НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ДЕРЖАВНА УСТАНОВА «НАУКОВИЙ ЦЕНТР АЕРОКОСМІЧНИХ ДОСЛІДЖЕНЬ ЗЕМЛІ ІНСТИТУТУ ГЕОЛОГІЧНИХ НАУК НАЦІОНАЛЬНОЇ АКАДЕМІЇ НАУК УКРАЇНИ»

На правах рукопису

АЛЬПЕРТ СОФІЯ ІОГАНІВНА

УДК 528.85(477.41)

МЕТОДИ КОНТРОЛЬОВАНОГО КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ В УМОВАХ НАВЧАЛЬНИХ ВИБІРОК ОБМЕЖЕНОЇ ЯКОСТІ

05.07.12 – Дистанційні аерокосмічні дослідження

Дисертація
на здобуття наукового ступеня
кандидата технічних наук

Науковий керівник – д. т. н., професор
Попов Михайло Олексійович

Київ – 2016

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ……5
ВСТУП……6
РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ. ПОСТАНОВКА ЗАВДАННЯ НА ДОСЛІДЖЕННЯ……15
1.1. Місце і роль процедури класифікування космічних зображень при розв'язанні тематичних задач дистанційного зондування Землі……..15
1.2. Огляд методів і алгоритмів класифікування гіперспектральних космічних зображень……16
1.3. Постановка завдання на дослідження……..……27
Висновки до розділу 1 ……31

РОЗДІЛ 2 РОЗРОБКА МЕТОДУ КОНТРОЛЬОВАНОГО КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ В УМОВАХ ЗАБРУДНЕНИХ НАВЧАЛЬНИХ ВИБІРОК……33
2.1. Основні положення теорії свідчень Демпстера-Шейфера……33
2.2. Метод класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок……...37
2.2.1. Модель гіперспектрального космічного зображення зображення та припущення……..……37
2.2.2. Організація і зміст методу класифікування гіперспектральнх космічних зображень в умовах забруднених навчальних вибірок ……38
2.2.2.1. Відбір зональних зображень……..40
2.2.2.2. Фільтрація навчальних вибірок……43
2.2.2.3. Розбиття ознакового простору……..45
2.2.2.4. Визначення класифікаційної цінності спектральних каналів……..47
2.2.2.5. Алгоритм класифікування……54

Висновки до розділу 2 ……..56

РОЗДІЛ 3 РОЗРОБКА МЕТОДУ КОНТРОЛЬОВАНОГО КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ В УМОВАХ ЗАБРУДНЕННЯ НАВЧАЛЬНИХ ВИБІРОК ТА КІЛЬКІСНИХ ОБМЕЖЕНЬ……...58

3.1. Обмеження, припущення, визначення……..58

3.2. Організація і процедури методу контрольованого класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень……60

3.2.1. Загальна схема класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень……60

3.2.2. Відбір зональних зображень……...62

3.2.3. Навчання……...72

3.2.4. Класифікування……75

3.3. Приклад застосування запропонованого методу класифікування……...76

Висновки до розділу 3 ……79

РОЗДІЛ 4 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТИ РОЗРОБЛЕНИХ МЕТОДІВ КОНТРОЛЬОВАНОГО КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ……80

4.1. Спеціалізоване програмне забезпечення для реалізації та дослідження розроблених методів контрольованого класифікування: архітектура і стислий опис……..80

4.2. Оцінка ефективності методу контрольованого класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок……91

4.3. Оцінка ефективності методу контрольованого класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень……100
Висновки до розділу 4 ……119
Загальні висновки по роботі…….……..121
Список використаних джерел…….....125
Додатки……138
Додаток 1……...138
Додаток 2……...140
Додаток 3……...148
Додаток 4……...152

ПЕРЕЛІК СКОРОЧЕНЬ

БД – база даних

БЗ – база знань

ГКЗ - гіперспектральне космічне зображення

ДЗЗ – дистанційне зондування Землі

СПЗ – спеціалізоване програмне забезпечення

ТСДШ – теорія свідчень Демпстера-Шейфера

DST – Dempster-Shafer Theory

SVM – Support Vector Machine (метод опорних векторів)

ВСТУП

Поступово, з кожним роком все більше актуальних і важливих задач вирішується із залученням методів і технологій дистанційного зондування Землі (ДЗЗ). Можливості ДЗЗ сьогодні використовуються при екологічному моніторингу територій, вивченні природних ресурсів, аналізі причин та наслідків надзвичайних ситуацій, при оперативному оцінюванні стану лісів і сільськогосподарських земель, тощо.

Новітні досягнення науки та техніки, такі як нанотехнології, поява принципово нових видових технічних засобів і методів цифрового оброблення зображень, удосконалення глобальних інформаційних комунікацій та ін., сприяють подальшому розвитку ДЗЗ. Тому за останні десятиліття суттєво збільшилися об’єми, різноманіття та якість матеріалів ДЗЗ, що, в свою чергу, сприяло розробці та впровадженню нових ефективних методів вивчення та дослідження геосистем.

Серед матеріалів, які одержуються шляхом знімання земної поверхні з борта космічного (рівно як й повітряного) літального апарату, гіперспектральні зображення є найбільш змістовними і вміщують в собі надзвичайно великий обсяг інформації про об'єкти зйомки та спостереження. Ця інформація дозволяє виявляти та розпізнавати (класифікувати) об'єкти на аерокосмічних знімках, оцінювати їх стан, фіксувати зміни, що відбуваються на місцевості, надавати обґрунтовані прогнозні оцінки і т. д. [1, 2].

Розв’язання подібних актуальних і важливих практичних завдань вимагає залучення великої кількості різних логіко-обчислювальних процедур, при цьому найбільш складною процедурою серед них є процедура класифікування об'єктів на гіперспектральному зображенні, оскільки, по-перше, розмірність ознакового простору при цьому є дуже великою (загальна кількість спектральних каналів обчислюється сотнями) і, по-друге, спектральні ознаки не відрізняються сталістю (величина зареєстрованого спектрального відгуку залежить від умов освітлення об'єкта, складу та прозорості атмосфери і т. д.).

В силу цього більшість комп’ютерних алгоритмів, які використовуються при класифікуванні гіперспектральних космічних зображень (ГКЗ), не зберігають свою ефективність у випадку гіперспектральних зображень. Тому останні роки розробляються алгоритми, які визначально орієнтуються на класифікування саме гіперспектральних зображень, проте їх точність поки залишається недостатньою, особливо в умовах, коли вхідна інформація, необхідна для проведення класифікування, має елементи невизначенності та/або неповноти. Одним з таких обмежень є недостатня якість навчальних вибірок.

У подібних умовах заслуговує на увагу підхід, оснований на використанні теорії свідчень Демпстера-Шейфера (ТСДШ). Математичний апарат ТСДШ [2–5] дозволяє будувати інтервальні оцінки достовірності гіпотез в умовах відсутності частини вхідних даних, пропонує просте правило для комбінування даних (об'єднання інформацій) від різних джерел, може у складних ситуаціях класифікування надавати багатоальтернативний розв’язок задачі (у вигляді кортежу з кількох гіпотез), а також має деякі інші переваги перед традиційним теоретико-ймовірнісним підходом. Означені властивості ТСДШ, зокрема, спосіб об’єднання даних за комбінаційним правилом Демпстера, покладені в основу розроблених автором нових методів класифікування ГКЗ.

Необхідно зазначити, що математичний апарат ТСДШ вже застосовувався рядом дослідників при аналізі та класифікуванні об'єктів на зображеннях. Запропоновані нами методи контрольованого класифікування відрізняються від відомих розв’язків однойменних задач введенням процедури оцінювання класифікаційної цінності спектральних каналів за допомогою спеціальної емпіричної функції і особливим способом розбиття спектрального ознакового простору, що, у сукупності, дозволяє застосовувати при класифікуванні лише найбільш інформативні спектральні канали і, таким чином, суттєво скоротити розмірність ознакового простору при одночасному підвищенні точності класифікування.

Зв'язок роботи з науковими програмами, планами, темами. Дослідження, які складають зміст дисертаційної роботи, виконувались у період 2012-2016 рр. у відповідності з планом наукових досліджень Наукового Центру аерокосмічних досліджень Землі ІГН НАН України і, зокрема відділу геоінформаційних технологій в ДЗЗ. Результати досліджень одержано автором в рамках виконання таких НДР: "Методологія комплексного геоінформаційного аналізу матеріалів аерокосмічних спостережень Землі і наземних даних в інтересах підвищення ефективності природокористування" (№ держреєстрації 0111U000030), "Методи інформаційного інтегрування даних багатоспектрального аерокосмічного знімання, польової спектрометрії і геолого-геофізичних даних при вирішенні завдань дистанційного пошуку покладів корисних копалин на суходолі і морському шельфі" (№ держреєстрації 0112U000702).

Мета та задачі дослідження. Метою дисертаційної роботи є підвищення точності контрольованого класифікування гіперспектральних космічних зображень в умовах навчальних вибірок обмеженої якості, зокрема, в умовах забруднення та малої кількості складових елементів.

Мета реалізується через постановку загального завдання розробити нові ефективні методи контрольованого класифікування ГКЗ в означених умовах.

Для вирішення поставленого завдання було сформульовано і досліджено такі задачі:

- аналіз сучасних методів класифікування ГКЗ;
- огляд відомих методів і підходів до класифікування, які основані на використанні положень ТСДШ;
- розробка нового методу контрольованого класифікування ГКЗ в умовах забруднення навчальних вибірок;

- розробка нового методу контрольованого класифікування ГКЗ в умовах забруднення навчальних вибірок та обмежень на їх кількість.

- алгоритмічна та програмна реалізація запропонованих методів контрольованого класифікування;

- проведення експериментальних досліджень розроблених методів контрольованого класифікування на реальних ГКЗ, зокрема, оцінювання точності пропонованих методів та їх порівняння із відомими сучасними методами і алгоритмами, такими, як метод опорних векторів (Support Vector Machine - SVM) та алгоритм класифікування об'єктів за їхніми спектрально-топологічними характеристиками.

Об'єкт дослідження: методи контрольованого класифікування гіперспектральних космічних зображень.

Предмет дослідження: точність методів контрольованого класифікування гіперспектральних космічних зображень в умовах навчальних вибірок обмеженої якості.

Методи дослідження:

- методи розпізнавання образів;
- математичний апарат ТСДШ;
- методи статистичного оцінювання;
- комп'ютерне моделювання.

Методи розпізнавання образів використано при формуванні загальних схем класифікування ГКЗ в умовах забруднених навчальних вибірок та обґрунтуванні організації оброблення даних.

Математичний апарат та положення ТСДШ дозволили запропонувати нові алгоритми формування ознакового простору та комбінування гіпотез, а також нове правило прийняття рішень.

Методи статистичного оцінювання були покладені в основу критеріального вибору та розрахунків показників ефективності запропонованих

методів класифікування, а також дозволили впровадити нову процедуру оцінювання класифікаційної цінності спектральних каналів за допомогою спеціальної емпіричної функції і скоротити таким чином розмірність ознакового простору при одночасному підвищенні точності класифікування.

Застосування комп'ютерного моделювання дозволило провести експериментальне дослідження розроблених методів на реальних ГКЗ.

Наукова новизна отриманих результатів. У результаті виконання дисертаційної роботи отримано нові наукові результати, які виносяться на захист:

1. Вперше, на основі математичного апарату та положень теорії свідчень Демпстера-Шейфера розроблені та обґрунтовані такі методи:

- метод класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок;

- метод класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок та обмежень на їх кількість.

2. Впроваджено нову функцію класифікаційної цінності спектральних каналів, що дозволяє на кількісній основі здійснювати відбір найбільш інформативних зональних зображень.

3. Запропоновано новий спосіб визначення репрезентативної комбінації зональних зображень при класифікуванні гіперспектрального знімка, який мінімізує навантаження на експерта і дає можливість зменшити вимоги до необхідного обчислювального ресурсу.

Обгрунтованість і достовірність наукових результатів підтверджується коректним використанням математичного апарату ТСДШ, експериментальною перевіркою розроблених методів на реальних гіперспектральних зображеннях, отриманих апаратурою Hyperion / EO-1,

проведенням оцінки точності класифікування запропонованих методів за допомогою загально прийнятих критеріїв.

Практичне значення отриманих результатів полягає у тому, що:

- показана можливість підвищення точності класифікування ГКЗ в умовах, коли існують обмеження на об'єм навчальних вибірок, а також на їх однорідність. Саме з такими умовами дослідник має справу при дослідженні змішаних лісів, вивченні природних та штучних об'єктів на урбанізованих територіях, у важко доступних місцях (наприклад, в гірській місцевості) тощо;

- сталість розроблених методів до забрудненості навчальних вибірок (робасність) дозволяє (зрозуміло, в певних границях) автоматично компенсувати помилки експерта-дослідника на етапі навчання системи і далі зберігати достатньо високу точність класифікування;

- розроблене спеціальне програмне забезпечення може бути використано при вирішенні багатьох тематичних задач ДЗЗ: оцінюванні стану лісів і сільськогосподарських земель, при вивченні природних ресурсів, екологічному моніторингу забруднених територій тощо.

