які застосовуються в випадку дисбалансу класів. В результаті, запропонований метод перевершив моделі, навчені на основі лише реальних даних та класичних підходів. Для більшості класів сільськогосподарських культур спостерігалось значне покращення точності для всіх мало представлених класів. Загальна точність моделі, навченої за запропонованим підходом, покращилася на 1.5% і 2.1% для OA та IoU, в той час як середні значення UA, PA та IoU для класів сільськогосподарських культур зросли на 11.2%, 6.4% та 10.2% порівняно з моделлю, навченою без синтетичних зображень.

Метод аугментації навчальних даних за допомогою генеративних змагальних мереж також вперше модифіковано до використання вагових масок, що використовувалися як під час навчання генеративної моделі, так і під час тренування сегментаційної модулі на згенерованих даних. Точність розпізнавання малих моделей незначно покращилась, при чому приріст точності був меншим ніж при використанні тільки реальних даних, а загальні метрики збільшилися на 0.7% та 1.0% для OA та IoU відповідно. Однак аугментація за допомогою генеративних мереж дає суттєвий приріст у точності для малих класів відносно базовою моделі, а використання вагових масок ще більше збільшує різницю. Це може вказувати

на досягнення межі найкращих можливих метрик на цьому наборі даних, оскільки неможливо постійно піднімати точність лише за рахунок покращення алгоритму навчання. Для задачі генерації маскету найкращі результати давали маски, де більш надійним пікселям присвоювалися менші значення вагових коефіцієнтів, що є протилежним результатом відносно задачі сегментації. В обох випадках, виділення крайових неоднозначних пікселів за допомогою морфологічних перетворень та присвоєння їм статусу надійних покращувало результат.

5 СЕМАНТИЧНА СЕГМЕНТАЦІЯ ГІСТОЛОГІЧНИХ ЗОБРАЖЕНЬ

Розроблений метод покращення результатів сегментації за допомогою вагових масок та GAN-агментації, досі був перевірений лише у області дистанційного зондування земної поверхні, а саме сегментації супутникових зображень.

Даний розділ присвячений дослідженню щодо можливості застосування розроблених методів у іншій прикладній області. Такою областю була вибрана сегментація гістологічних зображень. Вихідними даними є мікроскопічні знімки біологічних тканин, а задачею є коректно скласти сегментаційну карту, на якій відображено точну локалізацію різних типів тканин та окремих клітин [42].

Вихідні зображення в інших прикладних областях, наприклад такі як сегментація автомобілів чи пішоходів, значно відрізняються за своєю структурою та наповненням. Для будь-якого класу, відповідні йому пікселі можуть приймати будь-яке значення, і лише за одиничним пікселем неможливо вказали що саме це за клас. Приналежність до певного класу можливо визначити лише дивлячись на зображення в цілому та виділяючи складні високорівневі ознаки. На противагу, значення пікселю на супутниковому знімку вже містить в собі багато інформації. Для прикладу, висока яскравість у синьому спектральному каналі означає що це водні об'єкти, у зеленому – рослинність тощо. При цьому форма і розмір поля, засіяного певною культурою, може бути абсолютно довільним. Повертаючись до прикладу сегментації пішохода, ситуація протилежна – форма і розмір ніяк не можуть бути довільним, а натомість мають вирішальне значення при прийнятті рішення про наявність пішохода на зображенні.

Підсумовуючи вищенаведене: для супутникових знімків важливий колір та неважлива форма; для виділення пішоходів важлива форма і не так важливий його колір. Сегментація біологічних клітин на гістологічних у цьому розумінні є деяким перехідним етапом – різні типи клітин відрізняються між собою як формою та розміром, так і кольором.

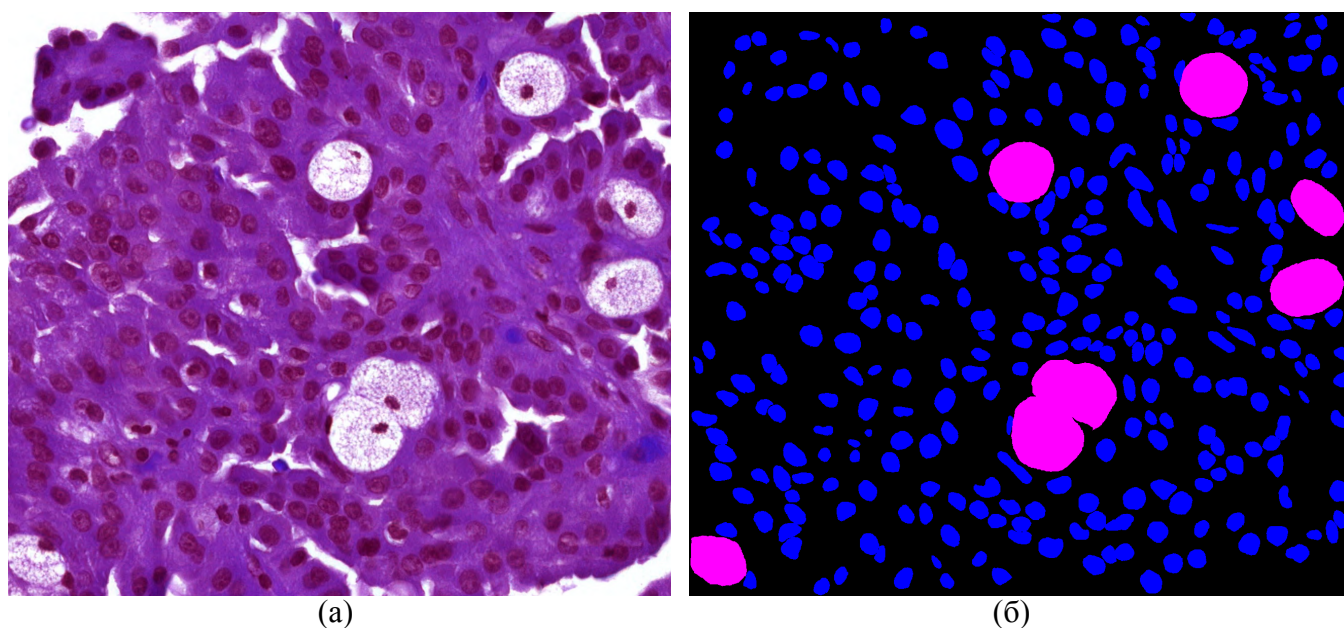


Рисунок 5.1 — Приклад зображення (а) та істинної маски сегментації (б) з датасету MoNuSAC2020

5.1 Використані дані та їхня попередня обробка

Для подальших експериментів у цьому розділі використано датасет MoNuSAC2020 [74], який був вперше представлений на змаганні з сегментації гістологічних знімків MoNuSAC (A Multi-organ Nuclei Segmentation and Classification Challenge) у 2020 році. Він представляє собою набір отриманих з мікроскопу зображень біологічних тканин, пофарбованих гематоксилином-еозином. У анотації датасету вказано клас та границі кожної клітини.

Загалом датасет містить 310 зображень різного розміру, з них 209 для навчання та 101 для тестування. На зображеннях присутні чотири типи клітин з чотирьох органів. Загальне число розмічених клітин перевищує 46 000, у зборі даних брали участь 37 різних лікувальних закладів та 71 пацієнт. Датасет має 5 класів: епітеліальні клітини, лімфоцити, нейтрофіли, макрофаги та один додатковий клас «Фон», що позначає відсутність перелічених клітин. Приклад розміченого зображення з датасету зображено на рисунку 5.1.

На рисунку 5.2 зображена гістограми розподілу значень пікселів в залежно-

сті від номеру каналу, а також індивідуальні гістограми для кожного класів. Значення є дискретними та лежать в діапазоні $[0, 255]$. Діаграми розподілу по класам показують, що представники різних класів дуже схожі між собою за спектральними характеристиками, зокрема такі пари як «Лімфоцит»-«Нейтрофіл» та «Фон»-«Макрофаг». У першому та третьому каналах середнє значення розподілу для деяких класів відрізняється, що дає можливість висувати гіпотези про їх часткову роздільність. Тим не менш, в усіх трьох каналах спостерігається сильне перекриття класів.

На рисунку 5.3 представлено візуалізації отриманого датасету. Через неможливість візуалізації тривимірних даних на площині, розмірність даних була понижена до двовимірної за допомогою алгоритмів PCA та tSNE. Наведені дані демонструють високий відсоток перекриття класів, та підтверджують висновки зі статистичного аналізу датасету (рисунок 5.2) про наявність пар класів, що дуже схожі між собою за спектральними характеристиками, наприклад «Лімфоцит»-«Нейтрофіл» та «Фон»-«Макрофаг».

Зображення з датасету поділено на фрагменти розміром по 256 кожне. Більшість оригінальних зображень мають значно більший розмір, аніж вказаний, тому у результаті загальне число фрагментів становило 1 131 у наборі для навчання, та 439 у наборі для тестування.

