



<https://doi.org/10.36023/ujrs.2023.10.3.244>

УДК 528.854.4:519.237.8

## Методика класифікування об'єктів на аеро- та космічних зображеннях в умовах низької розділимості розпізнавальних ознак

А. А. Андреев\*

ДУ “Науковий центр аерокосмічних досліджень Землі ІГН НАН України”, вул. Олеся Гончара, 55-Б, Київ, 01054, Україна

Методи класифікування використовуються в дистанційному дослідженні Землі (ДЗЗ) для дослідження природних ресурсів, моніторингу довкілля та вирішення багатьох інших завдань. Також вони відіграють важливу роль у залученні даних ДЗЗ до вирішення завдань, пов'язаних із досягненням цілей сталого розвитку. Серед них контрольовані та неконтрольовані методи класифікування об'єктів на аеро- та космічних зображеннях. Але для більшості тематичних завдань ДЗЗ доцільним є використання саме контрольованих методів класифікування, тому що розглянуті задачі потребують задання ознак вихідних класів. У контрольованих методах класифікування ознаки задаються навчальною вибіркою. Серед відомих підходів до обробки навчальної вибірки можна виокремити такі: кластерна вибірка; підходи, що дають змогу зменшити розмірність навчальної вибірки; підходи, що виявляють репрезенти, які були віднесені до вибірки не того класу, якому відповідає їхня класова належність. Проте, їх спільним недоліком є те, що вони не враховують фактор розділимості навчальної вибірки. Ця властивість безпосередньо впливає на достовірність класифікування. В дослідженні запропоновано методику, метою якої є підвищення достовірності класифікування об'єктів на аеро- та космічних зображеннях шляхом підвищення розділимості навчальної вибірки. Ця методика включає метод оцінювання розділимості навчальної вибірки. При цьому, є можливість оцінити розділивість як двох окремих класів, так і всього набору навчальної вибірки в цілому. Розроблена методика має дві гілки застосування: зменшення розмірності навчальної вибірки та кластеризація навчальної вибірки. В дослідженні було експериментально перевірено ефективність цієї методики на трьох прикладах. У двох прикладах застосування методики полягало в кластеризації навчальної вибірки. В першому загальна точність класифікації зросла на 4% (з 77 до 81%), а в другому – на 20% (з 63 до 83%). До третього прикладу було застосовано зменшення розмірності навчальної вибірки. Результатом чого стало зменшення розмірності вхідних даних із 167 до 57 шарів, тобто розмірність зменшилася у 2.92 раза. Також було підвищено загальну точність класифікування на 2% (з 91 до 93%).

**Ключові слова:** аеро- та космічні зображення, неконтрольоване класифікування, контрольоване класифікування, кластеризація, навчальна вибірка, розділивість навчальної вибірки.

© А. А. Андреев. 2023

### 1. Вступ

Новітні досягнення в галузі дистанційних методів класифікування створюють нові можливості для дослідження природних ресурсів, моніторингу довкілля та вирішення багатьох інших завдань (Melesse et al., 2007). Методи класифікування відіграють важливу роль у залученні даних ДЗЗ до вирішення завдань, пов'язаних із досягненням цілей сталого розвитку (Scott & Rajabifard, 2017). Фактично більшість методів, що використовуються для обробки даних ДЗЗ, засновані на методах класифікування (Andries et al., 2019, Landry et al., 2019).

Процедуру класифікування в ДЗЗ визначають як процес присвоєння різних категорій ґрунтового покриву або землекористування пікселям або сегментам в межах аеро- та космічних зображень. Цей процес передбачає використання комп'ютерних алгоритмів і методів для диференціації та класифікування різних об'єктів або ознак, які є на зображенні, на основі їхніх спектральних,

просторових і контекстних характеристик (Lillesand et al., 2014). При використанні аеро- та космічних зображень ознаки об'єктів визначаються каналами вхідних зображень та похідними шарами, що отримуються з каналів вхідних зображень або з додаткових джерел.

Методи класифікування в основному поділяють на дві групи, а саме: контрольовані (англ. *Supervised*) і неконтрольовані (*Unsupervised*) (Berry et al., 2019).

Методи неконтрольованого класифікування полягають в розподілі вхідних об'єктів на групи (класи) подібних об'єктів без залучення інформації про ознаки вхідних класів (Ferreira et al., 2017; Ruppert, 2004). Серед методів неконтрольованого класифікування найчастіше застосовують центроїдні методи та методи, що базуються на щільності ознак об'єктів.