Особистий внесок здобувача. Основою дисертаційної роботи стали результати особистих досліджень та персональних розробок автора. В опублікованих зі співавторами наукових роботах, у яких знайшли відображення деякі наукові результати дисертації, внесок здобувача полягав у наступному:

- участь у розробці методів відбору інформативних зональних зображень при класифікуванні ГКЗ;

- участь у розробці нових методів класифікування ГКЗ в умовах обмеженої якості навчальних вибірок, які базуються на використанні математичного апарату та прийомів, розвинених у ТСДШ;

- алгоритмічна та програмна реалізація розроблених методів класифікування.

Апробація результатів дисертації. Результати дисертаційних досліджень доповідались на наступних наукових конференціях і семінарах: XII Міжнародна конференція “Геоінформатика: теоретичні та прикладні аспекти” (Київ, 2013); XIII Міжнародна конференція “Геоінформатика: теоретичні та прикладні аспекти” (Київ, 2014); IV Всеукраїнська конференція “Аерокосмічні спостереження в інтересах сталого розвитку та безпеки GEO-UA” (Київ, 2014); VII Міжнародна науково-технічна конференція студентів та аспірантів “Перспективи розвитку інформаційно-телекомунікаційних технологій та систем” (Київ, 2015); XIV Міжнародна конференція “Геоінформатика: теоретичні та прикладні аспекти” (Київ, 2015); 16 Українська конференція з космічних досліджень (Одеса, 2016); науково-учбові семінари Наукового Центру аерокосмічних досліджень Землі ІГН НАН України (Київ, 2013-2015).

Публікації. Основні результати досліджень опубліковані у 15-ти друкованих роботах, серед яких 7 наукових статей у фахових виданнях та 8 публікацій в збірниках і працях конференцій [6–13, 112].

Структура і обсяг дисертації. Дисертація складається зі вступу, чотирьох розділів, загальних висновків, списку використаної літератури та додатків. Робота містить 153 сторінки, із них 123 сторінки основного тексту, в тому числі 36 рисунків та 12 таблиць. Список використаних джерел літератури нараховує 112 найменувань на 13 сторінках та 4 додатки на 17 сторінках.

У першому розділі обгрунтовується необхідність використання процедури класифікування даних, які отримуються за допомогою методів ДЗЗ. Методи класифікування використовуються для вирішення природоресурсних задач, моніторингу навколишнього середовища, для розв’язання численних задач з геології, лісогосподарства, сільського господарства, при здійсненні пошуку корисних копалин, проведенні оцінки стану грунтів, тощо. Також у

даному розділі розглянуто такі відомі методи контрольованого класифікування: метод мінімальної відстані, метод дистанції Махаланобіса, метод максимальної правдоподібності та метод неконтрольованого класифікування ISODATA. Були проаналізовані основні переваги та недоліки зазначених методів.

Основним недоліком розглянутих методів є те, що їх не можна застосовувати при наявності забруднених навчальних вибірок, що, в свою чергу, негативно впливає на точність класифікування. Тому ставиться задача розробити такі методи класифікування, які можуть ефективно працювати в умовах відсутності “чистих” навчальних вибірок та недостатньої апріорної інформації.

У другому розділі проведено огляд відомих підходів, де використовується математичний апарат ТСДШ, яка є певним узагальненням класичної теорії ймовірностей. Математичний апарат ТСДШ дозволяє коректно оперувати не тільки з окремими подіями, але й зі сполученнями подій. Метод Демпстера-Шейфера пристосований для використання в умовах неповної або неточної інформації, в умовах відсутності частини вхідних даних, при комбінуванні даних, отриманих від різних джерел. Також у цьому розділі розроблений і приведений новий метод контрольованого класифікування ГКЗ, який працює при наявності забруднених навчальних вибірок.

Запропонований метод відрізняється від відомих розв’язків однойменних задач введенням процедури оцінювання класифікаційної цінності спектральних каналів за допомогою спеціальної емпіричної функції і особливим способом розбиття спектрального ознакового простору, що, у сукупності, дозволяє застосовувати лише найбільш інформативні спектральні канали і, таким чином, суттєво скоротити розмірність ознакового простору при одночасному підвищенні точності класифікування.

У третьому розділі запропоновано метод класифікування ГКЗ в умовах забруднених і малих за обсягом навчальних вибірок. Даний розділ присвячений розробці та обгрунтуванню нового методу контрольованого класифікування

космічних зображень в умовах, коли навчальна вибірка є забрудненою, в тому числі і значною мірою. Запропонований метод враховує невизначеності, що можуть виникати при навчанні та класифікуванні, що досягається використанням математичного апарату та прийомів ТСДШ.

У четвертому розділі описується програмна реалізація запропонованих методів класифікування ГКЗ в умовах навчальних вибірок обмеженої якості, проводиться перевірка розроблених методів класифікування на реальних гіперспектральних зображеннях, отриманих апаратурою Hyperion супутника EO-1, та наводяться результати оцінювання точності класифікування, яка досягається в реальних умовах. При цьому результати класифікування запропонованих методів порівнюються із результатами класифікування інших відомих методів. Оцінка точності класифікування здійснюється за допомогою таких загально прийнятих показників точності, як повна точність класифікування, точність користувача, точність виробника, помилки пропуску класу, помилки ототожнення з класом, які отримуються із матриць помилок. Результати експериментальних досліджень свідчать про досить високу точність запропонованих методів класифікування.

РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ. ПОСТАНОВКА ЗАДАЧІ НА ДОСЛІДЖЕННЯ

1.1. Місце і роль процедури класифікування космічних зображень при розв'язанні тематичних задач дистанційного зондування Землі

У загальному вигляді ДЗЗ визначається як метод отримування даних (зображень) про Землю з космосу, використовуючи властивості електромагнітних хвиль, випромінюваних, відбиваних, поглинених чи розсіюваних об'єктами зондування [3–14]. При формуванні космічних зображень завжди мають місце різного роду похибки та спотворення, що обумовлено недосконалістю знімальних систем та зовнішніми чинниками. Тому перед тим, як проводити аналіз зображень, до них застосовують різноманітні процедури попереднього оброблення. Звичайно здійснюють радіометричне калібрування зображення, коригування геометричних спотворень та геопросторове прив'язування, усунення деяких видів систематичних похибок та ін. Все це достатнє складні процедури, але на сьогодні більшість з них повністю формалізовані (алгоритмізовані), тому за наявністю адекватної моделі сенсора, даних про стан атмосфери на момент знімання, необхідних наземних даних застосування цих процедур дозволяє суттєво підвищити якість зображень.

Аналіз космічного зображення передбачає пошук та визначення різних ознак (геометричних, яскравісних, спектральних та ін.) і за їх допомогою виявлення та класифікування об'єктів, які знаходяться в межах відзнятої сцени, по різних категоріях.

Найбільш складною частиною аналізу космічного зображення є його класифікування [13, 14], оскільки цей процес практично завжди має неформальну компоненту, а його реалізація потребує значних обчислювальних

витрат. Особливо випукло це проявляється в задачах класифікування ГКЗ, бо, як відомо, гіперспектральні зображення є найбільш змістовними і вміщують в собі надзвичайно великий обсяг інформації про об'єкти зйомки та спостереження.

Необхідно зазначити, що процедура класифікування ГКЗ є завершальною у ланцюгу всіх маніпуляцій із зображенням і саме вона визначає точність розв’язання тематичної задачі ДЗЗ, що вирішується. Тому кожного разу особлива увага приділяється вибору найбільш адекватного метода класифікування ГКЗ.

В силу вказаних вище чинників більшість методів і алгоритмів, які використовуються при класифікуванні багатоспектральних космічних зображень (наприклад, Landsat, MODIS, RapidEye, Sentinel), не зберігають свою високу ефективність у випадку гіперспектральних зображень. Тому останні роки розробляються методи і алгоритми, які визначально орієнтуються на класифікування саме ГКЗ.

Проведемо стислий огляд відомих методів і алгоритмів, які використовуються при аналізі та класифікуванні ГКЗ.

1.2. Огляд методів і алгоритмів класифікування гіперспектральних космічних зображень

Класифікування – це процес оброблення сигналів зображення, результатом якого є категорування (віднесення) кожного з об’єктів, що знаходяться в межах відзнятої сцени, до відповідного класу [15]. Гіперспектральні знімки поки мають доволі невисоку просторову розрізненність, тому звичайно у якості елементарного об’єкта розглядають кожний окремий піксел. При цьому виходять з того, що кожному пікселу гіперспектрального зображення відповідає набір значень спектральних ознак, або вектор в спектральному просторі, розмірність якого дорівнює повному

числу зональних зображень у складі гіперспектрального знімка. Тоді процес класифікування зводиться до розподілу за певним методом (алгоритмом) всіх пікселів по класах відповідно до відбивної здатністі (значенням спектральної яскравості) кожного об'єкта в одній або декількох зонах електромагнітного спектра.

Розрізняють два типи методів (алгоритмів) для здійснення такого розподілу – контрольованого і неконтрольованого класифікування [16, 17].

Будь-який метод контрольованого класифікування передбачає наявність навчальних вибірок, що дозволяє визначити еталонні спектральні ознаки кожного з наявних класів і сформувати правило (алгоритм) переходу від показників спектральної яскравості до класів об'єктів. Такий метод називают ще класифікуванням з навчанням [16, 17].

На той випадок, коли немає навчальних вибірок і тому не можна одержати еталонні спектральні ознаки кожного з наявних класів, існують методи і алгоритми неконтрольованого класифікування. Суть неконтрольованого класифікування полягає у тому, що воне дозволяє об’єднати піксели у групи за формальною ознакою, без врахування їх суттєвого змісту [16, 17]. При неконтрольованому класифікуванні відбувається розділення всіх пікселів зображення на групи (кластери), назва та спектральні характеристики яких заздалегідь невідомі.

Критерієм віднесення пікселів до того чи іншого кластеру є подібність спектральних характеристик. Вибір міри подібності лежить в основі правила віднесення елементів до області, що характеризується центром деякого кластера. Властивості кластера можна конкретизувати за допомогою трьох таких понять:

1) відстань між точками у просторі ознак;

2) відстань між групами точок (ймовірними кластерами);

3) критерій кластеризації [16, 17].

Найбільш розповсюдженою мірою відстані між точками є евклідова метрика

$$d_e(X_i, X_j) = \sqrt{\sum_{k=1}^{p}(x_{ki} - x_{kj})^2}, \qquad (1.1)$$

де $X_i = (x_{1i},...,x_{pi}), X_j = (x_{1j},...,x_{pj})$.

Відстань між групами точок може бути обчислена різними способами. Припустимо, необхідно знайти відстань між групами, що мають мітки *A* і *B* (групи *A* і *B* можуть бути кандидатами в кластери). Для визначення відстані між *A* і *B* потрібно визначити середню відстань між всіма парами точок, для яких один елемент кожної пари належить групі *A*, а другий - групі *B*.

Критерій кластеризації відображає віднесення точок до кластерів таким чином, щоб відстані між точками всередині кластера були мінімальними, а відстані між кластерами-максимальними.

Одним із поширених критеріїв кластеризації є критерій суми квадратів похибок – $\sum_{кп}$. Припустимо, ми маємо певну кількість кластерів, M_i – вектор математичного сподівання для *i*-го кластера, X_i – набір точок даних, що належать *i*-му кластеру. Тоді сума квадратів похибок обчислюється за наступною формулою:

$$\sum_{кп} = \sum_{i=1}^{m} \sum_{X \in C_i} \left\| X - M_i \right\|^2, \qquad (1.2)$$

де $\left\| X - M_i \right\|$ – евклідова відстань між $\boldsymbol{X}$ та $\boldsymbol{M}_i$.

Інший метод визначення критерія кластеризації пов’язується з визначенням величини розсіювання точок всередині кластера S_ω і визначення величини розсіювання кластерів S_b (міжкластерне розсіювання). Мета - мінімізувати S_ω при одночасній максимізації S_b.

Неконтрольоване класифікування є значно менш достовірним і точним у порівнянні із методом контрольованого класифікування, але його перевага – мінімум потрібних навчальних даних для виділення кластерів. Тому цей тип

класифікування знаходить застосування при обробленні гіперспектральних зображень сцен з рослиністю, лісами, болотними угіддями, відкритими грунтами, де малі спектральні контрасти і важко виділити однорідні ділянки для навчання класифікатора.

На сьогодні популярним методом неконтрольованого класифікування багатоспектральних і (частково) гіперспектральних аерокосмічних зображень є ISODATA (Iterative Self-Organizing Data Analysis Techniques), запропонований в [16, 17].