Для навчального набору даних обраховано відсоток неоднозначних точок по кожному класу та загалом, дані наведено у таблиці 5.1. Більша половина представників усі класи була визнана неоднозначними екземплярами. Класи «Нейтрофіл» та «Макрофаг» є найменш чисельними та одночасно найбільш неоднозначними класами. Незважаючи на це, явна тенденція що до залежності відсотку надійності від числа представників класу відсутня: наприклад, більш чисельний клас «Епітелій» має меншу частку надійних екземплярів ніж клас «Лімфоцит».

Візуалізацію надійних та неоднозначних точок представлено на рисунку 5.4. На перший погляд, закономірності на графіках відсутні, що підтверджує складну

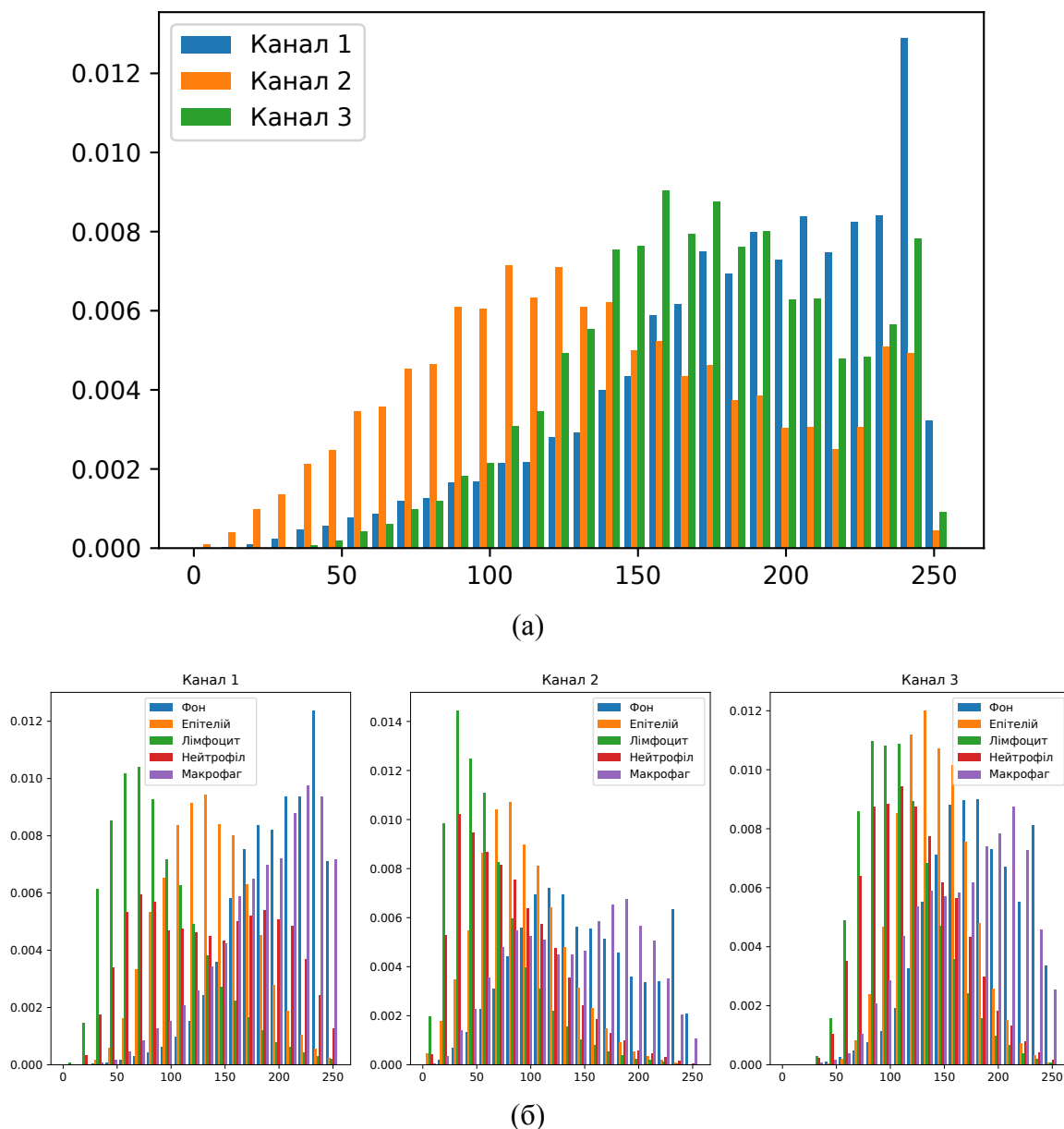


Рисунок 5.2 — Розподіл значень пікселів по каналам (а) та в залежності від істинного класу (б) для датасету MoNuSAC2020

структуру даних. Для успішного вирішення задачі сегментації, необхідна модель, яка враховуватиме не лише індивідуальні значення кожного пікселя, але й просторові властивості зображення, взаємне розташування різних пікселів та їхні значення. З огляду на це, модель, заснована на архітектурі U-Net, може бути не найкращим вибором, оскільки численні залишкові з'єднання фокусують увагу моделі саме на значеннях окремих пікселі [46].

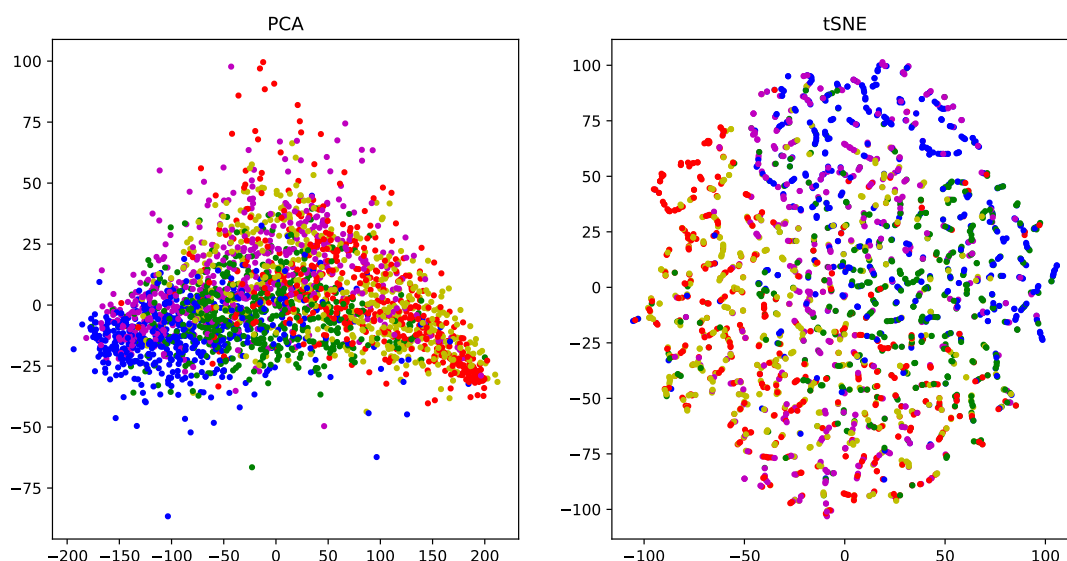


Рисунок 5.3 — Візуалізація датасету за допомогою PCA (а) та tSNE (б)

5.2 Покращення методу вагових масок

Перші експерименти на датасеті MoNuSAC2020 з використанням вагових масок показали необхідність вдосконалення даного методу, а саме в частині алгоритму виявлення неоднозначних екземплярів даних.

Оригінальна версія алгоритму розглядає кожен піксель як окремий елемент, жодним чином не пов'язаний з сусідніми пікселями, замінюючи задачу сегментації на задачу по-піксельної класифікації. Незважаючи на таке спрощення, це показувало хороший результат в області обробки супутникових знімків, де вирішальним фактором є колір, а не форма, розмір чи інші ознаки. У даній же області, колір хоч і є важливою ознакою, але й інші властивості, зокрема форма, мають більше значення ніж раніше. Тому постала необхідність врахувати хоча б частину просторових властивостей при визначенні неоднозначних пікселів.

При визначенні, які пікселі є сусідніми у просторі ознак, у першій версії алгоритму обчислювалася евклідова відстань між ними, а саме різниця значень кольорів у відповідних каналах:

Таблиця 5.1 — Загальне число пікселів та частка неоднозначних даних по кожному класу

Назва класу	Кількість пікселів, тис.	Відсоток надійних точок, %
Фон	60 719	44.68
Епітелій	8 990	34.66
Лімфоцит	3 138	42.28
Нейтрофіл	136	24.12
Макрофаг	1 136	24.06
Разом	74 121	43.01

$$\|\vec{X}_{i,j} - \vec{X}_{p,q}\|_{L2}$$

де $\hat{X} : \{\vec{X}_{i,j}\}$ – зображення \hat{X} , що складається з пікселів $\vec{X}_{i,j}$, а i,j та p,q – координати пікселів на зображенні. Довжина вектору $\vec{X}_{i,j}$ дорівнює числу каналів у зображенні, в даному випадку це 3. Значення каналу k в індивідуальному пікселі позначимо як $X_{i,j,k}$.