Методи контрольованого класифікування полягають у віднесенні об'єкта до того класу, ознаки якого є найбільш подібними до ознак означеного об'єкта серед інших класів (Lillesand et al., 2014). На відміну від неконтрольованого класифікування, ознаки вхідних класів задаються перед процедурою

\*E-mail: [artem.a.andreev@gmail.com](mailto:artem.a.andreev@gmail.com)

класифікування як частина вхідних даних. Ознаки класів можуть бути подані у формі набору навчальної вибірки, який містить вибірку кожного із класів.

Отже, основна відмінність між контрольованими і неконтрольованими методами полягає в тому, що в першому випадку ознаки вихідних класів задаються експертом, а в неконтрольованих методах класифікування вихідні класи формуються на основі подібності ознак вхідних об'єктів.

Більшість тематичних задач ДЗЗ потребують класифікування із заданими класами. Таким чином, ознаки класів, що визначені тематичною задачею, мають бути задані перед процедурою класифікування, тобто, для цього підходять саме методи контрольованого класифікування. В цих методах ознаки класів задаються у формі набору навчальної вибірки.

Серед властивостей навчальної вибірки можна виділити такі властивості, як повнота, рівномірність, чистота, суперечності, компактність, складність та розмірність (Subbotin, 2010).

Основними підходами та методами до обробки навчальної вибірки є: кластерна вибірка (Sedgwick, 2014); підходи, що дають змогу зменшити розмірність навчальної вибірки (Salem & Hussein, 2019; Green et al., 1988); підходи, що виявляють репрезенти, які були віднесені до вибірки не того класу, якому відповідає їх класова належність (Kang et al., 2018). Спільним недоліком розглянутих підходів є те, що вони не враховують фактор розділюваності навчальної вибірки. Ця властивість безпосередньо впливає на достовірність класифікування.

Саме тому підвищення розділюваності навчальної вибірки є актуальним завданням в класифікуванні об'єктів на аеро- та космічних зображеннях. Метою дослідження є підвищення достовірності класифікування об'єктів на аеро- та космічних зображеннях. Для досягнення зазначеної мети викладено методика, яка заснована на кластеризації та зменшенні розмірності навчальної вибірки.

## 2. Оцінювання розділюваності навчальної вибірки

Покроковий алгоритм оцінювання розділюваності навчальної вибірки зображений на Рис. 1. На першому етапі класифікується навчальна вибірка, при цьому вона використовується спочатку для навчання класифікатора, а потім кожен елемент цієї вибірки буде прокласифіковано цим класифікатором та віднесено до того чи іншого класу.



Рис. 1. Блок-схема алгоритму оцінювання розділюваності навчальної вибірки

Наступним етапом є формування матриці помилок (Роров, 2007). Потім на основі отриманої

матриці розраховується індекс розділюваності. Цей індекс може бути розраховано як для всього набору навчальної вибірки в цілому, так і для двох окремих класів. Оцінювання розділюваності двох класів навчальної вибірки базується на розрахунку середнього арифметичного значення показників чутливості (*sensitivity*) та специфічності (*specificity*) (Starovoitov & Golub, 2020). Результатом виконання цього алгоритму буде кількісна оцінка розділюваності навчальної вибірки.

Значення чутливості розраховується за формулою:

$$sensitivity = \frac{x_{ii}}{x_{ii} + x_{ji}}$$

де  $x_{ij}$  – це кількість репрезентів, які належать до класу  $i$ , але віднесені після класифікування до класу  $j$ , відповідно  $x_{ii}$  – це кількість правильно класифікованих репрезентів класу  $i$  (той, який після класифікування віднесений до того класу, до якого він дійсно належить).

Значення специфічності розраховується за формулою:

$$specificity = \frac{x_{jj}}{x_{jj} + x_{ij}}$$

Таким чином, формула розрахунку індексу розділюваності двох класів така:

$$SITS_{pair} = \frac{sensitivity + specificity}{2} = \frac{\frac{x_{ii}}{x_{ii} + x_{ji}} + \frac{x_{jj}}{x_{jj} + x_{ij}}}{2}$$

Розділюваність всього набору навчальної вибірки оцінюється на основі показників загальної точності класифікації (Роров, 2007) або капа-індексу (Bishop et al., 2007).