Зазначимо, що хоча контрольоване класифікування вимагає наявності навчальних вибірок і потребує значних обчислювальних ресурсів, саме цей тип класифікування звичайно застосовують при аналізі гіперспектральних зображень, оскільки він має суттєві переваги перед методом неконтрольованого класифікування по точності своїх результатів.

При класифікуванні гіперспектральних зображень найбільш часто використовують такі методи: паралелепіпедів, мінімальної відстані, дистанції Махаланобіса та максимальної правдоподібності, метод класифікування за спектральним кутом (Spectral Angle Mapper Classification - SAM), а також метод опорних векторів SVM.

Метод паралелепіпедів застосовується, коли області значень спектральної яскравості об’єктів не перетинаються. Алгоритм паралелепіпедів [16, 17] заснований на статистичних показниках навчальної вибірки для K спектральних діапазонів. Спочатку для кожного класу l та діапазону k обчислюється середнє значення яскравості в навчальній виборці x_{lk} та l_{lk}. Після цього для класифікування пікселів застосовують наступне правило. Піксел належить класу l тоді і тільки тоді, коли його яскравість BV_{ijk} задовольняє наступній умові:

$$\mu_{lk} - 2\delta_{lk} \le BV_{ijk} \le \mu_{lk} + 2\delta_{lk}, \qquad (1.3)$$

де $l = 1,2,3,\ldots,L$ – клас; $k = 1,2,3,\ldots,K$ – спектральний діапазон.

Якщо позначити нижню та верхню границі цієї нерівності як

$$L_{lk} = \mu_{lk} - 2\delta_{lk};$$

$$H_{lk} = \mu_{lk} + 2\delta_{lk},$$

то умову (1.3) можна записати у вигляді:

$$L_{lk} \le BV_{ijk} \le H_{lk}. \qquad (1.4)$$

Множина точок, які задовольняють умові (1.4), утворюють паралелепіпед в просторі спектральних ознак. Якщо значення спектральної яскравості піксела знаходяться всередині цього паралелепіпеда, то цей піксел буде віднесений до даного класу.

Перевага методу паралелепіпедів полягає в тому, що даний метод може працювати на базі простого та дуже швидкого алгоритму. Недоліком методу паралелепіпедів є те, що, область розподілу значень спектральних ознак кожного із класів має чіткі границі. Із-за цього на знімку може залишитися значна кількість некласифікованих пікселів, що не будуть віднесені до жодної області.

Метод мінімальної відстані [16, 17] використовується, коли спектральні ознаки різних класів схожі та діапазони значень їх яскравості перекриваються. Піксел відноситься до того еталонного класу, евклідова відстань до центра якого в просторі ознак мінімальна. При цьому непрокласифіковані піксели відсутні, оскільки в просторі ознак кожен піксел розташований ближче до середніх значень ознак одного з класів.

Правило спектральної відстані [16, 17] базується на обчисленні спектральної відстані між вектором значень піксела-кандидата та вектором середніх значень кожної навчальної вибірки за наступною формулою:

$$SD_{xyl} = \sqrt{\sum_{i=1}^{n} \left(\mu_{lk} - X_{xyl}\right)^2}, \qquad (1.5)$$

де K – число спектральних каналів (розмірність);

k – поточний спектральний канал;

l – поточний клас;

X_{xyk} – вихідне значення яскравості піксела з координатами x, y у каналі k;

μ_{lk} – середнє значення яскравості піксела в спектральному каналі k для зразка, що відповідає класу l;

SD_{xyl} – спектральна відстань від піксела з координатами x, y до точки, яка задає піксел-кандидат.

Згідно з правилом мінімальної відстані розраховуються спектральні відстані до всіх можливих значень l (відстань до всіх можливих класів). Потім кандидат-піксел відноситься до того класу, для якого відстань SD мінімальна.

Перевагою методу мінімальної відстані є те, що він є самим швидким після методу паралелепіпедів. При цьому не залишається некласифікованих пікселів, оскільки будь-який піксел є спектрально близьким до того чи іншого середнього значення. Недоліком даного методу є те, що для групи пікселів з невеликою варіацією яскравості виконується багато зайвих операцій, оскільки ці піксели є досить близькими до свого середнього значення.

Метод дистанції Махаланобіса [16, 17] відрізняється від методу мінімальної відстані тим, що під час класифікування вимірюється не евклідова відстань, а дистанція, яка дозволяє врахувати розподіл (дисперсію) значень яскравості пікселів на еталонних ділянках. Але обчислення за допомогою цього методу потребують значних часових затрат.

У методі максимальної правдоподібності [16, 17] розраховується ймовірність, з якою даний піксел належить до певного класу. Кількість та параметри класів задаються користувачем шляхом визначення навчальних вибірок. Кожен піксел відноситься до того класу, до якого він може належати з найбільшою ймовірністю. Під час обчислення ймовірності враховується спектральна яскравість піксела та яскравість тих пікселів, що знаходяться

навколо нього. Цей метод має високу точність, враховує дисперсію значень ознак класів і не залишає некласифікованих пікселів.

В методі максимальної правдоподібності дані із навчальної вибірки використовуються для обчислення середнього вектору вимірювань M_L та коваріаційної матриці V_L для кожного класу l та спектрального діапазону k.

Критерій того, що піксел належить до певного класу, формулюється наступним чином. Піксел x належить до класу l тоді і тільки тоді, якщо

$$p_l > p_i, \quad i = 1,...,L,$$

де L – число класів;

p_i – ймовірність того, що даний клас існує.

Ймовірність p_l визначається за наступною формулою:

$$p_l = \left[-0{,}5\log\{\det(V_l)\}\right] - 0{,}5\left((X - M_l)^T (V_l)^{-1} (X - M_l)\right). \qquad (1.6)$$

Недоліком даного методу є те, що для розрахунків треба задіювати великий об’єм машинної пам’яті та багато часу, при цьому результати класифікування можуть виявитися не досить точними.

Метод класифікування за спектральним кутом SAM [105] використовується для аналізу набору виміряних значень спектральних ознак (вектор в спектральному просторі) об’єкта з точки зору відповідності еталонним спектральним ознакам різних класів.

Під спектральним кутом θ розуміють значення, яке характеризує взаємне розташування вектора ознак x, що належить до контрольної вибірки, та вектора, що описує клас l:

$$\theta_l(x) = \cos^{-1}\left(\frac{x \cdot x^l}{\|x\| \cdot \|x^l\|}\right), \qquad (1.7)$$

де x^l – вектор емпіричних середніх класу l, побудований за навчальною вибіркою.

Вирішальне правило даного класифікатора полягає в мінімізації значення спектрального кута $\theta_l(x)$. При цьому значення спектрального кута показує, наскільки напрям вектора, що класифікується, близький до напряму вектора, що описує клас l.

В основі методу опорних векторів SVM лежить побудова оптимальної розділяючої гіперплощини між двома класами. Під оптимальною гіперплощиною розуміється така гіперплощина, відстань до якої від самих близьких точок двох класів максимальна.

Для даного методу кожен вектор ознак x характеризується парою (y, c), де $y-$нормований вектор ознак, а c приймає значення -1 або +1 в залежності від того, до якого класу належить вектор ознак.

Вираз для розділяючої площини має вигляд:

$$w \cdot y - b = 0, \tag{1.8}$$

де $w-$перпендикуляр до розділяючої площини, $b-$відстань від гіперплощини до початку координат. Тоді задача побудови оптимальної гіперплощини зводиться до задачі квадратичної оптимізації:

$$\begin{cases} \|w\| \to \min, \\ c(w \cdot y - b) \geq 1. \end{cases} \tag{1.9}$$

Недоліком методу SVM є те, що проведення процедури оптимізації потребує певних обчислювальних витрат. Разом з тим необхідно зазначити, що саме метод SVM найчастіше забезпечує найкращі за точністю результати при класифікуванні космічних зображень, як багатоспектральних, так і гіперспектральних [17, 18].

Як було зазначено вище, процедура класифікування ГКЗ є завершальною у ланцюгу всіх маніпуляцій із зображенням і саме вона визначає точність розв’язку тематичної задачі ДЗЗ. В літературі описано достатньо багато показників, які виражають точність результатів класифікування, але всі вони розраховуються на базі елементів матриці помилок (Error Matrix) [19–28].

Структура матриці помилок зображена на рис. 1.1. Дана матриця заповнюється статистичними результатами проведеного класифікування вибірки із n об'єктів при умові, що число класів в ній дорівнює L. Кожен рядок нумерується індексом i, а кожен стовпчик нумерується індексом j. При цьому $i, j = 1,2,...,L$. Елемент матриці n_{ij} позначає число об'єктів, що віднесені при класифікуванні до класу i, але в дійсності належать класу j. По головній діагоналі матриці стоять елементи, що позначають кількість об'єктів, що були правильно класифіковані.

		Завіркові дані (Reference data)					
		Class 1	Class 2	Class j	Class L	Сума елементів рядка (Row total) n_{i+}	**Точність користу-вача (User's Accuracy, UA)**
Класифіковані дані (Classified data)	Class 1	n_{11}	n_{12}	n_{1j}	n_{1L}	n_{1+}	UA_1
	Class 2	n_{21}	n_{22}	n_{2j}	n_{2L}	n_{2+}	UA_2
	Class i	n_{i1}	n_{i2}	n_{ij}	n_{iL}	n_{i+}	UA_i
	Class L	n_{L1}	n_{L2}	n_{Lj}	n_{LL}	n_{L+}	UA_L
	Сума елементів стовпчика (Column total) n_{+j}	n_{+1}	n_{+2}	n_{+j}	n_{+L}	n	
	Точність виробника (Producer's Accuracy, PA)	PA_1	PA_2	PA_j	PA_L		

Рис. 1.1. Структура матриці помилок

В матриці помилок використані наступні позначення:

$$n_{i+} = \sum_{j=1}^{K} n_{ij};$$ (1.10)

$$n_{+j} = \sum_{i=1}^{K} n_{ij}.$$ (1.11)

На основі даних, які містяться в матриці помилок, можна будувати різні показники ефективності процедури (методу) класифікування; серед них такі [29–32]:

$$OA = \frac{\sum_{i=1}^{K} n_{ii}}{n} -$$ загальна точність класифікування (Overall Accuracy); (1.12)

$$PA = \frac{n_{jj}}{n_{+j}} -$$ точність виробника (Producer’s Accuracy); (1.13)

$$UA = \frac{n_{ii}}{n_{i+}} -$$ точність користувача (User’s Accuracy). (1.14)

Загальна точність є оцінкою якості класифікування усього зображення в цілому, тобто оцінкою, усередненою по всім наявним класам.

Точність виробника та точність користувача обчислюються для кожного класу окремо. Результати обчислень записуються відповідно у нижній рядок та правий стовпчик матриці помилок.

Доволі часто кількісні представлення об’єктів різних класів у виборці сильно відрізняються. При цьому такі критеріальні показники, як загальна точність класифікування, точність виробника та точність користувача не завжди дають адекватне представлення про ефективність процедури класифікування, що розглядається. Але зазначений недолік можна виправити за допомогою капа-індексу (Kappa Index) [29–32], який обчислюється наступним чином:

$$KI = \frac{n\sum_{i=1}^{K} n_{ii} - \sum_{i=1}^{K} n_{i+} n_{+i}}{n^2 - \sum_{i=1}^{K} n_{i+} n_{+i}}.$$ (1.15)

Капа-індекс *KI* приймає значення від 0 до 1: чим ближче його значення до 1, тим більш точними є результати класифікування.

Матриця помилок дає можливість визначати ще два показника - помилку пропуску класу (Error of Omission) і помилку ототожнення з класом (Error of Commission).

Помилки пропуску класу (Error of Omission) вказують на невірну класифікацію, коли піксели, які в дійсності мають належати до певного конкретного класу, не були віднесені до цього класу. Помилка Error of Omission для певного конкретного класу рахується наступним чином: сума пікселів у стовпчику, що відповідає даному класу без елемента, що стоїть по головній діагоналі, ділиться на число завіркових пікселів.

Зазначимо, що помилка Error of Omission і показник точності виробника (Producer’s Accuracy) - обернено пропорційні величини. Висока точність виробника означає, що під час класифікування було мало помилок пропуску класу.

Помилки ототожнення з класом (Error of Commission) вказують на невірну класифікацію, коли піксел з одного класу був помилково віднесений до іншого класу, хоча в дійсності він до цього класу не належить. Помилка Error of Commission розраховується наступним чином: сума пікселів у рядку без елемента, що стоїть по головній діагоналі, ділиться на число завіркових пікселів.

Помилка Error of Commission і показник точності користувача (User’s Accuracy) - обернено пропорційні величини. Низька точність користувача означає, що при проведенні класифікування було допущено багато помилок щодо ототожнення.