Основна ідея вдосконалення алгоритму полягає у тому, що при визначенні відстані у просторі ознак між двома пікселями, враховувати значення їхніх сусідніх (за координатами на зображенні) пікселів в межах деякого радіусу R . Це дозволяє хоча б частково врахувати форми та структури в межах цього радіусу. Очевидно, що такі сусідні пікселі мають впливати на кінцеву величину відстані менше, ніж цільові два пікселі. Тому доцільно враховувати відстані між сусідніми пікселями з певним, нижчими за одиницю коефіцієнтами. Нехай K – така матриця коефіцієнтів, розмірами $[2R + 1 \times 2R + 1]$. Центральний її елемент дорівнює одиниці, решта знаходяться в межах $[0,1)$:

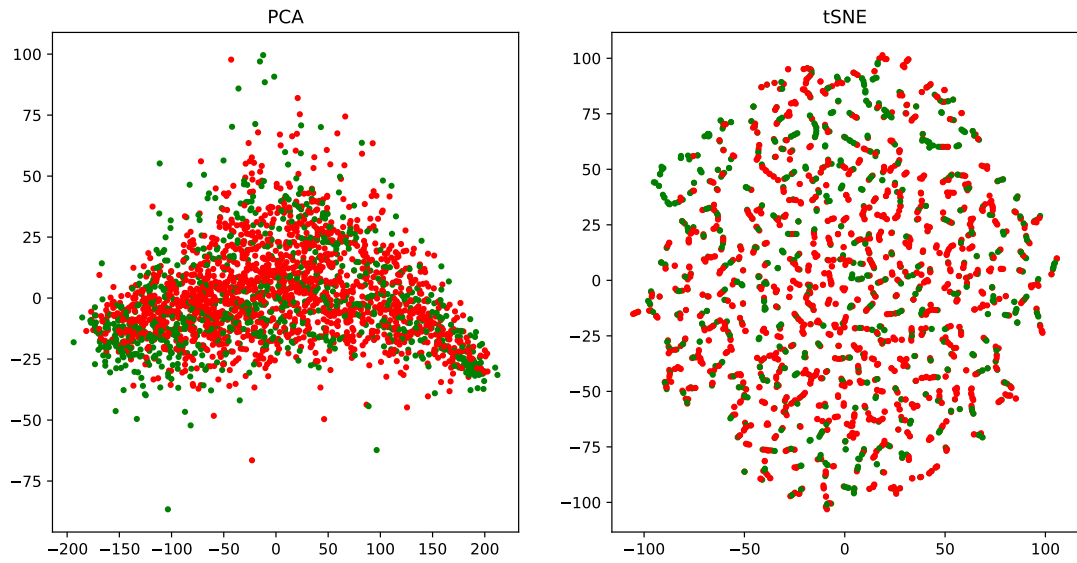


Рисунок 5.4 — Візуалізація визначених неоднозначних точок датасету за допомогою PCA (a) та tSNE (b). Зеленим кольором позначено надійні точки червоним – неоднозначні

$$K = \begin{bmatrix} K_{1,1} & \cdots & K_{1,R+1} & \cdots & K_{1,2R+1} \\ \vdots & \ddots & \vdots & & \vdots \\ K_{R+1,1} & \cdots & 1 & \cdots & K_{R+1,2R+1} \\ \vdots & & \vdots & \ddots & \vdots \\ K_{2R+1,1} & \cdots & K_{2R+1,R+1} & \cdots & K_{2R+1,2R+1} \end{bmatrix}$$

$$0 \leq K_{i,j} < 1, \begin{cases} i \neq R+1 \\ j \neq R+1 \end{cases}$$

Різниця між каналом k двох цільових пікселів $\vec{X}_{i,j}$ та $\vec{X}_{p,q}$, з урахуванням сусідніх пікселів обчислюється за наступною формулою:

$$\|X_{i,j,k} - X_{p,q,k}\|_{ch} = \sum_{f=-R}^{+R} \sum_{g=-R}^{+R} K_{f+R+1,g+R+1} (X_{i+f,j+g,k} - X_{p+f,q+g,k})$$

Це зважена середня різниця значень певного каналу пікселів. Застосування дужок замість модуля робить дану різницю інваріантною до поворотів фрагментів зображення. З урахуванням дистрибутивності множення, та комутативності додавання перепишемо ще в наступному вигляді:

$$\begin{aligned} \|X_{i,j,k} - X_{p,q,k}\|_{ch} &= \sum_{f=-R}^{+R} \sum_{g=-R}^{+R} (K_{f+R+1,g+R+1} X_{i+f,j+g,k} \\ &\quad - K_{f+R+1,g+R+1} X_{p+f,q+g,k}) = \\ &= \sum_{f=-R}^{+R} \sum_{g=-R}^{+R} K_{f+R+1,g+R+1} X_{i+f,j+g,k} - \sum_{f=-R}^{+R} \sum_{g=-R}^{+R} K_{f+R+1,g+R+1} X_{p+f,q+g,k} \end{aligned}$$

Отримані два доданки є нічим іншим як формулою згортки з ядром K над k каналом пікселів $\vec{X}_{i,j}$ та $\vec{X}_{p,q}$ відповідно. Для того, щоб уникнути обчислення згортки кожного разу при визначенні різниці між двома пікселями, є сенс обчислювати згортку заздалегідь для всього зображення. Нехай \hat{X}' – зображення, отримане шляхом по-канальної згортки ядра K з оригінальним зображенням \hat{X} :

$$X'_{i,j,k} = \sum_{f=-R}^{+R} \sum_{g=-R}^{+R} K_{f+R+1,g+R+1} X_{i+f,j+g,k}, \forall i,j,k$$

$$\hat{X}' = K \circledast \hat{X},$$

У результаті, відстань між двома пікселями у просторі ознак з урахуванням їх сусідів ($L2_{nn}$) обчислюється як:

$$\|\vec{X}_{i,j} - \vec{X}_{p,q}\|_{L2_{nn}} = \sqrt{\sum_k (X'_{i,j,k} - X'_{p,q,k})^2} = \|\vec{X}'_{i,j} - \vec{X}'_{p,q}\|_{L2}$$

Як висновок, визначення неоднозначних пікселів у наборі зображень з урахуванням просторовості, математично еквівалентне застосуванню згортки з ядром K до всього набору зображень, та запуску на ньому оригінального, по-піксельного алгоритму виявлення неоднозначних даних. Застосування згортки зводить дану задачу до попередньої і дозволяє хоча б частково врахувати просторові властивості в межах вибраного радіуса R .

5.3 Чисельні експерименти з сегментації гістологічних зображень

Нову версію алгоритму використано для визначення неоднозначних пікселів у датасеті MoNuSAC2020. У якості ядра згортки K , вибрано гауссівське ядро розмірами 5×5 . Застосування такого ядра до зображення є нічим іншим як гаус-

сівським розмиттям цього зображення. Далі застосовано звичайну, по-піксельну версію алгоритму на невеликій підмножині датасету. На її основі визначено приналежність до неоднозначних даних для повного датасету та згенеровано відповідні вагові маски. Статистика щодо частки неоднозначним даних по кожному класу та загалом представлено у таблиці 5.2.

Таблиця 5.2 — Загальне число пікселів та частка неоднозначних даних по кожному класу з урахуванням сусідніх пікселів

Назва класу	Кількість пікселів, тис.	Відсоток надійних точок, %
Фон	60 719	78.43
Епітелій	8 990	53.45
Лімфоцит	3 138	52.74
Нейтрофіл	136	1.23
Макрофаг	1 136	4.49
Разом	74 121	73.04

Для класів «Фон», «Епітелій» та «Лімфоцит» значно виросла частка надійних екземплярів. Це можна спробувати пояснити тим, що представники даних класів є подібними до інших, якщо розглядати задачу по-піксельної класифікації. Натомість, врахування просторових властивостей зображення, а саме значень у сусідніх пікселях, дає змогу краще розрізняти дані класи. Натомість, для класів «Нейтрофіл» та «Макрофаг» ситуація протилежна – збільшилася частка неоднозначних екземплярів, для цих класів врахування просторовості зменшило їх відмінності від інших класів. Це може бути пов'язаним з наявністю викидів та екстремальних значень серед представників цих класу, вплив яких зменшився через врахування просторовості. Іншим поясненням може бути малі розміри цих клітин: розглядаючи окремі екземпляри класів, можна зауважити що значна частка їхніх сусідів будуть вже представниками інших класів.

5.3.1 Набори вагових масок

Згідно з даними таблиці 5.2, майже всі пікселі у найбільшому класі «Фон» є надійними, а більшість пікселів з інших класів є неоднозначними. У попередніх експериментах з супутниковими знімками перевірялося, даним якого саме типу слід надати перевагу через вагові маски – чи надійним, чи неоднозначними. У даному випадку, з огляду на статистику датасету, очевидним є надати перевагу неоднозначним даним. Надати перевагу нормальним пікселям – означає зробити фокус моделі на найбільш представленому класі, який відповідає фону, та значно зменшити акцент моделі на найменш представлені класи, які відповідають різним типам клітин.