Формула, за якою розраховується цей показник:

$$OA = \frac{\sum_{i=1}^K x_{ii}}{\sum_{i=1}^K \sum_{j=1}^K x_{ij}}$$

де  $K$  – кількість класів.

Проте, цей показник не враховує класовий дисбаланс тестових об'єктів, за якими оцінювалася точність класифікування. Саме тому введено показник капа-індекс, який дає адекватну оцінку точності класифікування навіть при дисбалансі тестових об'єктів. Цей показник розраховується за формулою:

$$Kappa = \frac{\sum_{i=1}^K \sum_{j=1}^K x_{ij} * \sum_{i=1}^K x_{ii} - \sum_{i=1}^K x_{ij} \sum_{j=1}^K x_{ij} * \sum_{j=1}^K x_{ji}}{(\sum_{i=1}^K \sum_{j=1}^K x_{ij})^2 - \sum_{i=1}^K x_{ij} \sum_{j=1}^K x_{ij} * \sum_{j=1}^K x_{ji}}$$

Значення розглянутих індексів знаходяться від 0 до 1. При цьому, значення 0 показує, що отримана класифікація є повністю неточною, а значення 1 відповідає точній класифікації, яка не має жодного неправильно класифікованого об'єкта.

## 3. Методика

Розроблена методика класифікування об'єктів на аеро- та космічних зображеннях в умовах низької розділюваності розпізнавальних ознак має дві гілки

застосування. Спільним в них є те, що обидві націлені на підвищення розділимості навчальної вибірки.

Перша гілка полягає в зменшенні розмірності навчальної вибірки, друга – в кластеризації навчальної вибірки. Кожен з варіантів можна застосувати при будь-яких вхідних даних, проте, доцільно при виборі методики оцінити об'єм даних. У випадку, коли експерт вважає, що об'єм даних є великим (наприклад, перевищує наявні обчислювальні ресурси), то більш доречним є зменшення розмірності навчальної вибірки. Якщо експерт вважає, що кількість шарів вхідного куба геопросторових даних є малою, то обирається кластеризація навчальної вибірки.

### 3.1. Зменшення розмірності навчальної вибірки

Геопросторові дані, які отримуються з вхідних різномірних аеро- та космічних зображень чи інших джерел, можуть бути подані у вигляді растрових шарів. Для об'єднання растрових шарів геопросторових даних в один масив прийнято формувати куб геопросторових даних (Montero et al., 2023).

При певних задачах вхідний куб може мати занадто великий розмір. Серед основних проблем, які спричинені надлишковим об'ємом куба геопросторових даних, можна виділити дві (Li et al., 2022). По-перше, використання великої кількості шарів куба геопросторових даних потребує більше обчислювальних ресурсів. По-друге, через те, що сигнатури навчальної вибірки визначені в кожному з шарів куба геопросторових даних, розділимість навчальної вибірки може бути знижена, якщо присутні шари або неправильно створені, або не є актуальними для обраної тематичної задачі.

Розв'язанням вказаних проблем вбачається зменшення розмірності навчальної вибірки та куба геопросторових даних в цілому (Luo et al., 2020, Huang et al., 2019). Розроблена методика, окрім зменшення розмірності, спрямована також і на підвищення розділимості навчальної вибірки. Вона є процедурою оптимізації, в якій цільовою функцією розглядається показник, що визначає розділимість навчальної вибірки, а саме – загальний індекс розділимості навчальної вибірки, який зазначений вище. Таким чином, задачу цієї оптимізаційної процедури можна описати як пошук куба з найменшою кількістю шарів, при яких відповідна навчальна вибірка досягає найвищого значення загального індексу розділимості серед усіх інших кубів та їхніх відповідних наборів навчальної вибірки.

Блок-схему алгоритму зменшення розмірності навчальної вибірки зображено на Рис. 2. Як показано на цій блок-схемі, такий алгоритм є ітеративною процедурою.

На кожній ітерації шляхом відкидання від поточного куба по одному шару формується набір кубів. Кількість новоутворених кубів рівна кількості шарів поточного куба. Далі визначається той куб, який має найвище значення загального індексу розділимості навчальної вибірки. Якщо поточне

значення індексу розділимості не нижче за те, що було отримане на попередній ітерації, то цикл переходить до наступної ітерації, вже розкладаючи куб, який отриманий на цій ітерації. Інакше процедура зупиняється і оптимальним буде вважатися той куб геопросторових даних, який був отриманий на попередній ітерації. Також ітеративна процедура завершиться, якщо на цій ітерації отримано куб геопросторових даних, який містить тільки один шар.