1.3. Постановка завдання на дослідження

Як обговорювалось в п. 1.2, при вирішенні тематичних задач ДЗЗ на основі ГКЗ найбільш складною процедурою є класифікування, при цьому перевагу віддають контрольованому класифікуванню, яке потребує навчальних вибірок.

Щоб при класифікуванні забезпечити сталі та статистично достовірні результати, навчальна вибірка має відповідати певним вимогам, серед яких принциповими є репрезентативність, достатність та чистота [33–35].

Вимога репрезентативності декларує, що в навчальній виборці мають бути присутні, у приблизно рівних відсотках, представники всіх класів, що розглядаються.

Вимога достатності каже про той мінімальний об’єм навчальної вибірки N_{TS}^{sut}, який забезпечує статистичну коректність оцінок точності класифікування. Якщо позначити через N_{TS}. кількість елементів-репрезентів у навчальній виборці, то при формуванні навчальної вибірки має бути забезпечена умова

$$N_{TS} \geq N_{TS}^{sut}. \tag{1.16}$$

Існують різні способи розрахунку величини N_{TS}^{sut}, але найбільш поширеною є біноміальна модель [4, 35], за якою мінімально необхідний об'єм навчальної вибірки розраховується за формулою

$$N_{TS}^{sut} = \frac{Z_{\alpha}^{2} P_0 (1 - P_0)}{b^2}, \tag{1.17}$$

де Z_{α} – стандартна табульована оцінка для прийнятого значення ймовірності α двохсторонньої помилки 1-го роду; P_0 – оцінка ймовірності правильного класифікування; b – припустима похибка.

Величина P_0 визначається використовуваним алгоритмом класифікування. Рекомендації по вибору величин P_0, помилки α і допустимої

похибки b наводяться, зокрема, в [33–35]. При обраній величині помилки α процентиль Z_α знаходиться табличним шляхом.

Розрахунки по формулі (1.17) для типових даних показують, що для того, щоб забезпечити статистично надійні результати, навчальна вибірка має налічувати до кількох сотен елементів. На жаль, подібну велику навчальну вибірку на практиці вдається сформувати далеко не завжди. Якщо об’єм навчальної вибірки не задовольняє умові (1.16), тоді кажуть про “малі вибірки” [29–35].

Рівень чистоти навчальної вибірки розраховується як відношення кількості пікселів, справжня класова приналежність яких співпадає з тією, що їм була присвоєна при формуванні навчальної вибірки, до всього обсягу навчальної вибірки. Характеристикою, протилежною чистоті навчальної вибірки, є її забрудненість. Показник забрудненості вимірюється для вибірки часткою (відсотком) пікселів, які входять до її складу, але реально не належать до класу, з яким ці піксели ототожнюються.

Зазначимо, що виділити в природному середовищі повністю чисті навчальні вибірки проблематично; серед репрезентів класів зазвичай присутні сторонні елементи. Наприклад, навчальна підвибірка "рослинність" у вигляді майданчика з трав'яним покриттям може містити ділянки з оголеним грунтом, доріжки, штучні об'єкти і т. д. У подібних випадках говорять про забруднені вибірки [29–35].

Іноді проблема визначення класу ускладнюється тим, що деякі піксели є “змішаними”, тобто їх сигнал є сумішшю сигналів об’єктів двох або більше класів. У вибірку можуть потрапляти ще так звані “збійні” піксели.

Також слід мати на увазі, що навчальна вибірка формується експертом на основі візуального аналізу зображення, з урахуванням його знань та досвіду, а також із залученням довідкових матеріалів. Безумовно, при цьому деякі рішення експерта щодо класової приналежності окремих пікселів можуть бути помилковими.

Таким чином, при контрольованому класифікуванні ГКЗ частіше усього приходиться мати справу з навчальними вибірками, які лише частково задовольняють теоретично необхідним вимогам. Безумовно, це впливає на точність та сталість результатів класифікування.

Так, в [33–36] показано залежність точності результатів класифікування від об’єму навчальної вибірки. В ряді робіт [37–40], де вивчался вплив забрудненості навчальної вибірки на точність класифікування, було показано, що навіть незначні вкраплення сторонніх елементів в навчальну вибірку можуть призводити до серйозних помилок в оцінці параметрів і ознак.

З наведеного слідує, що потрібні нові підходи і методи до класифікування, які будуть більш ефективно, ніж відомі, працювати в умовах навчальних вибірок обмеженої якості. Тому у дисертації ставиться завдання розробити та обґрунтувати нові ефективні методи контрольованого класифікування ГКЗ в умовах, коли навчальна вибірка є забрудненою, в тому числі сильно, а її об’єм є дуже обмеженим [17–37].

У якості критерієв ефективності пропонованих нами методів контрольованого класифікування будемо використовувати загально визнані показники, які досить просто розраховуються виходячи з елементів матриці помилок (п. 1.2).

Математичною платформою для розробки нових методів контрольованого класифікування нами обрано апарат та положення ТСДШ [37–41]. Вибір пов'язаний з тим, що математичний апарат ТСДШ дає можливість в умовах забрудненості та неповноти даних працювати зі складними багатоальтернативними гіпотезами, пропонує просте правило для комбінування даних від різних джерел, дозволяє будувати інтервальні оцінки достовірності гіпотез з урахуванням існуючих ризиків, а також має деякі інші переваги перед традиційним теоретико-ймовірнісним підходом [42–48].

Для вирішення поставленого в дисертації наукового завдання формулюються та досліджуються такі часткові задачі:

- провести аналіз сучасних методів класифікування ГКЗ, їх переваг та недоліків;
- розробити метод контрольованого класифікування ГКЗ в умовах забруднених навчальних вибірок;
- розробити метод контрольованого класифікування ГКЗ в умовах забруднених навчальних вибірок та кількісних обмежень;
- провести експериментальні дослідження для оцінки ефективності запропонованих методів класифікування на реальних ГКЗ.

ВИСНОВКИ ДО РОЗДІЛУ 1

1. У розділі обговорюються місце і роль процедури класифікування космічних зображень при розв’язанні тематичних задач дистанційного зондування Землі. Проведено огляд сучасних методів і алгоритмів класифікування гіперспектральних космічних зображень, аналізуються притаманні ним особливості та недоліки.

2. Показано, що сучасні алгоритми контрольованого класифікування гіперспектральних космічних зображень досить чутливі до характеристик навчальних вибірок. Більш конкретно, точність відомих алгоритмів суттєво залежить від об’ємів навчальних вибірок та рівня їх забрудненості. Тому відзначається, що потрібні нові підходи до класифікування, які будуть більш ефективно, ніж відомі, працювати в умовах навчальних вибірок обмеженої якості.

3. Сформульовано завдання розробити та обґрунтувати нові ефективні методи контрольованого класифікування гіперспектральних космічних зображень в умовах, коли навчальна вибірка є забрудненою, в тому числі сильно, а її об’єм є дуже обмеженим. Наведено критерії, за якими передбачається оцінювати ефективність пропонованих автором нових методів контрольованого класифікування.

4. Математичною платформою для розробці нових методів контрольованого класифікування обрано апарат та положення теорії свідчень Демпстера-Шейфера. Вибір пов'язаний з тим, що математичний апарат ТСДШ дає можливість в умовах забрудненості та неповноти даних працювати зі складними багатоальтернативними гіпотезами, пропонує просте правило для комбінування даних від різних джерел (у тому числі зображень), дозволяє будувати інтервальні оцінки достовірності гіпотез з урахуванням існуючих ризиків, а також має деякі інші переваги перед традиційним теоретико-ймовірнісним підходом.

5. Сформульовано часткові задачі, вирішення яких потребує поставлене в дисертації наукове завдання.

РОЗДІЛ 2 РОЗРОБКА МЕТОДУ КОНТРОЛЬОВАНОГО КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ В УМОВАХ ЗАБРУДНЕНИХ НАВЧАЛЬНИХ ВИБІРОК

У розділі розглянуто новий метод класифікування ГКЗ, який базується на положеннях ТСДШ. Метод відрізняється від відомих розв'язків однойменних задач введенням процедури оцінювання класифікаційної цінності спектральних каналів за допомогою спеціальної емпіричної функції і особливим способом розбиття спектрального ознакового простору, що, у сукупності, дозволяє застосовувати лише найбільш інформативні спектральні канали і, таким чином, суттєво скоротити вимірність ознакового простору при одночасному підвищенні точності класифікування. Метод реалізован програмно. Результати експериментальних досліджень точності класифікування методу буде наведено далі у розділі 4.

2.1. Основні положення теорії свідчень Демпстера-Шейфера

Підхід ТСДШ є певним узагальненням класичного ймовірнісного підходу [49–65], оскільки математичний апарат ТСДШ дозволяє коректно працювати (оперувати) не тільки з окремими подіями (сінглтонами) або гіпотезами, але й зі сполученнями подій (гіпотез), причому кожне таке сполучення розглядається за методологією ТСДШ як окрема нерозкладна подія (гіпотеза) *A*. У ТСДШ ключовим поняттям є поняття "маси" (mass), яке є узагальненням класичного поняття ймовірності. На відміну від байєсівської ймовірності, масою можна визначити ситуацію при неповноті даних та відокремити поняття відсутності довіри від недовіри.

Сукупність вихідних (базових) гіпотез відносно стану об'єкта та всі можливі їх сполучення утворюють множину Ω, яка називається у ТСДШ "основою аналізу" (Frame of Discernment). Вважається, що всі гіпотези з

множини Ω (й базові, й їх сполучення) є взаємно незалежними та у сукупності повністю характеризують об'єкт. Якщо число базових гіпотез рівно Q, то загальна кількість підмножин у множині Ω складає величину 2^Q (сюди входять пуста множина Ø та сама множина Ω).

Кожному елементу множини Ω надається, за певними правилами, відповідна маса *m*, причому область значень маси лежить в інтервалі від нуля до одиниці. Формально це записується таким чином: $m: 2^Q \to [0,1]$, при цьому вважається, що:

- маса пустої множини дорівнює нулю: $m(\varnothing) = 0;$
- сума усіх мас для кожної підмножини $A \subseteq \Omega$ дорівнює одиниці: $\sum_{A \subseteq \Omega} m(A) = 1.$

На змістовному рівні масу *m* можна розглядати як міру довіри до пов'язаної з нею гіпотези; тому m(A) ще називають базовою ймовірністю, а будь-яку підмножину A, для якої m(A)>0, називають фокальною множиною або елементом (Focal Set) [42–48].

Різниця між теорією Демпстера-Шейфера і теорією ймовірностей полягає у тому, що для теорії ймовірностей виконуються такі умови:

1) $P(\Omega) = 1;$

2) Якщо $X \subset Y,$ то обов'язково виконується умова: $P(X) \le P(Y);$

3) $P(X) + P(\overline{X}) = 1,$ де $\overline{X}$ – доповнення до множини X, тобто виконується: $X \cup \overline{X} = \Omega, \quad X \cap \overline{X} = \varnothing.$

Для теорії Демпстера-Шейфера виконуються такі умови:

1) $m(\Omega) \ne 1;$

2) Якщо $X \subset Y,$ то не обов'язково, щоб виконувалася умова: $m(X) \le m(Y);$

3) Не вимагається взаємозв'язок між $m(X)$ та $m(\overline{X})$.

Для представлення суб'єктивної ненадійності, яку не можна виразити за допомогою байєсівської ймовірності, Демпстером було введено такі поняття, як нижня та верхня ймовірності [18–48]. Згодом Шейфер, удосконалюючи теорію Демпстера, перейменував їх відповідно у функцію довіри та міру правдоподібності.

Функції довіри (Belief) та міри правдоподібності (Plausibility) [18–48], для будь-якої підмножини $X \subseteq \Omega$ обчислюються відповідно за виразами:

$$Bel(X) = \sum_{A_i \subseteq X} m(A_i), \tag{2.1}$$

$$Pls(X) = \sum_{A_i \cap X \neq \varnothing} m(A_i). \tag{2.2}$$

Інакше кажучи, нижня ймовірність являє суму базових ймовірностей, які замкнуті у підмножині X і не виходять з неї. Верхня ймовірність являє собою суму базових ймовірностей, що хоча б частково можуть увійти до підмножини X.

Функція довіри $Bel(X)$ визначає (сумарний) рівень впевненості, що надається підмножині $X \subseteq \Omega$. Функція правдоподібності $Pls(X)$ визначає рівень розширення, до якого підмножина X ще може вважатися правдоподібною.

Для функції довіри та міри правдоподібності виконуються наступні властивості:

1) $Bel(\Omega) = 1;$

2) $Pls(\Omega) = 1;$

3) $Bel(X) \leq Pls(X), \ \forall X \subseteq \Omega;$

4) $Pls(X) = 1 - Bel(\overline{X}); \ \forall X \subseteq \Omega;$

5) $Bel(X) + Bel(\overline{X}) \leq 1; \ \forall X \subseteq \Omega,$ де $\overline{X}$ – доповнення до множини X, тобто виконується: $X \cup \overline{X} = \Omega, \quad X \cap \overline{X} = \varnothing.$

6) $Bel(X) \leq \Pr ob(X) \leq Pls(X), \ \forall X \subseteq \Omega.$

Також значення $Bel(X)$ та $Pls(X)$ дають нижню та верхню границі інтервалу, який містить точну оцінку значення можливості підмножини X.