Згенеровано 2 набори вагових масок, де неоднозначні пікселі мали більші значення відповідних коефіцієнтів у вагових масках. Для повноти експерименту, також було згенеровано ще один набір вагових масок, де фокус надавався нормальним даним. Морфологічні перетворення було вирішено не застосовувати через неоднорідну структуру вагових масок. Повна інформація про набори представлена у таблиці 5.3.

Приклади вагових масок з різних наборів зображені на рисунку 5.5. З цих зображень можна зробити висновок, що деякі клітини повністю віднесені до неоднозначних даних, а інші – лише частково. У таких клітинах до неоднозначних даних віднесені лише їх периферичні області, що вказує на особливу складність розпізнавання межі клітини та фону. Більшість пікселів, що відповідають фону, віднесені до надійних даних, окрім невеликої частки структур, що схожі за своїми спектральними властивостями на клітини.

Таблиця 5.3 — Список перетворень, застосованих до різних наборів вагових масок

№	Назва набору	Фокус вагових коефіцієнтів	Гауссове розмиття
1	Normal focus	нормальні пікселі (4:1)	-
2	Hardcase focus	неоднозначні пікселі (4:1)	-
3	Hardcase focus Gauss	неоднозначні пікселі (4:1)	5×5

5.3.2 Архітектура моделі та схема експерименту

У зв'язку з набагато складнішою структурою даних на зображенні та великою варіабельністю текстур, прийнято рішення використати іншу архітектуру моделі, а саме SegNet [75], який має вдвічі більше параметрів за попередню модель, засновану на U-Net. На рисунку 5.6 представлена архітектура моделі. Вона приймає на вхід дані розмірністю $[N \times 256 \times 256 \times 3]$, та видає результат розмірністю $[N \times 256 \times 256 \times 5]$, де N – кількість зображень у навчальному пакеті.

Оптимізація параметрів моделі виконувалася з використанням оптимізатора Adam за параметрами $\beta_1 = 0.9$ та $\beta_2 = 0.999$. Швидкість навчання змінювалася динамічно в межах $[1 \cdot 10^{-1}, 1 \cdot 10^{-5}]$. Для навчання базової моделі у якості функції втрат використовувалося стандартна крос-ентропія. Під час навчання моделі з використанням методу вагових масок, використовувалася модифікована крос-ентропія:

$$BCE(\hat{Y}, \hat{Y}', \hat{M}) = -\frac{1}{n} \sum_{i,j} M_{ij} \sum_{k=1}^C Y_{ijk} \log Y'_{ijk}$$

де \hat{Y} , \hat{Y}' – сегментаційні маски, істинна і передбачена відповідно, \hat{M} – вагова маска, n – кількість пікселів в одному фрагменті, C – кількість класів.

Всього було натреновано 4 моделі – одна базова і 3 – з використанням різних наборів вагових масок.

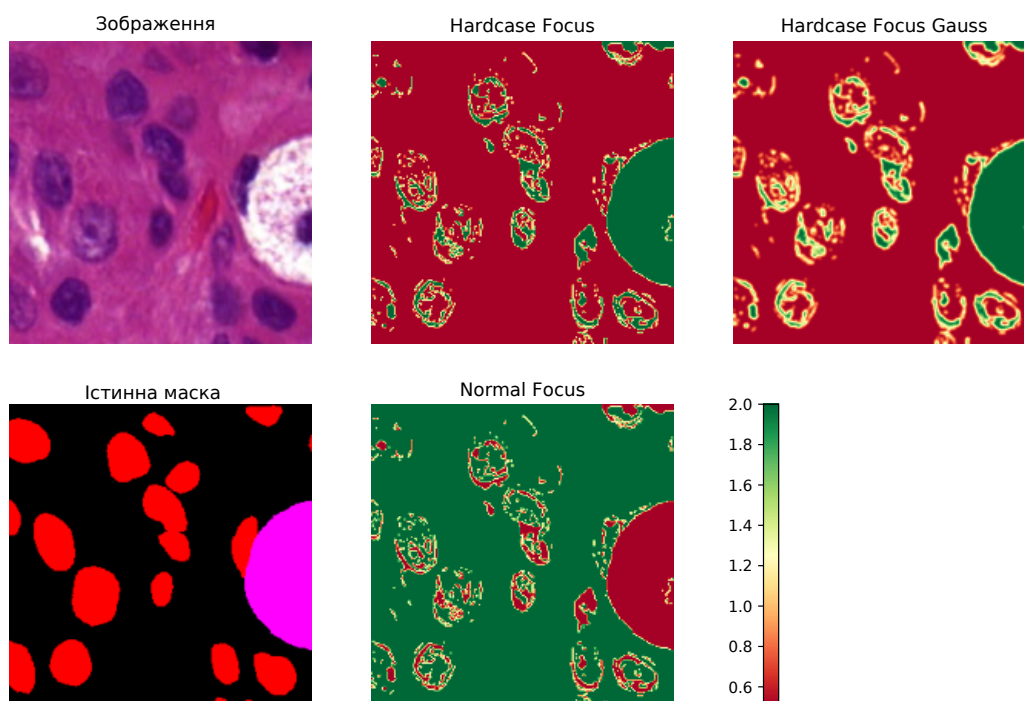


Рисунок 5.5 — Візуалізація гістологічного зображення та відповідної йому маски з набору

5.4 Результати та обговорення

Навчені моделі оцінено на незалежному наборі даних, який не використовувався при навчанні. Обраховано метрики, такі як точність, влучність (precision), повнота (recall) та IoU. Детальні результати наведені у таблиці 5.4. Як і очікувалось, модель, що тренувалася з акцентом на нормальні дані, не продемонструвала покращення метрик малих класів у порівнянні з базовою моделлю. Інші дві моделі розрізняють малі класи суттєво краще за базову модель. Вагові маски цих моделей мали більше значення коефіцієнтів для неоднозначних пікселів, ніж для нормальних, а до набору масок останньої моделі також було застосовано гаусове розмиття з ядром 5×5 . Модель, що вчилася з використанням набору масок Hardcase focus, перевищує базову модель майже в усіх метриках, які стосуються

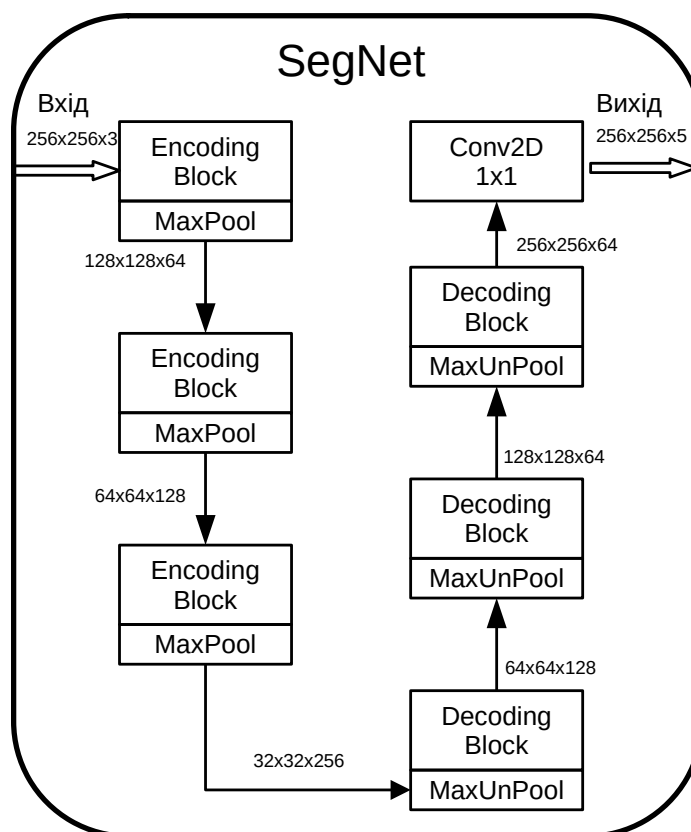


Рисунок 5.6 — Архітектура сегментаційної моделі

малих класів, а загальні метрики змінилися незначно. Особливо слід відзначити те, що базова модель нездатна розпізнати клас «Нейтрофіл», а використання вагових масок дало можливість правильно розпізнавати цей клас майже в половині випадків.

У попередніх дослідях, для задачі сегментації найкраще працювали вагові маски, де більші значення коефіцієнтів відповідали неоднозначним пікселям. У даному експерименті ситуація протилежна. Це може бути пов'язано з наявністю домінантного класу «Фон», число екземплярів якого набагато більше за усі інші класи разом узяті. Алгоритм визначення неоднозначним даних відніс більшість екземплярів цього класу до однієї категорії. Як наслідок, модель мала шанси вивчити малі класи лише якщо цій категорії був присвоєний низький пріоритет.