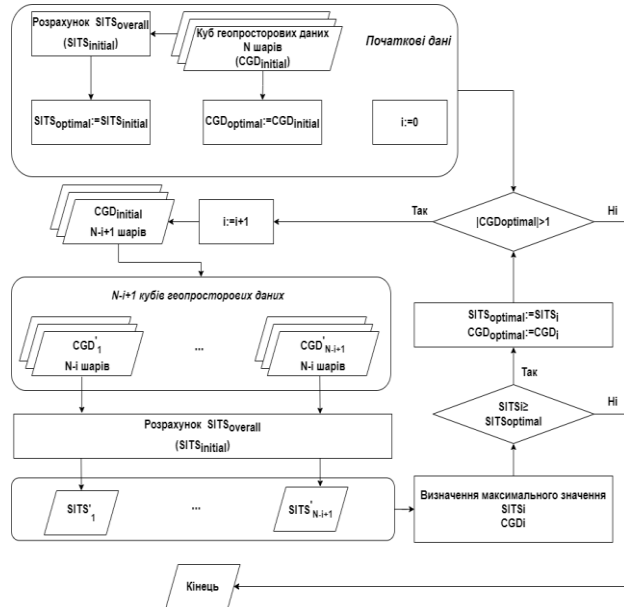


Рис. 2. Блок-схема алгоритму зменшення розмірності навчальної вибірки

Результатом виконання цієї процедури є оптимальний куб геопросторових даних та відповідний набір навчальної вибірки, розмірність яких нижче за початкову, а розділимість – вище. Отримані дані будуть використані для подальшого класифікування.

### 3.2. Кластеризація навчальної вибірки

Кластеризація набору навчальної вибірки здійснюється для підвищення розділимості цього набору. Через те, що розроблена методика при кластеризації розглядає центроїдні методи (Omran et al., 2007), мету можна визначити як пошук оптимальної кількості кластерів для кожного з класів навчальної вибірки. Такий пошук є задачею оптимізації, в якій цільовою функцією є загальний індекс розділимості навчальної вибірки. Обраний показник дає кількісну оцінку розділимості між сигнатурами різних класів. Таким чином, кількість кластерів для кожного з класів навчальної вибірки є оптимальною за умови, що відповідний кластеризований набір навчальної вибірки матиме найвище серед усіх інших наборів значення загального індексу розділимості навчальної вибірки.

Алгоритм кластеризації навчальної вибірки зображено на блок-схемі (Рис. 3).

Як показано на цій блок-схемі, якщо значення загального індексу розділимості початкового набору навчальної вибірки ( $SITS_{Overall}$ ) дорівнює своєму абсолютному максимуму, а саме 1, то цей набір вже

є оптимальним, тому кластеризація проводиться не буде. В іншому випадку оптимізація полягатиме в ітеративному пошуку оптимального числа кластерів для тих пар, які мають найнижче значення індексу розділимості двох класів  $SITS_{pair}$ .

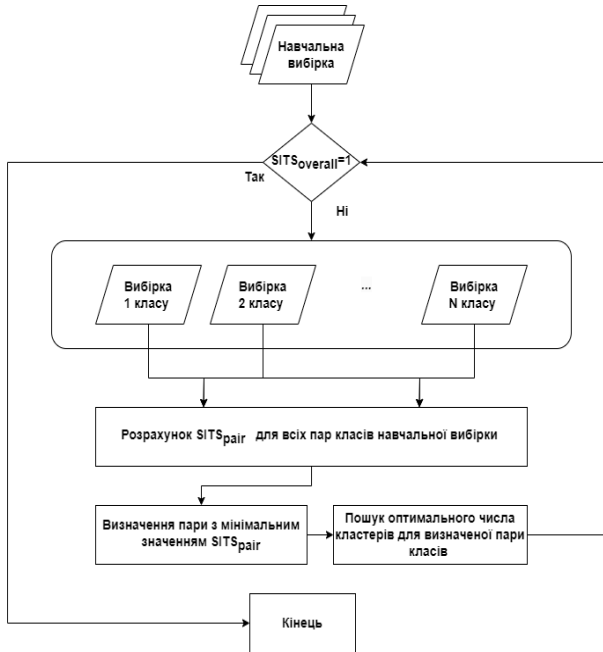


Рис. 3. Блок-схема алгоритму кластеризації навчальної вибірки

Серед умов зупинки цієї оптимізаційної процедури можна виділити досягнення значення 1 загального індексу розділимості набору навчальної вибірки на будь-якому кроці. Також поточне число кластерів для кожного з класів навчальної вибірки буде вважатися оптимальним тоді, якщо при пошуку оптимальної кількості кластерів для всіх поспіль пар класів не було жодного підвищення загального індексу розділимості.