Таким чином, невизначеність будь-якої гіпотези A представляється значеннями на відрізку $[Bel(A), Pls(A)]$, який називається “інтервалом довіри”, а ширина інтервалу визначає похибку виміру невизначеності.

Слід зауважити, що базову масу події A можна отримати через функцію довіри як [18–48]:

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} Bel(B), \quad \forall A \subseteq \Omega, \tag{2.3}$$

де $|A-B|$ – потужність множини $(A-B)$.

Гіпотези, або свідчення можуть одночасно формуватися кількома різними джерелами. Для випадку, коли ці джерела взаємно незалежні, в ТСДШ розроблено просте правило об’єднання (комбінування) свідчень.

Припустимо, ми маємо множину - основу аналізу Ω з числом свідчень (гіпотез) Q, а також маємо P незалежних джерел ID_p ($1 \le p \le P$) цих свідчень. При цьому $p-$те джерело дає свою оцінку ймовірності (можливості) свідчення A_n у вигляді відповідної маси $m_p(A_n)$ або оцінку ймовірності (можливості) сполучення свідчень X у вигляді $m_p(X)$. Тоді об’єднання свідчень, або отримання комбінованого свідчення здійснюється за допомогою правила Демпстера [49–69]:

$$m_D(X) = \frac{1}{1-C} \sum_{\substack{X_1 \cap \ldots \cap X_R = X \\ \forall (X \neq \varnothing) \in 2^Q}} \prod_{1 \le p \le P} m_p(X_r). \tag{2.4}$$

де C – коефіцієнт конфліктності,

$$C = \sum_{\substack{X_1 \cap \ldots \cap X_R = \varnothing \\ \forall (X \neq \varnothing) \in 2^Q}} \prod_{1 \le p \le P} m_p(X_r).$$

Коефіцієнт конфліктності C вказує, наскільки протилежними між собою є джерела (точніше, їх свідчення). Область значень коефіцієнта C лежить в інтервалі [0;1], при цьому нульове значення свідчить про

відсутність протиріч в оцінках джерел, а чим сильніше ці протиріччя, тим ближче до одиниці стає величина *C*.

Ще слід враховувати, що в розрахунках за правилом Демпстера (2.4) приймається, що $m_D(\varnothing)=0$.

Розглянуті положення ТСДШ є математичною основою нових методів, контрольованого класифікування ГКЗ, що пропонуються та обґрунтовуються у дисертації.

2.2. Метод класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок

2.2.1. Модель гіперспектрального космічного зображення та припущення

Будемо записувати ГКЗ у складі *K* зональних зображень S_k як

$$HSI=\left\{S_k \mid k=1,2,...,K\right\}=\left\{\left(\pi_n,\mathbf{u_n}\right) \mid n=1,2,...,N_\pi\right\}, \qquad (2.5)$$

де S_k − k-те зональне зображення; *K* – загальна кількість зональних зображень (спектральних каналів, або зон); π_n − n-ий піксел з повним (векторним) сигналом $\mathbf{u}_n=(u_{1n},u_{2n},...u_{kn},...,u_{Kn})$; u_{kn} − величина сигналу піксела π_n у *k*-му спектральному каналі (зоні); N_π − загальна кількість пікселів в кадрі ГКЗ.

В запису (2.5) передбачається, що зональні зображення із сусідніми номерами, наприклад, S_k та S_{k+1}, належать відповідним суміжним спектральним зонам $\Delta\lambda_k=\lambda_{k_2}-\lambda_{k_1}$ і $\Delta\lambda_{k+1}=\lambda_{k_3}-\lambda_{k_2}$.

Передбачається також, що на ГКЗ представлені (одним або більшим числом пікселів) об'єкти різних класів, тобто кожен об'єкт - піксел $\boldsymbol{\pi}_n$ даного зображення належить до одного з *L* заданих класів. Вважається, що експерту-досліднику доступна інформація у вигляді вибірки об'єктів з репрезентів наявних класів.

2.2.2. Організація і зміст методу

Схема організації виконання логіко-обчислювальних процедур при класифікуванні ГКЗ в умовах забруднених навчальних вибірок представлена на рис. 2.1.

Метод включає такі головні процедури: формування бази знань (БЗ); здійснення класифікування ГКЗ; оцінювання точності класифікування; перевіряння на задоволення встановленим (заданим) вимогам.

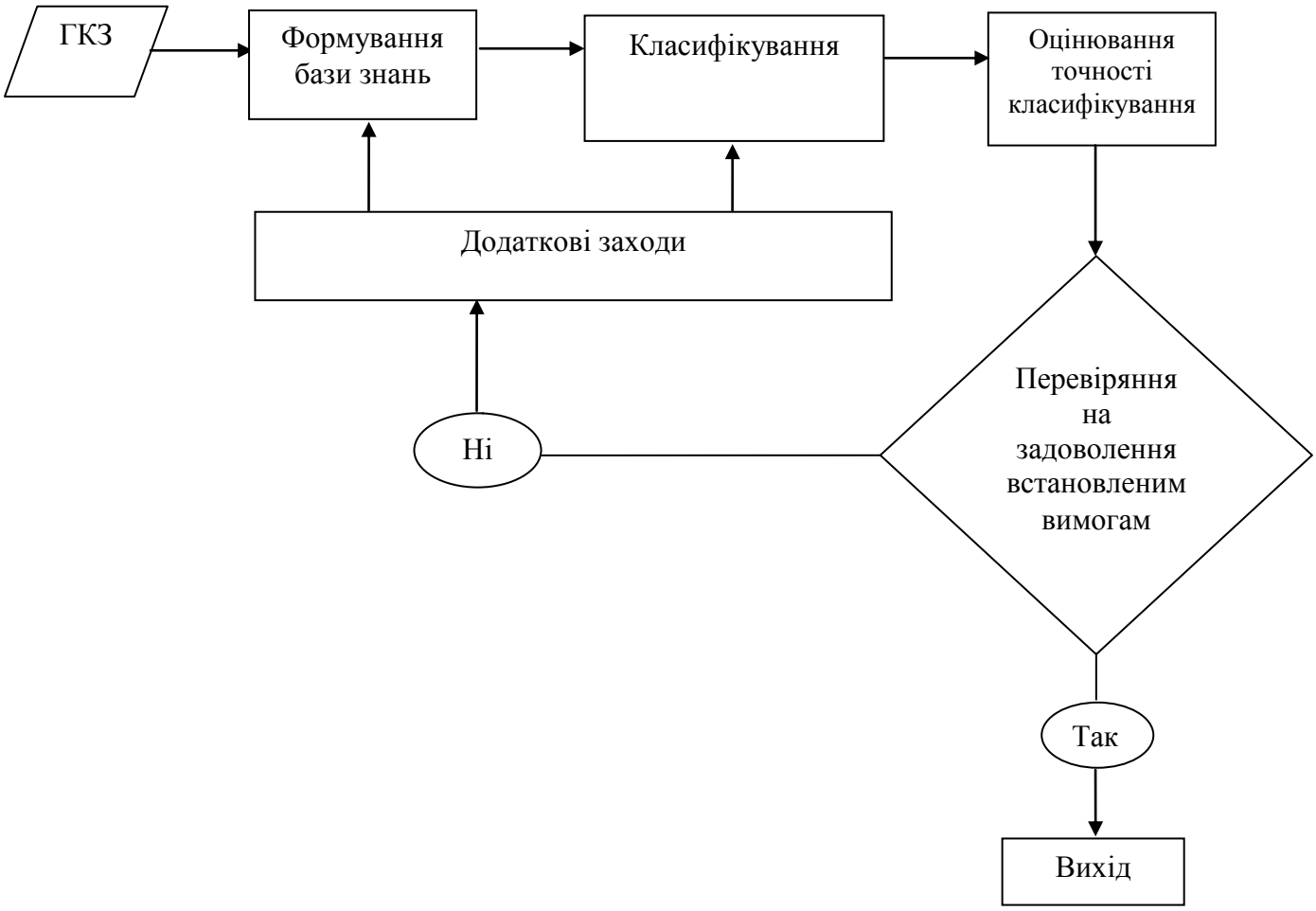


Рис. 2.1. Схема організації виконання логіко-обчислювальних процедур при класифікуванні ГКЗ в умовах забруднених навчальних вибірок

Формування БЗ здійснюється на основі навчальної вибірки, яка одержується експертом-дослідником шляхом аналізу ГКЗ і виділення на ньому пікселів-репрезентів для кожного з наявних класів.

Комплексний аналіз сигналів пікселів одержаної навчальної вибірки дозволяє створити БЗ, інформація з якої надалі використовується при класифікуванні ГКЗ. Створення БЗ потребує виконання низки операцій, зміст яких буде розглянутий нижче.

Процес власне класифікування ГКЗ здійснюється в автоматичному режимі.

За результатами класифікування, при наявності референсних (завіркових) даних, будується матриця помилок (error matrix), на основі елементів якої обчислюються оцінки досягнутої точності. Якщо ті оцінки задовольняють встановленим (заданим) вимогам, процедура класифікування вважається завершеною. В іншому випадку здійснюються додаткові заходи, які можуть сприяти підвищенню точності класифікування (збільшення навчальної вибірки, уточнення описів класів і т. д.).

Процедура формування БЗ (бази знань) відображена на рис. 2.2 і включає наступні роботи:

Відбір зональних зображень;

Виділення навчальної вибірки;

Фільтрація навчальної вибірки;

Розбиття ознакового простору;

Внутрішньоінтервальні оцінки для класів;

Визначення класифікаційної цінності спектральних каналів.

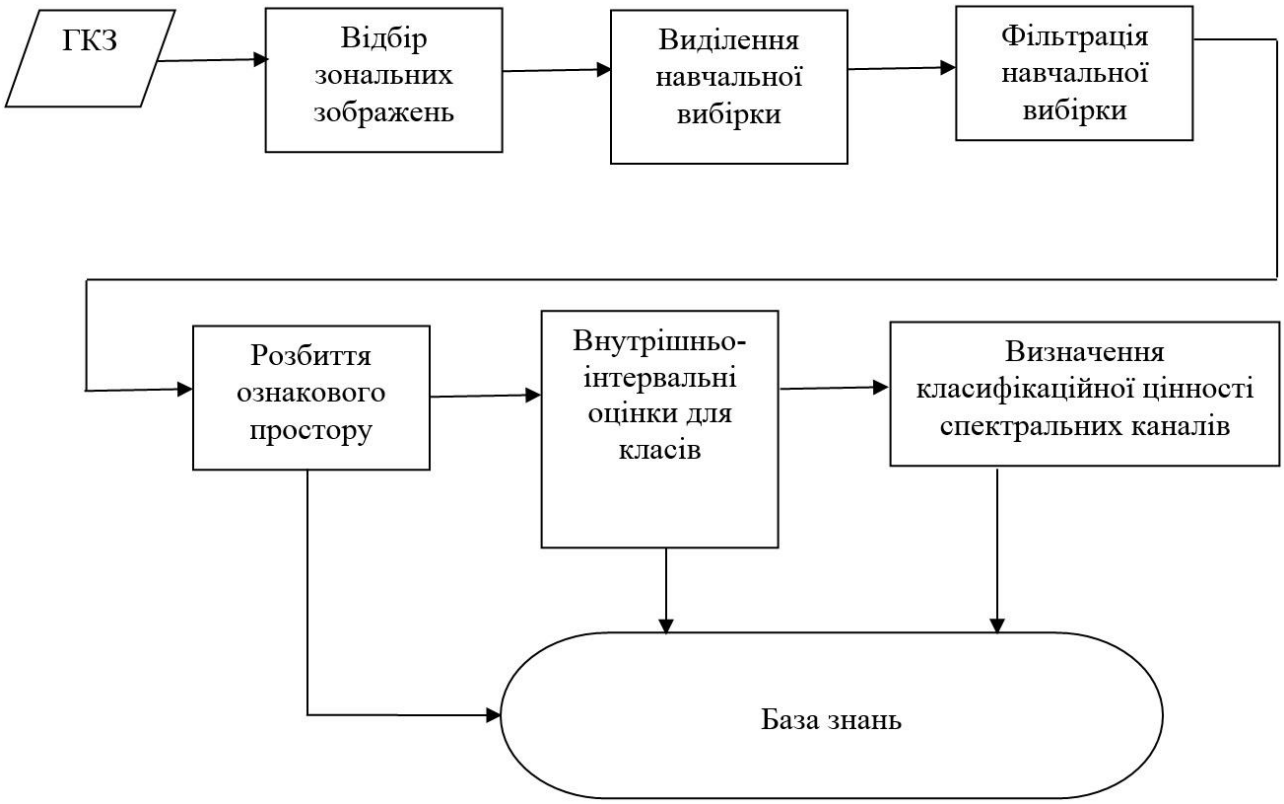


Рис. 2.2. Схема процесу створення БЗ

Розглянемо зміст та особливості виконання кожного з названих видів робіт.