Представники домінантного класу знаходяться у просторі ознак у великих

Таблиця 5.4 — Порівняння метрик моделей сегментації гістологічних зображень (метрики, вищі за базову модель, виділено жирним шрифтом)

Назва класу	Базова модель			Модель Normal focus			Модель Hardcase focus			Модель Hardcase focus Gauss		
	Метрики											
	R,%	P,%	IoU	R,%	P,%	IoU	R,%	P,%	IoU	R,%	P,%	IoU
Фон	93.0	96.1	89.6	93.6	96.0	90.1	93.0	95.7	89.3	92.9	94.5	88.1
Епітелій	81.9	71.9	62.1	76.9	74.3	60.8	82.7	73.5	63.7	83.7	69.7	61.4
Лімфоцит	81.0	71.3	61.1	77.5	72.5	59.9	84.8	68.2	60.7	79.0	74.4	62.1
Нейтрофіл	0.0	0.0	0.0	0.0	0.0	0.0	48.3	42.9	29.4	23.0	10.1	7.6
Макрофаг	4.9	3.0	1.9	24.8	2.3	2.1	11.4	11.9	6.2	5.4	9.6	3.6
AAC (%)	52.1	48.5	42.9	54.6	49.0	42.6	64.0	58.4	49.9	56.8	51.7	44.6
OA (%)	90.5		83.4	90.7		83.6	90.3		83.4	89.1		82.6

кількостях, і у разі наявності проблеми перекриття класів, з більшою імовірністю будуть вибрані як сусіди під час роботи алгоритму виявлення неоднозначних даних. Це узгоджується з реальними результатами, коли модель має тенденцію вивчати найбільш представлений клас, ігноруючи малі класи. У даному датасеті, приблизно половина пікселів класів «Епітелій» та «Лімфоцит» знаходилися у просторі ознак достатньо далеко, щоб бути чітко розрізненими від інших класів та не отримати такого суттєвого фокусу під час навчання, як класи «Нейтрофіл» чи «Макрофаг». Тим не менш, метрики цих класів також збільшилися у порівнянні з базовою моделлю.

5.5 Післяопераційна обробка сегментаційних масок

Використання вагових масок значно покращило метрики розпізнавання малих класів. Тим не менш, вивчення візуальних представлень передбачених сегментаційних масок, наприклад зображених на рисунку 5.7, звертає увагу на їх неоднорідність. У значній частині випадків, правильно класифікована лише ча-

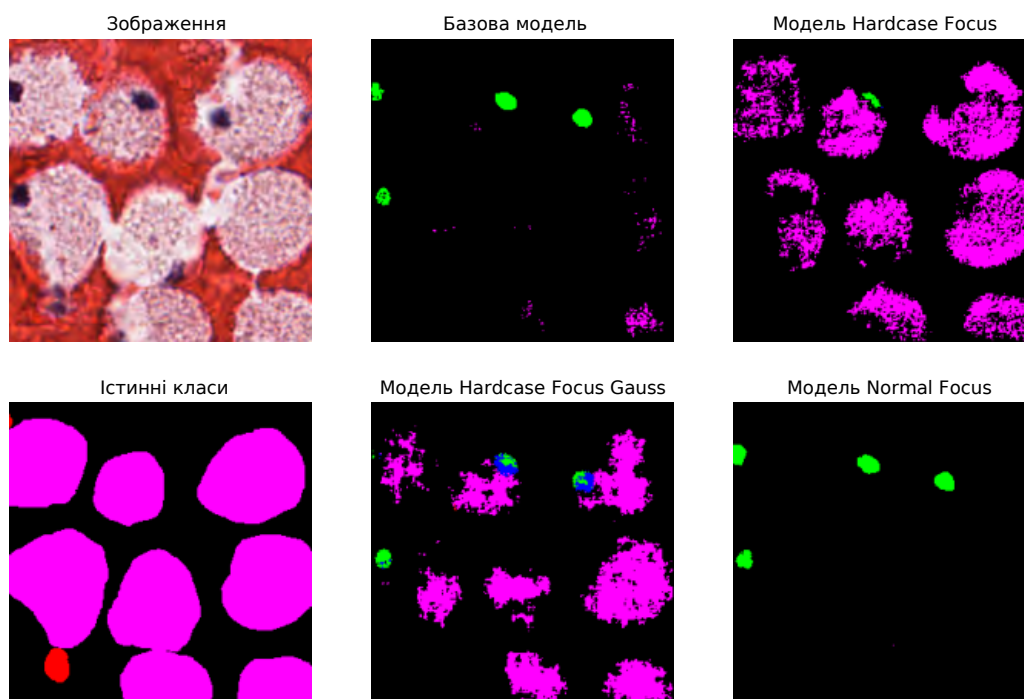


Рисунок 5.7 — Візуалізація гістологічних зображень, істинних масок сегментації (істинні класи), результатів роботи базової моделі та результати роботи моделей, що навчені з використанням вагових масок

стина пікселів, що відповідають певній клітині, іншим же пікселям присвоєно клас «Фон». Ці дві групи пікселів хаотично розподілені по всій площі розглянутих клітини.

Також слід взяти до уваги, що для класів, які відповідають клітинам, метрика влучності практично завжди менша за повноту, для класу «Фон» ситуація протилежна. Це дає підстави стверджувати, що ті пікселі, які віднесені до класу клітин, з високою ймовірністю і справді до них належать, тобто частка хибних спрацювань невелика. У той же час, для пікселів фону, особливо тих що знаходяться поруч з пікселями інших класів, частка хибних спрацювань може бути великою.

Для подолання цієї проблеми запропоновано використання морфологічних перетворень, а саме операції морфологічного закриття (morphological closing) з

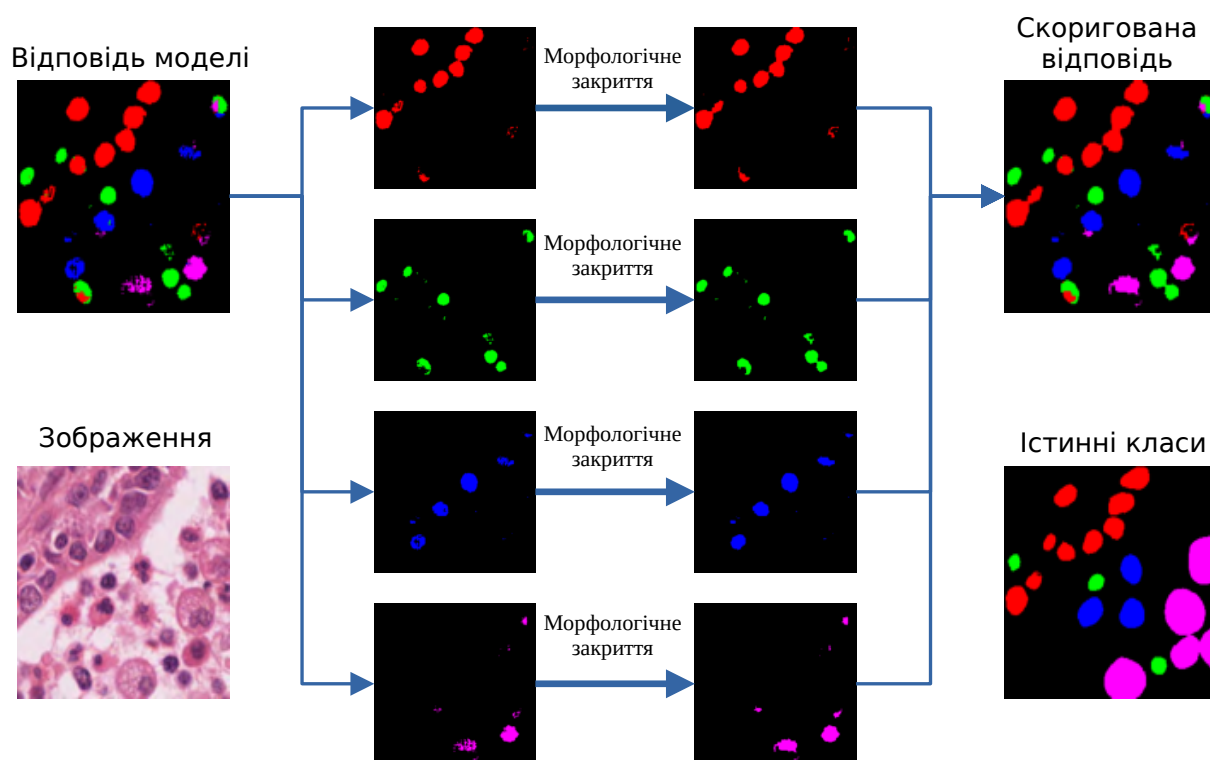


Рисунок 5.8 — Схема алгоритму післяопераційної обробки сегментаційної маски та приклад результату

ядром 5×5 відносно класів, які представляють клітини. Схема даного перетворення зображена на рисунку 5.8. Вихідна сегментаційна маска розбивається на 4 окремих зображень, що відповідає числу класів клітин. Отримані зображення по своїй суті є маскою, де пікселям цільового класу відповідають значення 1, а решті пікселів – 0. До кожної такої маски застосовується операція морфологічного закриття, яка має на меті заповнити невеликі проміжки між пікселями цільового класу. З ядром 5×5 , максимальний теоретично можливий розмір таких проміжків складає 10 пікселів. Далі, пусте зображення по чергово заповнюється класами клітин відповідно до обробленої маски, решті пікселів, які залишилися незаповненими, присвоюється клас «Фон». Таким чином формується скоригована сегментаційна маска.