Результатом виконання такої процедури є кластеризована навчальна вибірка, яка буде використана для подальшого класифікування.

#### 4. Практичне застосування розробленої методики

Розроблена методика була застосована до 3 задач. Однією з них є класифікування боліт, де територією дослідження обрано Шацький контрольно-калібрувальний полігон ДЗЗ (Andreiev & Kozlova, 2021). Тематичною задачею було визначено 6 класів, а саме: болота, ліси, луки, водні об'єкти, штучні поверхні та інші землі. Як вхідні дані використано лише канали одного багатоспектрального космічного знімка Sentinel-2 без даних з додаткових джерел. Методом керованого класифікування обрано відстань Махаланобіса (Bruzzone & Demir, 2014). Застосування методики полягало в кластеризації навчальної вибірки. Для цієї процедури обрано метод некерованого класифікування K-Середніх (Jain & Dubes, 1988). Перевірка ефективності розробленої методики полягала в проведенні двох процедур класифікування та їх порівнянні між собою. Перша процедура – перед кластеризацією навчальної

вибірки, друга – після кластеризації. Таким чином, загальна точність класифікації зросла на 4 % (з 77 до 81 %).

Іншим прикладом застосування розробленої методики є класифікування, де територією дослідження була ділянка в Івано-Франківській області (Andreiev, 2020). Цією тематичною задачею було визначено 6 класів: сільськогосподарські угіддя, ліси, луки, водні об'єкти, штучні поверхні та інші землі. Вхідні дані – канали багатоспектрального космічного знімка Landsat-OLI8. Аналогічно до попереднього експерименту було проведено кластеризацію навчальної вибірки та порівняно між собою дві процедури класифікування: до та після кластеризації навчальної вибірки. Методом керованого класифікування обрано відстань максимальна правдоподібність. Для кластеризації навчальної вибірки обрано метод некерованого класифікування – K-Середніх. У підсумку, загальна точність класифікації зросла на 20 % (з 63 до 83 %).

Гілка розробленої методики, яка полягає в зменшенні розмірності навчальної вибірки, була застосована до аналогічної з першого прикладу задачі. Визначені такі ж 6 класів, територія дослідження – Шацький контрольно-калібрувальний полігон ДЗЗ. Проте, як вхідні дані було використано часовий ряд, який містив спектральні канали космічних знімків Sentinel-2 за три дати, відповідні спектральні індекси та допоміжні геоморфологічні параметри. Методом керованого класифікування обрано відстань Махаланобіса. Після застосування методики до вхідного набору даних вдалося скоротити його розмірність із 167 до 57 шарів, тобто розмірність зменшилася у 2.92 раза. Разом з тим, було підвищено загальну точність класифікування за отриманим набором даних щодо початкового набору, а саме на 2 % (з 91 до 93 %).

#### 5. Висновки

Розглянуто контрольовані та неконтрольовані методи класифікування об'єктів на аеро- та космічних зображеннях. Обґрунтовано, що для більшості тематичних задач ДЗЗ доцільно використати саме контрольовані методи класифікування, адже вони дають змогу задати ознаки вихідних класів у вигляді набору навчальної вибірки. Проаналізовано властивості набору навчальної вибірки. Виділено такі властивості, як повнота, рівномірність, чистота, суперечності, компактність, складність та розмірність. Проведено аналіз підходів до обробки навчальної вибірки. Серед них: кластерна вибірка; підходи, що дають змогу зменшити розмірність навчальної вибірки; підходи, що виявляють репрезенти, які були віднесені до вибірки не того класу, якому відповідає їхня класова належність. Згідно з проведеним аналізом, їх спільним недоліком є те, що вони не враховують фактор розділимості навчальної вибірки. Проте, ця властивість безпосередньо впливає на достовірність класифікування.