2.2.2.1. Відбір зональних зображень

Практично завжди ГКЗ є інформаційно надлишковими [49–69]. Знизити надлишок даних, що містяться у ГКЗ, можна відсіканням зональних зображень з високим рівнем шумів та усуненням частини зональних зображень зі значною взаємною кореляцією. Зазначимо, що останній шлях важливий також у світлі

наведених вище вимог до взаємної незалежності джерел даних, що комбінуються за правилом Демпстера.

У пропонованому методі використовується процедура відбору зональних зображень $\{S_k\}$, метою якої є отримання деякого їх списку довжиною K^*, який є суттєво меншим за початкову кількість K у вхідному гіперспектральному зображенні (k = 1,2, …, K^*; $K^* << K$), але при цьому набір зображень, що залишились у списку, зберігає інформацію, яка необхідна для розв'язання задачі класифікування.

Процедура відбору зональних зображень $\{S_k\}$ складається з наступних кроків:

Крок 1. Зі складу ГКЗ виключаються зональні зображення, які є сильно зашумленими і мають малі середні градієнти (останнє характерно для рівномірно засвітлених полів). Після візуального контролю і відбраковування експертом формується скорочений список зональних зображень, з якими виконуються подальші маніпуляції (кроки).

Крок 2. Задаються порогове значення коефіцієнту взаємної кореляції $r_{\max}$ між сусідніми (суміжними) зональними зображеннями та (попередньо) мінімальне число каналів K_{min}, які у сукупності ще достатні для успішного розв'язання задачі класифікування. Обидва названих числа встановлюються експертом з урахуванням інформації про характер району зйомки.

Крок 3. Зональні зображення об'єднуються у пари (S_1 і S_2), (S_3 і S_4),…, (S_k і S_{k+1}) і т. д., і для кожної пари розраховується відповідний коефіцієнт взаємної кореляції r_{12},. r_{34}, …, $r_{k,k+1}$ і т. д.

Крок 4. Проводиться послідовне порівняння розрахованих коефіцієнтів взаємної кореляції із заданим пороговим значенням $r_{\max}$. У випадку, якщо величина коефіцієнту, наприклад, $r_{k,k+1}$, не перевищує порогового значення $r_{\max}$ ($r_{k,k+1} \leq r_{\max}$), у створюваний список заносяться обидва зональні зображення; в протилежному випадку до списку заноситься зональне зображення, яке має більш високу інформативність (у найпростішому випадку може бути оцінена

величиною середньоквадратичного відхилення піксельних сигналів). За результатами порівнянь формується проріджений список зональних зображень довжиною K^*.

Крок 5. Довжина K^* сформованого на попередньому кроці списку зіставляється з числом K_{min}, яке попередньо задається експертом. В залежності від результату зіставлення можливі такі варіанти:

а) якщо $K^* < K_{min}$, то задача відбору вважається невиконаною; й слід повертатися до кроку 2 із внесенням відповідних корективів у значення $r_{\max}$ та/або K_{min};

б) за умови $K^* > 2 \cdot K_{min}$ повертаємося до *кроку 3*, при цьому сформований на кроці 4 список зональних зображень розглядається як вихідний;

в) якщо $K^* \leq 2 \cdot K_{min}$, то сформований на кроці 4 список зональних зображень розглядається як остаточний для подальшого класифікування, тобто задача відбору зональних зображень вважається успішно виконаною.

Алгоритмічна реалізація розглянутої процедури відбору зональних зображень у псевдокоді приведена на рис. 2.3.

```
N: number of images S_i
r_max : threshold
K_min : minimum number of spectral bands
i = i + 2 : step size
For i = 1 to N
S is a vector of images S_i
Sort S
begin
write ("Для зонального изображения S_i считаем: a_i, cov_i, δ_i");
readln (a_i, cov_i, δ_i);
write ("Для зонального изображения S_i+1 считаем: a_i+1, cov_i+1, δ_i+1");
readln (a_i+1, cov_i+1, δ_i+1);
write ("Для зональных изображений S_i и S_i+1 считаем: r_i,i+1");
```

```
readln ( r_{i,i+1} );
if r_{i,i+1} ≤ r_max then we choose images S_i and S_{i+1}
else
begin
if δ_i > δ_{i+1} then we choose image S_i ;
else: we choose image S_{i+1}
end
write ("we get a new number of spectral bands, K*);
end.
begin
if K* < K_min then we correct r_max and/or K_min
else
begin
if K* > 2·K_min then write ("Для списка зональных изображений длины K*
считаем: r_{i,i+1}")
else: we use these K* spectral images for following classification
end
end.
```

Рис. 2.3. Алгоритм відбору зональних зображень у псевдокоді

2.2.2.2. Фільтрація навчальної вибірки

Вимога чистоти вибірки пов'язана з тим, що, як відзначалось вище, дуже рідко вдається виділити підвибірки, кожна з яких містить об’єкти тільки одного відповідного класу. Присутність сторонніх пікселів у навчальній вибірці негативно впливає на точність класифікування, тому доцільно усунення їх зі складу вибірки. Для цього застосовується спеціальна процедура фільтрації, у процесі якої кожна підвибірка TS_l (l=1, 2,…, L) перевіряється на предмет, чи всі піксели, що складають компактну групу, належать лише відповідному класу [66–69]. Таку контрольну перевірку здійснюють за допомогою одного з алгоритмів класифікування [66–69]. Якщо встановлюється, що деякий піксел з

підвибірки TS_l фактично не належить до класу l, він усувається з цієї підвибірки.

У методі, що пропонується, використана процедура фільтрації, представлена схемою на рис. 2.4; при цьому для контролю фактичної приналежності пікселів застосовується алгоритм класифікування з використанням ТСДШ.

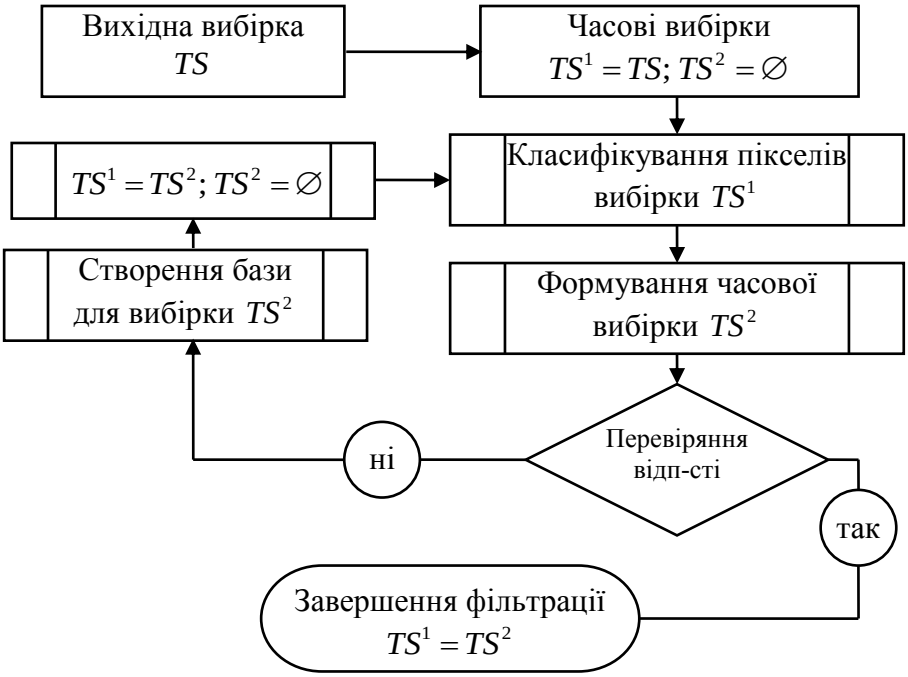


Рис. 2.4. Схема фільтрації навчальної вибірки з метою підвищення її чистоти

Фільтрація навчальної вибірки з метою підвищення її чистоти здійснюється у такій послідовності:

Крок 1. На вхід алгоритму класифікування поступає навчальна вибірка TS.

Крок 2. Введемо дві проміжні вибірки TS^1 и TS^2, такі, що $TS^1 = TS; TS^2 = \varnothing$.

Крок 3. За допомогою обраного алгоритму класифікування визначаємо класову приналежність всіх пікселів часової вибірки TS^1, і, у випадку їх відповідності заявленим класам, вносимо їхні властивості (належність до відповідного класу і координати) у часову вибірку TS^2.

Крок 4. Перевіряємо відповідність пікселів, що належать вибіркам TS^1 и TS^2.

Крок 5. У випадку позитивного результату ($TS^2 = TS^1$) вважаємо задачу фільтрації завершеною, приймаємо рішення для навчальної вибірки у вигляді $TS = TS^1$ і виходимо з процедури.

Крок 6. Виконуємо процедури розбиття ознакового простору і внутрішньоінтервальної оцінки у відповідності з одержаними у п. 3 класами і параметрами вибірки TS^1.

Крок 7. Здійснюємо заміщення $TS^1 = TS^2$, обнуляємо TS^2 і повертаємось до виконання п. 3.

Піксели, що не пройшли таку перевірку, виключаються зі складу навчальної вибірки TS, і в результаті отримують уточнену, більш чисту вибірку. Далі, використовуючи уточнену навчальну вибірку, виконують редагування (корекцію) описів класів.

2.2.2.3. Розбиття ознакового простору

Розбиття ознакового простору здійснюється з метою представити кожен клас деякою областю у просторі спектральних ознак розмірністю за числом задіяних для класифікування зональних зображень. Області формуються на основі значень сигналів пікселів навчальної вибірки, при цьому припускається, що числове значення ознаки класу у кожному спектральному каналі (зональному зображенні) розташовується у деякому інтервалі, положення і межі якого визначаються величинами сигналів пікселів класу в цьому каналі (зоні). Загальна кількість інтервалів в каналі визначається числом класів L; інтервали межують один з іншим, але не перетинаються.

Для будь-якого з розглянутих класів положення інтервалу на вісі спектральних значень задається середнім значенням сигналів пикселів даного класу у спектральному каналі (наприклад, у *k*-ому спектральному каналі (зоні) для класів *l* та *l+1* це будуть відповідно величини $\overline{u}_{k,l}$ та $\overline{u}_{k,l+1}$, рис. 2.6), а

положення граничної точки $A_{l,l+1}$, що розділяє ці два класи з близькими значеннями сигналів, визначається зі співвідношення

$$\frac{a_l}{a_{l+1}} = \frac{\sigma_{k,l}}{\sigma_{k,l+1}}, \qquad (2.6)$$

де $\sigma_{k,l}$ і $\sigma_{k,l+1}$ – середньоквадратичне відхилення сигналів пікселів відповідних класів у *k*-ому спектральному каналі (зоні).

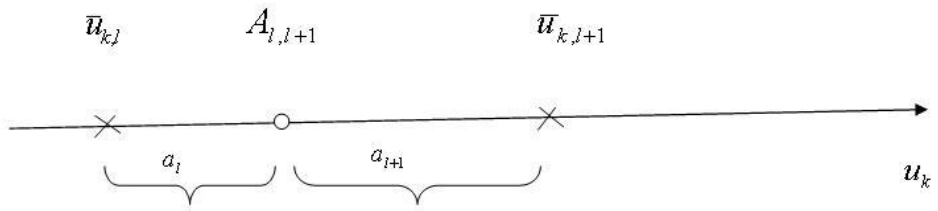


Рис. 2.5. До визначення положення границі, що розділяє класи на спектральній вісі

Аналогічно формуються границі та інтервальні описи для всіх інших класів, за винятком двох - класу з мінімальним середнім значенням й класу з максимальним середнім значенням сигналів пікселів у даному спектральному каналі. Для першого з названих класів не встановлюється нижня (ліва на рис. 2.6) гранична точка на спектральній осі, а для другого - верхня (права на рис. 2.6) гранична точка, тобто ці два інтервали є напіввідкритими.

Визначення: інтервал у *k*-му спектральному каналі, положення якого визначається величиною $\bar{u}_{k,l}$, будемо називати власним інтервалом класу *l*.