Описаний вище алгоритм застосований для післяопераційної обробки результатів роботи натренованих моделей. Приклади сегментаційних масок до, та після обробки зображено на рисунку 5.9. З нього видно, що для базової моделі внесок даного алгоритму у загальні метрики невеликий. Натомість для моделі, що навчалася з використанням вагових масок, даний алгоритм вніс суттєвий вклад у розпізнавання загальної форми клітин. Однак, наявні наступні неточності: залишаються окремі анклавні пікселі фону всередині клітин, а також остаточна форма клітин не є опуклою. Як висновок, представлений алгоритм суттєво покращує результати роботи моделі, однак має сенс розробка більш досконалих методів післяопераційної обробки, які будуть враховувати типові геометричні форми цільових клітин.

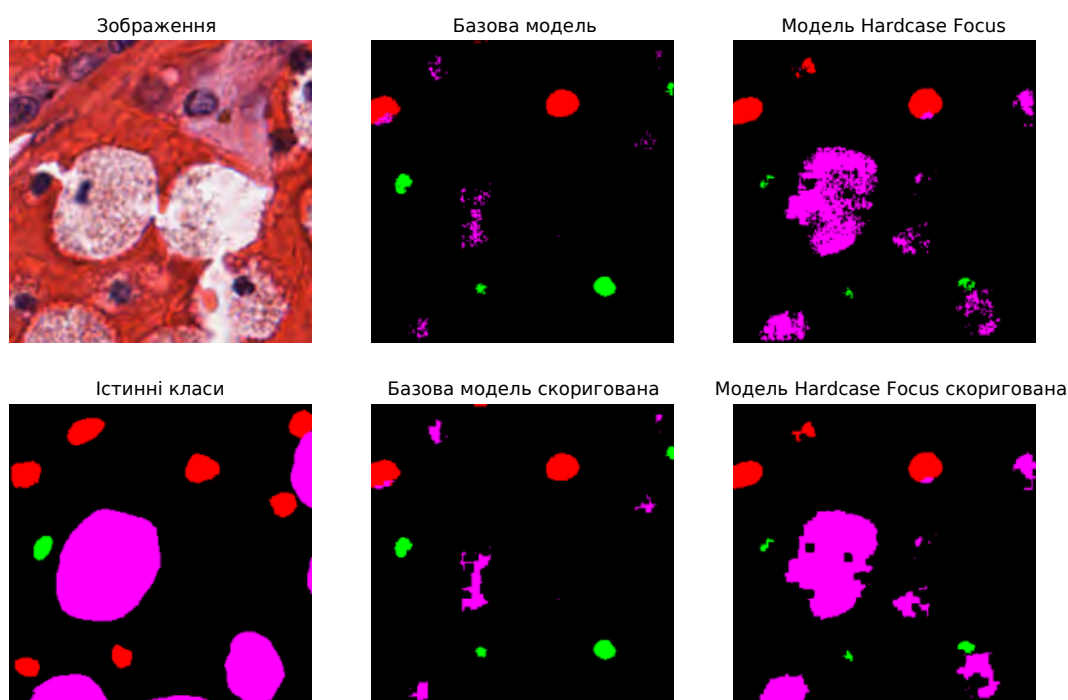


Рисунок 5.9 — Візуалізація гістологічних зображень, істинних масок сегментації, результатів роботи базової та найкращої моделей; а також їх скориговані версії

За допомогою вищенаведеного алгоритму скориговано передбачення усіх моделей для всього тестового набору та перераховано метрики. Результати наведені у таблиці 5.5. Загальні метрики залишилися майже без змін, а серед метрик, обрахованих по окремим класам, спостерігається як зростання, так і падіння. В середньому, збільшення метрик спостерігалось частіше, ніж зниження. Окрім того, метрики, які відповідають за повноту, у основному збільшилися. Натомість метрики, які відповідають за точність, або залишилися такими ж самими, або незначно зменшилися. Можна зробити висновок, що відбулася певна зміна балансу між метриками зі сторони влучності в сторону повноти, що виправляє початкову зміщеність рівноваги в сторону влучності (таблиця 5.4). Загалом, суттєвого збільшення метрик не відбулося, однак покращилася візуальне представлення результатів та їхня придатність для подальшої обробки.

Таблиця 5.5 — Порівняння метрик моделей сегментації гістологічних зображень зі скоригованою сегментаційною маскою (метрики, вищі за базову модель, виділено жирним шрифтом)

Назва класу	Базова модель			Модель Normal focus			Модель Hardcase focus			Модель Hardcase focus Gauss		
	Метрики											
	UA	PA	IoU	UA	PA	IoU	UA	PA	IoU	UA	PA	IoU
Фон	93.3	95.6	89.4	94.0	95.5	90.0	93.4	94.8	88.9	93.2	94.0	88.0
Епітелій	81.0	73.6	62.8	75.8	76.3	61.4	81.6	75.3	64.4	82.9	71.4	62.2
Лімфоцит	79.5	73.2	61.5	75.7	73.4	59.4	83.4	69.9	61.4	77.1	76.4	62.3
Нейтрофіл	0.0	0.0	0.0	0.0	0.0	0.0	47.7	44.7	30.0	27.9	9.6	7.7
Макрофаг	5.2	4.0	2.3	15.6	2.4	2.1	11.5	16.1	7.2	6.0	11.4	4.1
AAC (%)	51.8	49.3	43.2	52.2	49.5	42.6	63.5	60.2	50.4	57.4	52.6	44.9
OA (%)	90.3		83.4	90.6		83.6	90.0		83.4	89.0		82.6

Висновки до розділу 5

У даному розділі розглянуто застосування розробленого нового методу вагових масок в іншій предметній області з відмінними властивостями вихідних даних. Дані з обраного датасету гістологічних знімків, мали сильний дисбаланс у розподілів класів, а подальший аналіз датасету вказав на наявність проблеми перекриття класів. Запропоновано вдосконалену версію алгоритму виявлення неоднозначних екземплярів даних, яка враховує просторові властивості зображення, а саме значення сусідніх пікселів при обрахунку відстані у просторі ознак. Доведено, що дану версію алгоритму можна звести до попередньої шляхом застосування до зображення операції згортки з ядром коефіцієнтів.

Проведено порівняння розробленого методу зі стандартною процедурою навчання моделі на тих самих даних і тією самою архітектурою моделі. За результатами порівняння, використання розробленого методу збільшує точність розпізнавання усіх малих класів, які відповідають різним типам клітин, однак значення загальних метрик залишається практично незмінними. Зокрема для класу «Нейтрофіл», який не могла розпізнати базова модель, використання вагових масок збільшило влучність до 48.3%, а повноту до 42.9%. Дещо гірші результати для класів «Епітелій» та «Макрофаг», де влучність зросла на 1.8% та 6.5% відповідно.

Також запропоновано алгоритм післяопераційної обробки результатів роботи моделі, який змішує балансує метрики влучності та повноти, покращує візуальне представлення сегментаційних масок та робить їх більш придатними для подальшого використання. Для даної задачі найкращі результати давали маски, де неоднозначним пікселям присвоювалися більші значення вагових коефіцієнтів. Це пояснюється наявністю домінантного класу, більшість представників якого мали статус надійних.

ВИСНОВКИ

Дисертацію присвячено вирішенню проблеми незбалансованих наборів даних та перекриття класів у задачах сегментації складних зображень, яка є актуальною при навчанні моделей машинного навчання. Основною науковою новизною дослідження є розробка нових методів для покращення продуктивності моделей сегментації при наявності незбалансованих наборів даних та проблеми перекриття класів, зокрема, шляхом використання генеративних змагальних мереж (GAN) для аугментації даних і застосування нового методу вагових масок для покращення навчання моделей.

- 1. Вперше розроблено метод виявлення неоднозначних екземплярів даних** на основі методу К найближчих сусідів (KNN), який дозволяє ефективно визначати дані, що викликають проблеми при навчанні моделей машинного навчання через подібність між класами. Цей метод підвищив влучність класифікації малих класів у межах від 4.4% до 52% та повноту у межах від 5.8% до 54% для задач сегментації з високим перекриттям класів. Він ефективно зменшив кількість помилкових класифікацій через подібність між класами, особливо в задачах класифікації сільськогосподарських культур.
- 2. Вперше запропоновано метод аугментації датасетів багатоспектральних зображень з використанням генеративних змагальних мереж**, що дозволяє ефективніше вирішувати проблему незбалансованості класів. Зокрема, в задачах класифікації сільськогосподарських культур за супутниковими даними загальна точність зростає на 1.5%, а показник загального IoU на 2.1%. Цей підхід також покращив баланс недостатньо представлених класів у задачах класифікації без ризику перенавчання, що значно покращило метрики малих класів.
- 3. Удосконалено метрики для оцінки якості датасетів**, які включають врахування незбалансованості даних та перекриття класів. Ці метрики дозво-

ляють покращити процес формування навчальних наборів даних, зокрема для задач сегментації багатоспектральних зображень, та роблять можливим використання методу вагових масок.