Розроблено методику класифікування об'єктів на аеро- та космічних зображеннях в умовах низької

розділимості розпізнавальних ознак. Методика включає в себе метод оцінювання розділимості навчальної вибірки. Цей метод дає змогу оцінити розділимість як двох окремих класів навчальної вибірки, так і всього набору в цілому. Перше оцінюється на основі розрахунку середнього арифметичного показників чутливості та специфічності. Друге – на основі різних критеріїв, які оцінюють достовірність класифікування в цілому, наприклад, на основі загальної точності класифікації або капа-індексу. Також, ця методика розглядає дві гілки застосування: зменшення розмірності навчальної вибірки та кластеризація навчальної вибірки.

Методика була практично застосована до трьох задач. В кожному з прикладів перевірка ефективності розробленої методики полягала в проведенні двох процедур класифікування та їх порівнянні між собою. Перша процедура – до застосування розробленої методики, друга – після. В двох прикладах застосування методики полягало в кластеризації навчальної вибірки. В першому прикладі загальна точність класифікації зросла на 4 % (з 77 до 81 %), а в другому – на 20 % (з 63 до 83 %). До третього прикладу було застосовано зменшення розмірності навчальної вибірки. Результатом чого стало зменшення розмірності вхідних даних із 167 до 57 шарів, тобто розмірність зменшилася у 2.92 раза. Також було підвищено загальну точність класифікування на 2 % (з 91 до 93 %).

Таким чином, показано, що застосування розробленої методики є ефективним до задач класифікування об'єктів на аеро- та космічних знімках в умовах низької розділимості їх ознак.

## Література / References

- Andreev, A. A. (2020). Hybrid approach to classification of remote sensing data. *CERes Journal*, 6(2), 32–37.
- Andreev, A., & Kozlova, A. (2021). Enhancement of land cover classification by training samples clustering. In *Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference*, 223–227. Minsk, Belarus: UIIP NASB. ISBN 978-985-7198-07-8.
- Andries, A., Morse, S., Murphy, R. et al (2019) Translation of Earth observation data into sustainable development indicators: an analytical framework. *Sustain Dev*, 27, 366–376. <https://doi.org/10.1002/sd.1908>.
- Berry, M., Mohamed, A., & Yap, B. W. (2019). Supervised and Unsupervised Learning for Data Science. *Springer, Cham*.
- Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). Discrete multivariate analysis: Theory and Practice. *Springer Science & Business Media*.
- Bruzzzone, L. & Demir, B. A. (2014). A review of modern approaches to classification of remote sensing data. In *Land use and land cover mapping in Europe*, I. Manakos, M. Braun, Eds. Springer: Dordrecht, Netherlands, 127–143.
- Ferreira, B., Silva, R. G., & Pereira, V. (2017). Feature selection using non-binary decision trees applied to condition monitoring. *22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. <https://doi.org/10.1109/etfa.2017.8247642>.
- Green, A., Berman, M., Switzer, P., & Craig, M. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65–74. <https://doi.org/10.1109/36.3001>.
- Huang, H., Shi, G., He, H., Duan, Y., & Luo, F. (2019). Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning. *IEEE transactions on cybernetics*, 50(6), 2604–2616.
- Jain, A. K., & Dubes, R. C. (1988). Algorithms for Clustering Data. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Kang, X., Xiang, X., Li, S., & Benediktsson, J. A. (2018). Detection and correction of mislabeled training samples for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10), 5673–5686. <https://doi.org/10.1109/tgrs.2018.2823866>.
- Landry T, Sotir M, Rajotte J-F, et al (2019) Applying machine learning to earth observations in a standards based workflow. *IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE*, 5567–5570.
- Li, H., Cui, J., Zhang, X., Han, Y., & Cao, L. (2022). Dimensionality reduction and classification of hyperspectral remote sensing image feature extraction. *Remote Sensing*, 14(18), 4579. <https://doi.org/10.3390/rs14184579>.
- Lillesand, T. M., Kiefer, R. W., & Chipman, J. W. (2014). *Remote Sensing and Image Interpretation*. Wiley.
- Luo, F., Zhang, L., Du, B., & Zhang, L. (2020). Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(8), 5336–5353.
- Melesse, A. M., Weng, Q., Thenkabail, P. S., & Senay, G. B. (2007). Remote sensing sensors and applications in environmental resources mapping and modelling. *Sensors*, 7(12), 3209–3241. <https://doi.org/10.3390/s7123209>.
- Montero, D., Kraemer, G., Anghela, A., Camacho, C. A., Brandt, G., Camps-Valls, G., Cremer, F., Flik, I., Gans, F., Habershon, S., Ji, C., Kattenborn, T., Martínez-Ferrer, L., Martinuzzi, F., Reinhardt, M., Söchtting, M., Teber, K., & Mahecha, M. (2023a). Data Cubes for Earth System research: Challenges ahead. *EarthArXiv (California Digital Library)*. <https://doi.org/10.31223/x58m2v>.
- Omran, M. G. H., Engelbrecht, A. P., & Salman, A. A. (2007). An overview of clustering methods. *Intelligent Data Analysis*, 11(6), 583–605. <https://doi.org/10.3233/ida-2007-11602>.
- Popov, M. O. (2007). Methodology of accuracy assessment of classification of objects on space images. *Journal of Automation and Information Sciences*, 39, 1–10. <https://doi.org/10.1615/JAutomatInfScien.v39.i1.50>.
- Ruppert, D. (2004). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Journal of the American Statistical Association*, 99(466), 567. <https://doi.org/10.1198/jasa.2004.s339>.
- Salem, N., & Hussein, S. (2019). Data dimensional reduction and principal components analysis. *Procedia Computer Science*, 163, 292–299. <https://doi.org/10.1016/j.procs.2019.12.111>.
- Scott, G., & Rajabifard, A. (2017). Sustainable development and geospatial information: a strategic framework for integrating a global policy agenda into national geospatial capabilities. *Geo-spatial Information Science*, 20(2), 59–76. <https://doi.org/10.1080/10095020.2017.1325594>.
- Sedgwick, P. (2014). Cluster sampling. *BMJ*, 34, 1215. <https://doi.org/10.1136/bmj.g1215>.
- Starovoitov, B. B., & Golub, Y. I. (2020). Comparative study of quality estimation of binary classification. *Informatika*, 17(1), 87–101. <https://doi.org/10.37661/1816-0301-2020-17-1-87-101>.