Вирішальне правило, на основі якого проводиться класифікування, працює з оцінками ймовірності приналежності об'єкта класу (або одночасно кільком класам). Ці оцінки обчислюються для всіх інтервалів кожного зі спектральних каналів за даними навчальної вибірки наступним чином. Для кожного інтервалу підраховуються дві величини - базова ймовірність власних

піксельних об'єктів (один фокальний елемент) та сумарна базова ймовірність об'єктів інших класів (всі ці об'єкти разом складають другий фокальний елемент). Нехай, наприклад, у деякому *j*-му інтервалі нараховується у цілому M_j пікселів, з них M_1 - це піксели класу, для якого даний інтервал – власний, і M_2 - загальне число пікселів інших класів. Тоді оцінка базової ймовірності піксельних об’єктів у власному інтервалі вказаного класу розраховується як

$$m^{(c)} = \frac{M_1}{M_j}, \tag{2.7}$$

а сумарна базова ймовірність піксельних об’єктів інших класів у цьому ж інтервалі розраховується як

$$m^{(нс)} = \frac{M_2}{M_j} \tag{2.8}$$

Очевидно, що у кожному інтервалі $m^{(c)} + m^{(нс)} = 1$.

2.2.2.4. Визначення класифікаційної цінності спектральних каналів

Досить складно провести розбиття ознакового простору таким чином, щоб у власних інтервалах знаходилися виключно піксели відповідного класу. Значно частіше у більшості інтервалів одночасно присутні піксели з кількох класів навчальної вибірки, що, безумовно, негативно впливає на класифікаційну здатність спектрального каналу та точність класифікування у цілому.

При оцінюванні інформативності спектральних каналів будемо вважати, що нам відома навчальна вибірка для кожного з K_m класів, $m = 1, 2, ..., M$. Кожна кількісна ознака π_n $(n = 1, 2, N)$ у відповідному спектральному каналі має певний діапазон значень, позначимо його d_{π_n}. У межах загальної навчальної вибірки I діапазон значень $n-ї$ ознаки розраховується як

$$d_{\pi_n} = \max_{\bar{I}} \pi_n - \min_{\bar{I}} \pi_n, \tag{2.9}$$

де $\max\limits_{\bar{I}} \pi_n$ та $\min\limits_{\bar{I}} \pi_n$ – відповідно найбільше та найменше значення цієї ознаки.

Для дослідження кореляції значень ознаки π_n з класами $K_1, K_2, ..., K_M$ розділимо діапазон значень ознаки на однакові за шириною інтервали. Ширина інтервалу визначається як

$$\Delta d_{\pi_n} = \frac{d_{\pi_n}}{\left|\bar{I}\right|}, \tag{2.10}$$

де $\left|\bar{I}\right|$ – потужність загальної навчальної вибірки.

Клас K_M будемо називати зіставленим певному інтервалу ознаки π_n, яка розглядається, якщо в навчальній виборці I_m позитивних прикладів по класу K_M матимемо приклад об'єкта, значення ознаки якого потрапляє до цього інтервалу. Якщо кількісна ознака для об'єктів різних класів в кожній з N спектральних зон розподілена на відповідні інтервали, то, використовуючи загальну навчальну вибірку, для будь-якої спектральної зони можна визначити, в який інтервал потрапили конкретні представники різних класів з цієї вибірки, а потім підрахувати кількість попадань класів у відповідні інтервали. Зрозуміло, розподіл попадань класів по інтервалах буде відрізнятися для різних спектральних каналів, а це означає, що різні спектральні канали неоднаково ефективні при класифікуванні об'єктів. Ефективність класифікування об'єктів може бути платформою для визначення й порівняння інформативних властивостей різних спектральних каналів [70–73].

Впровадимо функцію F, яку будемо далі називати критеріальною функцією інформативності спектрального каналу:

$$F = 1 - \frac{1}{L\ (L-1)} \cdot \sum_{m=1}^{L} \frac{\sum\limits_{j=1}^{nz} \left(\delta_{mj} \cdot \sum\limits_{\substack{k=1 \\ k \neq m}}^{L} \delta_{kj} \right)}{\sum\limits_{j=1}^{L} \delta_{mj}}, \tag{2.11}$$

де L – кількість класів, nz – число інтервалів для ознаки спектрального каналу, що розглядається, δ_{kj} – індикатор зіставлення,

$$\delta_{kj} = \begin{cases} 1, & \textit{якщо об'єкт класу } k \textit{ пов'язаний з інтервалом } j; \\ 0, & \textit{в інших випадках}. \end{cases} \quad (2.12)$$

Можна показати, що функція інформативності (2.11) має значення в діапазоні від 0 до 1, а її конкретна величина залежить від того, як зіставлені між собою класи та ознаки. Серед спектральних каналів той буде мати найбільше значення функції F, для ознаки якого число класів, одночасно зіставлених кожному інтервалу, мінімально. Саме той спектральний канал і вважається найбільш інформативним. Таким чином, формула (2.11) надає можливість кількісно порівнювати інформативності спектральних каналів та відбирати серед них найбільш ефективні для класифікування об’єктів.

Разом з тим в дисертації запропонована модифікація критеріальної функції інформативності, яка дозволяє враховувати кількість об’єктів кожного класу. Здійснюється це наступним чином. Для кожного інтервалу j обчислюємо відсоток (долю) “неправильних” пікселів, тобто тих пікселів, що помилково потрапили в даний інтервал. Для цього кількість “неправильних” пікселів ділимо на загальну кількість пікселів, що знаходяться в інтервалі. Отримуємо S_j – відношення кількості “неправильних” пікселів до загальної кількості пікселів в інтервалі j. Після того, як ми знайдемо значення всіх величин S_j для кожного інтервалу, нам треба їх просумувати та поділити на загальну кількість інтервалів.

Модифікована критеріальна функція інформативності має вигляд:

$$F^* = 1 - \left(\frac{S_1 + S_2 + S_3 + ... + S_j}{J} \right). \quad (2.13)$$

Формулу (2.12) можна записати у іншому вигляді, через індикатори:

$$F^* = 1 - \frac{1}{J}\left(S_1 + S_2 + S_3 + \ldots + S_j\right) = 1 - \frac{1}{J}\sum_{j=1}^{J}\left(\frac{\sum_{\substack{k=1 \\ k \neq j}}^{M} l_{kj} n_{kj}}{\sum_{k=1}^{M} l_{kj} n_{kj}}\right), \qquad (2.14)$$

де M – кількість класів, J – число інтервалів для ознаки спектрального каналу, що розглядається, l_{kj} – індикатор зіставлення,

$$l_{kj} = \begin{cases} 1, & \textit{якщо клас } K_k \textit{ зіставлений з інтервалом } j, \\ 0, & \textit{в іншому випадку.} \end{cases}$$

n_{kj} – кількість елементів з k – го класу, що потрапили в інтервал j.

Властивості модифікованої критеріальної функції інформативності (2.13) розглянемо на деяких прикладах.

Приклад 1. Припустимо, ми маємо два класи $(M = 2)$. Інтервал значень ознак у кожному спектральному каналі буде розбитий на 2 інтервали.

а) Нехай для першого спектрального каналу ми маємо наступні індикатори зіставлення:

$$l_{11} = 1, l_{12} = 0, l_{21} = 0, l_{22} = 1.$$

Також маємо значення n_{kj}, що вказують на число елементів з k – го класу, що потрапили в інтервал j:

$$n_{11} = 5, n_{12} = 0, n_{21} = 0, n_{22} = 5.$$

1) Спочатку обчислюємо звичайну критеріальну функцію інформативності для першого спектрального каналу. Для даного випадку формула (2.13) буде мати такий вигляд:

$$F_1 = 1 - \frac{1}{2}\left[\frac{l_{11} \cdot l_{21} + l_{12} \cdot l_{22}}{l_{11} + l_{12}} + \frac{l_{21} \cdot l_{11} + l_{22} \cdot l_{12}}{l_{21} + l_{22}}\right].$$

Тоді $F_1 = 1 - \frac{1}{2}\left[\frac{1 \cdot 0 + 0 \cdot 1}{1 + 0} + \frac{0 \cdot 1 + 1 \cdot 0}{0 + 1}\right] = 1.$

2) Тепер обчислюємо модифіковану критеріальну функцію інформативності із врахуванням кількості об’єктів, що потрапили в ці два інтервали:

$$F_1^* = 1 - \frac{1}{2}\left[\frac{l_{21}n_{21}}{l_{11}n_{11} + l_{21}n_{21}} + \frac{l_{12}n_{12}}{l_{12}n_{12} + l_{22}n_{22}}\right].$$

$$F_1^* = 1 - \frac{1}{2}\left[\frac{0 \cdot 0}{1 \cdot 5 + 0 \cdot 0} + \frac{0 \cdot 0}{0 \cdot 0 + 1 \cdot 5}\right] = 1.$$

Отже, в даному випадку ми маємо 100-відсотково вірну класифікацію, значення критеріальної функції інформативності та модифікованої критеріальної функції інформативності збігаються.

б) Тепер розглянемо другий спектральний канал з такими індикаторами зіставлення:

$l_{11} = 1, l_{12} = 0, l_{21} = 1, l_{22} = 1$ та такими кількостями об'єктів $k-$го класу, що потрапили в інтервал j:

$n_{11} = 5, n_{12} = 0, n_{21} = 1, n_{22} = 4.$

1) Тоді звичайна критеріальна функція інформативності для другого спектрального каналу буде мати вигляд:

$$F_2 = 1 - \frac{1}{2}\left[\frac{1 \cdot 1 + 0 \cdot 1}{1 + 0} + \frac{1 \cdot 1 + 1 \cdot 0}{1 + 1}\right] = 0,25.$$

2) Тепер обчислюємо модифіковану функцію інформативності для другого спектрального каналу:

$$F_2^* = 1 - \frac{1}{2}\left[\frac{1 \cdot 1}{1 \cdot 5 + 1 \cdot 1} + \frac{0 \cdot 0}{0 \cdot 0 + 1 \cdot 4}\right] = 0,9166666 \approx 0,92.$$

в) Розглянемо третій спектральний канал, що має такі ж самі індикатори зіставлення як попередній: $l_{11} = 1, l_{12} = 0, l_{21} = 1, l_{22} = 1,$ але має інші значення n_{kj}:

$n_{11} = 5, n_{12} = 0, n_{21} = 2, n_{22} = 3.$

Тобто, у даному випадку ми маємо гірший результат класифікування, ніж в попередньому, оскільки в цьому випадку кількість об'єктів, що була помилково віднесена з другого класу до першого інтервалу складає "2", а в першому випадку вона становила "1".

1) Розраховуємо звичайну критеріальну функцію інформативності:

$$F_3 = 1 - \frac{1}{2}\left[\frac{1\cdot 1 + 0\cdot 1}{1+0} + \frac{1\cdot 1 + 1\cdot 0}{1+1}\right] = 0,25.$$

Звідси ми бачимо, що отримана звичайна критеріальна функція інформативності співпадає із критеріальною функцією інформативності для другого спектрального каналу, тобто: $F_3 = F_2$, хоча, в дійсності, другий і третій канал дають різні результати класифікування. Тобто, звичайна критеріальна функція не спроможна вловити та показати різницю в точності класифікування для другого і третього спектрального каналів, оскільки вона не враховує кількості об'єктів.

2) Тепер обчислюємо модифіковану функцію інформативності для третього спектрального каналу:

$$F_3^* = 1 - \frac{1}{2}\left[\frac{1\cdot 2}{1\cdot 5 + 1\cdot 2} + \frac{0\cdot 0}{0\cdot 0 + 1\cdot 3}\right] = 0,8571428 \approx 0,86.$$

З отриманого результату ми бачимо, що значення модифікованої критеріальної функції інформативності для третього спектрального каналу менше ніж для другого: $F_3^* < F_2^*$, тобто модифікована критеріальна функція спроможна показати різницю між другим та третім спектральними каналами. З отриманого результату ми можемо зробити висновок, що результат класифікування з використанням третього спектрального каналу є менш точним ніж з використанням другого. Отже, другий канал є більш інформативним ніж третій.

Приклад 2. Тепер розглянемо приклад, коли ми маємо 3 класи $(M = 3)$. У кожному спектральному каналі інтервал значень ознак буде розбитий на 3 інтервали.

Припустимо, що ми маємо два спектральні канали $(N = 2)$. Оцінимо інформативність кожного з цих каналів за допомогою двох функцій інформативності та порівняємо результати.