4. **Модифіковано функцію втрат в моделях глибинного навчання через введення вагових масок для задач сегментації та навчання GAN.** Застосування вагових масок під час навчання сегментаційних та генеративних моделей, покращили виявлення менш представлених класів, а також підвищили показник загальної точності (OA) на 2.0% та збільшення IoU на 3.1% у порівнянні зі стандартними моделями сегментації.
5. **Отримав подальший розвиток метод вагових масок для компенсації незбалансованості та перекриття класів,** що дозволяє покращити точність класифікації малопоширених класів без необхідності розширення датасету. Покращена версія алгоритму враховує частину просторових ознак на зображенні, що дозволяє уникнути втрати інформації та покращити результати моделей сегментації гістологічних зображень. Метод покращив точність класифікації малих класів на величину від 3.8% до 48%.
6. Результати дисертаційного дослідження використано в проекті програми Horizon Europe «Satellites for Wilderness Inspection and Forest Threat Tracking» (SWIFTT) (грантова угода 101082732, 2022-2025); у проекті НФД України «Методи і моделі глибинного навчання для прикладних задач супутникового моніторингу» (2020.02/0292, 2020-2023); та у НТР «Інформаційні технології геопросторового аналізу розвитку сільських територій і громад» (Договір РН/27-2023 від 25 травня 2023 р.) за рахунок зовнішнього інструменту допомоги Європейського Союзу для виконання зобов'язань України у Рамковій програмі Європейського Союзу з наукових досліджень і інновацій «Горизонт 2020».

Таким чином, запропоновані нові методи та підходи дозволяють суттєво підвищити точність моделей сегментації в умовах незбалансованих наборів да-

них та проблеми перекриття класів, що робить дослідження вагомим внеском у розвиток методів прикладної математики та обробки зображень.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. L. Shumilo, A. Okhrimenko, N. Kussul, S. Drozd, and O. Shkalikov, “Generative adversarial network augmentation for solving the training data imbalance problem in crop classification,” *Remote Sensing Letters*, vol. 14, pp. 1131–1140, 2023.
2. A. Okhrimenko and N. Kussul, “Using weight reliability masks on imbalanced datasets for satellite image segmentation,” *Cybernetics and Systems Analysis*, vol. 60, pp. 700–710, 9 2024.
3. A. Okhrimenko and N. Kussul, “Data mining of machine learning datasets for hard case identification,” *International Scientific Technical Journal «Problems of Control and Informatics»*, vol. 68, pp. 84–95, 5 2023.
4. А. О. Охріменко, “Методи сегментації зображень різної природи для випадку незбалансованого набору даних,” in *Методи комп’ютерного зору і глибоких нейронних мереж для еколого-економічного аналізу*, pp. 143–171, Наукова думка, 2024.
5. A. Okhrimenko and N. Kussul, “Knn-based algorithm of hard case detection in datasets for classification,” in *Proceedings of the 11th International Conference on Applied Innovations in IT*, vol. 11, Anhalt University of Applied Sciences, 3 2023.
6. A. Okhrimenko and N. Kussul, “Unbalanced datasets management for the problem of segmentation of satellite images,” in *2024 IEEE Fourth International Conference on System Analysis Intelligent Computing (SAIC)*, 2024.
7. A. Mikołajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pp. 117–122, 2018.

8. C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, 2019.
9. W. Almutairi and R. Janicki, “On relationships between imbalance and overlapping of datasets,” in *Proceedings of 35th International Conference on Computers and Their Applications* (G. Lee and Y. Jin, eds.), vol. 69 of *EPiC Series in Computing*, pp. 141–150, EasyChair, 2020.
10. O. Kramer, “Dimensionality reduction with unsupervised nearest neighbors,” *Intelligent Systems Reference Library*, vol. 51, 2013.
11. V. García, R. A. Mollineda, and J. S. Sánchez, “On the k-nn performance in a challenging scenario of imbalance and overlapping,” *Pattern Analysis and Applications*, vol. 11, 2008.
12. M. M. Nwe and K. T. Lynn, *KNN-Based Overlapping Samples Filter Approach for Classification of Imbalanced Data*, pp. 55–73. Cham: Springer International Publishing, 2020.
13. L. Chen, B. Fang, Z. Shang, and Y. Tang, “Tackling class overlap and imbalance problems in software defect prediction,” *Software Quality Journal*, vol. 26, 2018.
14. Y. Tang and J. Gao, “Improved classification for problem involving overlapping patterns,” *IEICE Transactions on Information and Systems*, vol. E90-D, 2007.
15. N. Lässig, S. Oppold, and M. Herschel, “Metrics and algorithms for locally fair and accurate classifications using ensembles,” *Datenbank-Spektrum*, vol. 22, 2022.
16. L. Wang, M. Han, X. Li, N. Zhang, and H. Cheng, “Review of classification methods on unbalanced data sets,” *IEEE Access*, vol. 9, 2021.
17. A. Kumar, S. Goel, N. Sinha, and A. Bhardwaj, “A review on unbalanced data classification,” in *Proceedings of International Joint Conference on Advances in*

- Computational Intelligence* (M. S. Uddin, P. K. Jamwal, and J. C. Bansal, eds.), (Singapore), pp. 197–208, Springer Nature Singapore, 2022.
18. X. Li and L. Zhang, “Unbalanced data processing using deep sparse learning technique,” *Future Generation Computer Systems*, vol. 125, 2021.
 19. A. Vilorio, O. B. P. Lezama, and N. Mercado-Caruzo, “Unbalanced data processing using oversampling: Machine learning,” *Procedia Computer Science*, vol. 175, pp. 108–113, 2020.
 20. S. Hido, H. Kashima, and Y. Takahashi, “Roughly balanced bagging for imbalanced data,” *Statistical Analysis and Data Mining*, vol. 2, 2009.
 21. M. Lango and J. Stefanowski, “Multi-class and feature selection extensions of roughly balanced bagging for imbalanced data,” *Journal of Intelligent Information Systems*, vol. 50, 2018.
 22. Y. Cui, M. Jia, T. Y. Lin, Y. Song, and S. Belongie, “Class-balanced loss based on effective number of samples,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2019-June, 2019.
 23. T. H. Phan and K. Yamamoto, “Resolving class imbalance in object detection with weighted cross entropy losses,” *CoRR*, vol. abs/2006.01413, 6 2020.
 24. X. Qiao and Y. Liu, “Adaptive weighted learning for unbalanced multiclass classification,” *Biometrics*, vol. 65, 2009.
 25. Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 904, 1995.

26. W. Hu, Y. Yin, Y. K. Tan, A. Tran, H. Kruppa, and R. Zimmermann, "Gan-assisted road segmentation from satellite imagery," *ACM Trans. Multimedia Comput. Commun. Appl.*, 11 2023. Just Accepted.
27. M. Shah, M. Gupta, and P. Thakkar, "Satgan: Satellite image generation using conditional adversarial networks," in *Proceedings - International Conference on Communication, Information and Computing Technology, ICCICT 2021*, pp. 1–6, 2021.
28. W. Li, J. Li, J. Polson, Z. Wang, W. Speier, and C. Arnold, "High resolution histopathology image generation and segmentation through adversarial training," *Medical Image Analysis*, vol. 75, 2022.
29. B. Franch, E. Vermote, S. Skakun, A. Santamaria-Artigas, N. Kalecinski, J. C. Roger, I. Becker-Reshef, B. Barker, C. Justice, and J. A. Sobrino, "The arya crop yield forecasting algorithm: Application to the main wheat exporting countries," *International Journal of Applied Earth Observation and Geoinformation*, vol. 104, 2021.
30. N. Kussul, M. Lavreniuk, A. Kolotii, S. Skakun, O. Rakoid, and L. Shumilo, "A workflow for sustainable development goals indicators assessment based on high-resolution satellite data," *International Journal of Digital Earth*, vol. 13, 2020.
31. C. Boryan, Z. Yang, R. Mueller, and M. Craig, "Monitoring us agriculture: The us department of agriculture, national agricultural statistics service, cropland data layer program," *Geocarto International*, vol. 26, 2011.
32. P. Defourny, S. Bontemps, N. Bellemans, C. Cara, G. Dedieu, E. Guzzonato, O. Hagolle, J. Inglada, L. Nicola, T. Rabaute, M. Savinaud, C. Udriou, S. Valero, A. Bégué, J. F. Dejoux, A. E. Harti, J. Ezzahar, N. Kussul, K. Labbassi, V. Lebourgeois, Z. Miao, T. Newby, A. Nyamugama, N. Salh, A. Shelestov,