Subbotin, S. (2010). The training set quality measures for neural network learning. *Optical Memory and Neural*

*Networks*, 19(2), 126–139. <https://doi.org/10.3103/s1060992x10020037>.

#### AN OBJECT CLASSIFICATION TECHNIQUE ON AERIAL AND SPACE IMAGERY UNDER LOW SEPARABILITY OF RECOGNITION FEATURES

A. A. Andreiev

*Scientific Centre for Aerospace Research of the Earth of the Institute of Geological Sciences of the National Academy of Sciences of Ukraine, Olesia Honchara Str., 55-b, Kyiv, 01054, Ukraine*

Classification methods are used in remote sensing of the Earth to study natural resources, monitor the environment, and solve many other problems. Also, they play an important role in involving the remotely sensed data in solving tasks related to achieving sustainable development goals. Among them are supervised and unsupervised classification methods on aerial and space imagery. However, for most thematic problems, it is advisable to use precisely supervised classification methods because the considered problems require setting the characteristics of the initial classes. In supervised classification methods, the features are given by the training sample. Among the well-known approaches to processing the training sample, the following can be distinguished: cluster sampling; approaches to reduce the size of the training sample; approaches that detect representatives who were assigned to the sample of the wrong class to which their class membership corresponds. However, their common disadvantages is that they do not consider the factor of separation of the training sample. This property directly affects the reliability of the classification. The research proposed a technique to increase the reliability of object classification on aerial and space imagery by increasing the separability of the training sample. This technique includes a method of assessing the separability of the training sample. At the same time, it is possible to assess the separability of both two separate classes and the entire set of the educational sample. The developed technique has two branches of application: reducing the size of the training sample and clustering the training sample. In the study, the effectiveness of this technique was experimentally tested on three examples. In two examples, the technique was used to cluster the training sample. In one of these examples, the overall accuracy of the classification increased by 4 % (from 77 to 81%), and in the second one by 20 % (from 63 to 83 %). A reduction in the size of the training sample was applied to the third example. As a result, the dimensionality of the input data was reduced from 167 to 57 layers. That is, the dimensionality decreased by 2.92 times. Also, the overall accuracy of the classification was increased by 2 % (from 91 to 93 %).

**Keywords:** aerial and space images, unsupervised classification, supervised classification, clustering, training sample, training sample separability.

*Рукопис статті отримано 01.09.2023  
Надходження остаточної версії: 08.09.2023  
Публікація статті: 29.09.2023*