- V. Simonneaux, P. S. Traore, S. S. Traore, and B. Koetz, “Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the sen2-agri automated system in various cropping systems around the world,” *Remote Sensing of Environment*, vol. 221, 2019.
33. S. I. Moazzam, U. S. Khan, M. I. Tiwana, J. Iqbal, W. S. Qureshi, and S. I. Shah, “A review of application of deep learning for weeds and crops classification in agriculture,” in *2019 International Conference on Robotics and Automation in Industry, ICRAI 2019*, 2019.
 34. W. Xu, X. Deng, S. Guo, J. Chen, L. Sun, X. Zheng, Y. Xiong, Y. Shen, and X. Wang, “High-resolution u-net: Preserving image details for cultivated land extraction,” *Sensors (Switzerland)*, vol. 20, 2020.
 35. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, pp. 139–144, 10 2020.
 36. V. D. Pham and Q. T. Bui, “Spatial resolution enhancement method for landsat imagery using a generative adversarial network,” *Remote Sensing Letters*, vol. 12, 2021.
 37. L. Abady, M. Barni, A. Garzelli, and B. Tondi, “Gan generation of synthetic multispectral satellite images,” in *Image and Signal Processing for Remote Sensing XXVI* (L. Bruzzone, F. Bovolo, and E. Santi, eds.), vol. 11533, p. 115330L, International Society for Optics and Photonics, SPIE, 2020.
 38. M. Alibani, N. Acito, and G. Corsini, “Multispectral satellite image generation using stylegan3,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, 2024.

39. Z. Xu, J. Du, J. Wang, C. Jiang, and Y. Ren, "Satellite image prediction relying on gan and lstm neural networks," in *IEEE International Conference on Communications*, vol. 2019-May, 2019.
40. L. E. Christovam, M. H. Shimabukuro, M. de Lourdes B.T. Galo, and E. Honkavaara, "Pix2pix conditional generative adversarial network with mlp loss function for cloud removal in a cropland time series," *Remote Sensing*, vol. 14, 2022.
41. H. J. Andrade and B. J. Fernandes, "Synthesis of satellite-like urban images from historical maps using conditional gan," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, 2022.
42. Y. Wu, M. Cheng, S. Huang, Z. Pei, Y. Zuo, J. Liu, K. Yang, Q. Zhu, J. Zhang, H. Hong, D. Zhang, K. Huang, L. Cheng, and W. Shao, "Recent advances of deep learning for computational histopathology: Principles and applications," *Cancers*, vol. 14, p. 1199, 2 2022.
43. M. Kadaskar and N. Patil, "Image analysis of nuclei histopathology using deep learning: A review of segmentation, detection, and classification," *SN Computer Science*, vol. 4, 2023.
44. N. Y. Ahmed, "Real-time and accurate deep learning-based multi-organ nucleus segmentation in histology images," *Journal of Real-Time Image Processing*, vol. 21, 2024.
45. E. S. Nasir and M. M. Fraz, "Nurisc: Nuclei radial instance segmentation and classification," in *Lecture Notes in Electrical Engineering*, vol. 810 LNEE, 2023.
46. S. Liang, S. Tian, L. Yu, and X. Kang, "Improved u-net based on contour attention for efficient segmentation of skin lesion," *Multimedia Tools and Applications*, vol. 83, 2024.

47. E. Hancer, M. Traoré, R. Samet, Z. Yıldırım, and N. Nemati, “An imbalance-aware nuclei segmentation methodology for h&e stained histopathology images,” *Biomedical Signal Processing and Control*, vol. 83, 2023.
48. H. Li, X. Zhao, A. Su, H. Zhang, J. Liu, and G. Gu, “Color space transformation and multi-class weighted loss for adhesive white blood cell segmentation,” *IEEE Access*, vol. 8, 2020.
49. N. V. Chawla, “Data mining for imbalanced datasets: An overview,” *Data Mining and Knowledge Discovery Handbook*, 2006.
50. H. Abdi and L. J. Williams, “Principal component analysis. wiley interdisciplinary reviews: computational statistics,” *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010.
51. L. V. D. Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of Machine Learning Research*, vol. 9, 2008.
52. K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” *Neural and Evolutionary Computing*, 2015.
53. T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, 1967.
54. V. N. Balasubramanian, S. Chakraborty, and S. Panchanathan, “Conformal predictions for information fusion: A comparative study of p-value combination methods,” *Annals of Mathematics and Artificial Intelligence*, vol. 74, 2015.
55. J. H. Friedman, J. L. Bentley, and R. A. Finkel, “An algorithm for finding best matches in logarithmic expected time,” *ACM Transactions on Mathematical Software (TOMS)*, vol. 3, 1977.

56. Y. Chen, L. Zhou, Y. Tang, J. P. Singh, N. Bouguila, C. Wang, H. Wang, and J. Du, "Fast neighbor search by using revised k-d tree," *Information Sciences*, vol. 472, 2019.
57. F. Zheng, R. Tao, H. R. Maier, L. See, D. Savic, T. Zhang, Q. Chen, T. H. Assumpção, P. Yang, B. Heidari, J. Rieckermann, B. Minsker, W. Bi, X. Cai, D. Solomatine, and I. Popescu, "Crowdsourcing methods for data collection in geophysics: State of the art, issues, and future directions," *Reviews of Geophysics*, vol. 56, 2018.
58. J. C. L. Bayas, L. See, S. Fritz, T. Sturn, C. Perger, M. Dürauer, M. Karner, I. Moorthy, D. Schepaschenko, D. Domian, and I. McCallum, "Crowdsourcing in-situ data on land cover and land use using gamification and mobile technology," *Remote Sensing*, vol. 8, 2016.
59. N. Kussul, L. Mykola, A. Shelestov, and S. Skakun, "Crop inventory at regional scale in ukraine: developing in season and end of season crop maps with multi-temporal optical and sar satellite imagery," *European Journal of Remote Sensing*, vol. 51, 2018.
60. H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Computing Surveys*, vol. 52, 2019.
61. N. Kussul, A. Shelestov, M. Lavreniuk, I. Butko, and S. Skakun, "Deep learning approach for large scale land cover mapping based on remote sensing data fusion," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2016-November, pp. 198–201, 2016.
62. M. Banko and E. Brill, "Scaling to very very large corpora for natural language disambiguation," in *Proceedings of the 39th Annual Meeting on Association for*

Computational Linguistics - ACL '01, pp. 26–33, Association for Computational Linguistics, 2001.

63. V. Makarichev, I. Vasilyeva, V. Lukin, B. Vozel, A. Shelestov, and N. Kussul, “Discrete atomic transform-based lossy compression of three-channel remote sensing images with quality control,” *Remote Sensing*, vol. 14, 2022.
64. M. Lavreniuk, N. Kussul, and A. Novikov, “Deep learning crop classification approach based on sparse coding of time series of satellite data,” in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, vol. 2018-July, pp. 4812–4815, 2018.
65. A. Shelestov, M. Lavreniuk, N. Kussul, A. Novikov, and S. Skakun, “Large scale crop classification using google earth engine platform,” in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 2017-July, pp. 3696–3699, 2017.
66. V. S. F. Garnot, L. Landrieu, S. Giordano, and N. Chehata, “Time-space tradeoff in deep learning models for crop classification on satellite multi-spectral image time series,” in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 6247–6250, 2019.
67. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, 2015.
68. A. Mao, M. Mohri, and Y. Zhong, “Cross-entropy loss functions: Theoretical analysis and applications,” in *Proceedings of Machine Learning Research*, vol. 202, 2023.

69. C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, and D. Rueckert, "Gan augmentation: Augmenting training data using generative adversarial networks," *CoRR*, vol. abs/1810.10863, 10 2018.
70. P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, 2017.
71. P. Wang and X. Bai, "Thermal infrared pedestrian segmentation based on conditional gan," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 28, 2019.
72. D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
73. W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation*, vol. 29, 2017.
74. R. Verma, N. Kumar, A. Patil, N. C. Kurian, S. Rane, S. Graham, Q. D. Vu, M. Zwager, S. E. Raza, N. Rajpoot, X. Wu, H. Chen, Y. Huang, L. Wang, H. Jung, G. T. Brown, Y. Liu, S. Liu, S. A. F. Jahromi, A. A. Khani, E. Montahaei, M. S. Baghshah, H. Behroozi, P. Semkin, A. Rassadin, P. Dutande, R. Lodaya, U. Baid, B. Baheti, S. Talbar, A. Mahbod, R. Ecker, I. Ellinger, Z. Luo, B. Dong, Z. Xu, Y. Yao, S. Lv, M. Feng, K. Xu, H. Zunair, A. B. Hamza, S. Smiley, T. K. Yin, Q. R. Fang, S. Srivastava, D. Mahapatra, L. Trnavska, H. Zhang, P. L. Narayanan, J. Law, Y. Yuan, A. Tejomay, A. Mitkari, D. Koka, V. Ramachandra, L. Kini, and A. Sethi, "Monusac2020: A multi-organ nuclei segmentation and classification challenge," *IEEE Transactions on Medical Imaging*, vol. 40, 2021.

75. V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, 2017.