а) Нехай для першого спектрального каналу ми маємо такі індикатори зіставлення:

$$l_{11}=1, l_{12}=0, l_{13}=0, l_{21}=1, l_{22}=1, l_{23}=0, l_{31}=0, l_{32}=0, l_{33}=1.$$

Значення n_{kj}, що вказують на число елементів з $k-$го класу, що потрапили в інтервал j будуть мати вигляд:

$$n_{11}=5, n_{12}=0, n_{13}=0, n_{21}=1, n_{22}=4, n_{23}=0, n_{31}=0, n_{32}=0, n_{33}=5.$$

1) Тоді звичайна критеріальна функція інформативності згідно з формулою (2.7) для першого спектрального каналу буде мати вигляд:

$$F_1=1-\frac{1}{6}\left[\frac{l_{11}\left(l_{21}+l_{31}\right)+l_{12}\left(l_{22}+l_{32}\right)+l_{13}\left(l_{23}+l_{33}\right)}{l_{11}+l_{12}+l_{13}}+\right.$$

$$+\frac{l_{21}\left(l_{11}+l_{31}\right)+l_{22}\left(l_{12}+l_{32}\right)+l_{23}\left(l_{13}+l_{33}\right)}{l_{21}+l_{22}+l_{23}}+$$

$$\left.+\frac{l_{31}\left(l_{11}+l_{21}\right)+l_{32}\left(l_{12}+l_{22}\right)+l_{33}\left(l_{13}+l_{23}\right)}{l_{31}+l_{32}+l_{33}}\right].$$

Тоді маємо:

$$F_1=1-\frac{1}{6}\left[\frac{1\cdot(1+0)+0\cdot(1+0)+0\cdot(0+1)}{1+0+0}+\frac{1\cdot(1+0)+1\cdot(0+0)+0\cdot(0+1)}{1+1+0}+\right.$$

$$\left.+\frac{0\cdot(1+1)+0\cdot(0+1)+1\cdot(0+0)}{0+0+1}\right]=0,58.$$

2) Тепер обчислюємо модифіковану функцію інформативності для першого спектрального каналу:

$$F_1^*=1-\frac{1}{3}\left[\frac{1\cdot1+0\cdot0}{1\cdot5+1\cdot1+0\cdot0}+\frac{0\cdot0+0\cdot0}{0\cdot0+1\cdot4+0\cdot0}+\frac{0\cdot0+0\cdot0}{0\cdot0+0\cdot0+1\cdot5}\right]=0,9444445\approx0,94.$$

б) Розглянемо другий спектральний канал, що має такі ж самі індикатори зіставлення як попередній:

$$l_{11}=1, l_{12}=0, l_{13}=0, l_{21}=1, l_{22}=1, l_{23}=0, l_{31}=0, l_{32}=0, l_{33}=1.$$

але має інші значення n_{kj}:

$$n_{11}=5, n_{12}=0, n_{13}=0, n_{21}=2, n_{22}=3, n_{23}=0, n_{31}=0, n_{32}=0, n_{33}=5.$$

Тобто, в даному випадку ми маємо менш точний результат класифікування, ніж в попередньому, оскільки в цьому випадку було

помилково віднесено не до того класу на один об’єкт більше, ніж в попередньому випадку.

1) Тоді звичайна критеріальна функція інформативності для другого спектрального каналу буде мати вигляд:

$$F_2 = 1 - \frac{1}{6}\left[\frac{1\cdot(1+0)+0\cdot(1+0)+0\cdot(0+1)}{1+0+0} + \frac{1\cdot(1+0)+1\cdot(0+0)+0\cdot(0+1)}{1+1+0} + \right.$$
$$\left. + \frac{0\cdot(1+1)+0\cdot(0+1)+1\cdot(0+0)}{0+0+1}\right] = 0,58.$$

Маємо, що $F_1 = F_2$, тобто звичайна критеріальна функція не спроможна вловити різницю класифікування між випадками а) та б).

2) Тепер обчислюємо модифіковану функцію інформативності для другого спектрального каналу:

$$F_2^* = 1 - \frac{1}{3}\left[\frac{1\cdot 2+0\cdot 0}{1\cdot 5+1\cdot 2+0\cdot 0} + \frac{0\cdot 0+0\cdot 0}{0\cdot 0+1\cdot 3+0\cdot 0} + \frac{0\cdot 0+0\cdot 0}{0\cdot 0+0\cdot 0+1\cdot 5}\right] = 0,904762 \approx 0,9.$$

Отже, маємо, що: $F_2^* < F_1^*$. З цього випливає, що результат класифікування із використанням другого спектрального каналу є менш точним ніж з використанням першого. Перший канал є більш надійним та інформативним, ніж другий канал.

З даних прикладів можна зробити висновок, що запропонована модифікована критеріальна функція інформативності враховує кількість об’єктів кожного класу і забезпечує досить точну оцінку інформативності спектральних каналів.

Після визначення класифікаційної цінності спектральних каналів за допомогою формули (2.11) або (2.13) процес створення БЗ завершується, система підготовлена до проведення класифікування.

2.2.2.5. Алгоритм класифікування

Класифікування ГКЗ здійснюється по-піксельно, тому достатньо розглянути процедуру класифікування одного піксела. Припустимо, піксельний

об’єкт π_n характеризується векторним сигналом $\mathbf{u}_n = (u_{1n}, u_{2n}, \cdots u_{kn}, \cdots, u_{Kn})$. Потрібно, грунтуючись на цьому сигналі, визначити, до якого (або яких) з L класів даний піксел найбільш ймовірно належить.

Алгоритмічна процедура класифікування включає наступні кроки.

Крок 1. У кожному спектральному каналі визначається інтервал, з яким пов'язаний відповідний сигнал u_{kn} піксела π_n ($k=1,2,\ldots,K$).

Крок 2. Встановлюються імена усіх класів об'єктів, що знаходяться у кожного такому інтервалі, після чого з БЗ отримуються значення базової ймовірності власних об'єктів інтервалу та сумарної базової ймовірності об'єктів інших класів всередині інтервалу.

Крок 3. Розраховують за допомогою правила Демпстера комбіноване значення мас для результуючих фокальних елементів (у вигляді окремих класів або підмножин класів).

Крок 4. Використовуючи правило максимальної правдоподібності, приймається рішення про найбільш ймовірний клас (класи) даного піксельного об'єкта.

Така процедура послідовно застосовується до всіх пікселів гіперспектрального зображення. По закінченні класифікування усього кадру ГКЗ будується, при наявності референсних (завіркових) даних, матриці помилок і на основі її елементів розраховуються показники точності проведеного класифікування, які далі перевіряються на відповідність встановленим (заданим) вимогам.

Програмне забезпечення для реалізації розробленого метода буде представлено в розділі 4. Там же будуть представлені результати оцінювання ефективності даного метода, а також ще одного нового методу, який розглянемо у наступному розділі.

ВИСНОВКИ ДО РОЗДІЛУ 2

1. Розглянуто основні положення теорії свідчень Демпстера-Шейфера, яка покладена в основу побудови методів класифікування гіперспектральних космічних зображень, що розглядаються у цьому та наступному розділах.

2. Розроблено новий метод класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок. Введено модель гіперспектрального космічного зображення та припущення. Детально розглянуто організацію і зміст методу класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок. Запропонований метод відрізняється від відомих розв'язків однойменних задач введенням процедури оцінювання класифікаційної цінності спектральних каналів за допомогою спеціальної емпіричної функції і особливим способом розбиття спектрального ознакового простору, що, у сукупності, дозволяє застосовувати лише найбільш інформативні спектральні канали і, таким чином, суттєво скоротити розмірність ознакового простору при одночасному підвищенні точності класифікації.

3. Детально описано процедуру відбору зональних зображень для виключення сильно зашумлених зображень із використанням коефіцієнту кореляції між сусідніми зональними зображеннями, описано процедуру виділення навчальної вибірки, перераховано основні вимоги до навчальних вибірок, такі як: чистота / забрудненість, достатність і репрезентативність, описано процедуру фільтрації навчальної вибірки з метою підвищення її чистоти, описано процедуру розбиття ознакового простору, наведено розрахунки внутрішньоінтервальних оцінок для класів.

4. Введено функцію інформативності для визначення класифікаційної цінності спектральних каналів, за допомогою якої відбираються найбільш інформативні спектральні канали для класифікування. Після визначення

класифікаційної цінності спектральних каналів надано детальний опис процедури проведення класифікування.

РОЗДІЛ 3 РОЗРОБКА МЕТОДУ КОНТРОЛЬОВАНОГО КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ В УМОВАХ ЗАБРУДНЕННЯ НАВЧАЛЬНИХ ВИБІРОК ТА КІЛЬКІСНИХ ОБМЕЖЕНЬ

У розділі розглянуто новий метод класифікування ГКЗ на основі ТСДШ, який орієнтований на вирішення цієї задачі в умовах, коли навчальна вибірка є не тільки забрудненою, а ще й малою за кількістю елементів, що її складають. Обмежена кількість елементів у навчальній виборці заважає застосовувати процедуру її фільтрації, яка входила до складу схеми метода, запропонованого і розглянутого у розділі 2 (див. п. 2.2.2.2).

Особливостями метода є удосконалена процедура навчання та впровадження правила прийняття рішення на основі пігністичної ймовірності.

Також показано, що процедура відбору зональних зображень може бути автоматизована. Розроблено спосіб визначення репрезентативної комбінації зональних зображень при класифікуванні гіперспектрального космічного знімка, який мінімізує навантаження на експерта і дає можливість об’єктивізувати і прискорити процес вирішення задачі за рахунок зменшення вимог до необхідного обчислювального ресурсу.

Метод реалізован програмно. Результати експериментальних досліджень точності класифікування методу буде наведено у розділі 4.

3.1. Обмеження, припущення, визначення

Як і раніше у розділі 2, будемо спиратися на модель ГКЗ у вигляді

$$HSI = \left\{ S_k \,\middle|\, k = 1,2,\ldots,K \right\} = \left\{ \left(\pi_n, \mathbf{u_n}\right) \,\middle|\, n = 1,2,\ldots,N_\pi \right\}, \qquad (3.1)$$

де S_k – k-те зональне зображення; $S_k = \left\{ \boldsymbol{\pi}_n \,,\, u_{nk} \right\}_{n=1}^{N_\pi}$;

K – загальна кількість зональних зображень (спектральних каналів, або зон); π_n – n-ий піксел з повним (векторним) сигналом $\mathbf{u}_n = (u_{1n}, u_{2n}, \ldots u_{kn}, \ldots, u_{Kn})$; u_{kn} – величина сигналу піксела π_n у k-му спектральному каналі (зоні); N_π – загальна кількість пікселів в кадрі ГКЗ.

Кожен піксел ГКЗ має свій унікальний номер n, а рівень сигналу піксела визначається відображеним фізичним об'єктом (або елементом об'єкта). Приналежність об'єкту та/або відповідного піксела до того чи іншого класу будемо позначати позначкою l. Якщо L –загальна кількість наявних класів, тоді позначка l конкретного піксела ГКЗ приймає яке-небудь одне із значень в діапазоні від 1 до L.

Із врахуванням прийнятих позначень задача класифікування ГКЗ зводиться до того, щоб шляхом аналізу сигналів $\{u_n\}_{n=1}^{N_\pi}$ пікселів зображення π_n максимально вірно надати відповідну позначку кожному із них.

В данному розділі задача класифікування ГКЗ розв'язується при наступних обмеженнях та припущеннях:

1. Номенклатура класів, до яких можуть належати об'єкти на зображенні, відома, при цьому експерту-досліднику доступна навчальна вибірка із представниками (репрезентами) кожного з класів.

2. Серед елементів навчальної вибірки можуть бути присутні сторонні елементи, а об'єм навчальної вибірки може не задовольняти умові (1.11).

3. Кожен конкретний піксельний об'єкт реально належить тільки якомусь одному з L класів.

4. При прийнятті рішення щодо приналежності кожного окремого піксельного об'єкта відповідному классу з числа наявних будемо використовувати критерій максимума пігністичної ймовірності (*Pignistic Probability) BetP* [74–77].

Визначення. Пігністична ймовірність *BetP* одноелементної фокальної підмножини (*Singleton Subset*) $B \in \Omega$ обчислюється як

$$BetP\ (B) = \sum_{A \in \Omega} m(A) \frac{|B \cap A|}{|A|}. \qquad (3.2)$$

Значення пігністичної ймовірності завжди знаходиться в межах інтервалу довіри:

$$Bel\ (B) \le BetP\ (B) \le Pl\ (B).$$

Як і будь-яка інша ймовірність, пігністична ймовірність нормалізована до одиниці.

Критерій максимума пігністичної ймовірності має вигляд:

$$\mathrm{D} = \arg\left(\max_{\mathrm{A} \in 2^{\mathrm{N}\Omega}} \left(\mathrm{BetP}\,(\mathrm{A}) \right) \right), \qquad (3.3)$$

де D – рішення щодо класової приналежності піксела, яке приймається на множині всіх наявних класів.

Задача – розробити для вказаних умов, обмежень та припущень простий та ефективний метод класифікування ГКЗ.

3.2. Організація і процедури методу контрольованого класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень

3.2.1. Загальна схема класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень

Загальна схема організації виконання логіко-обчислювальних процедур при класифікуванні ГКЗ в умовах забруднення навчальних вибірок та кількісних обмежень представлена на рис. 3.1. Метод включає такі основні процедури: відбір зональних зображень, навчання, формування бази даних (БД); здійснення класифікування ГКЗ.

Процедура навчання виконується експертом-дослідником, який виділяє навчальну вибірку з представників різних класів. Після цього шляхом аналізу повних (векторних) сигналів пікселів виділеної вибірки формуються описи еталонів кожного з наявних класів. Ці описи накопичуються у БД і використовуються при класифікуванні.

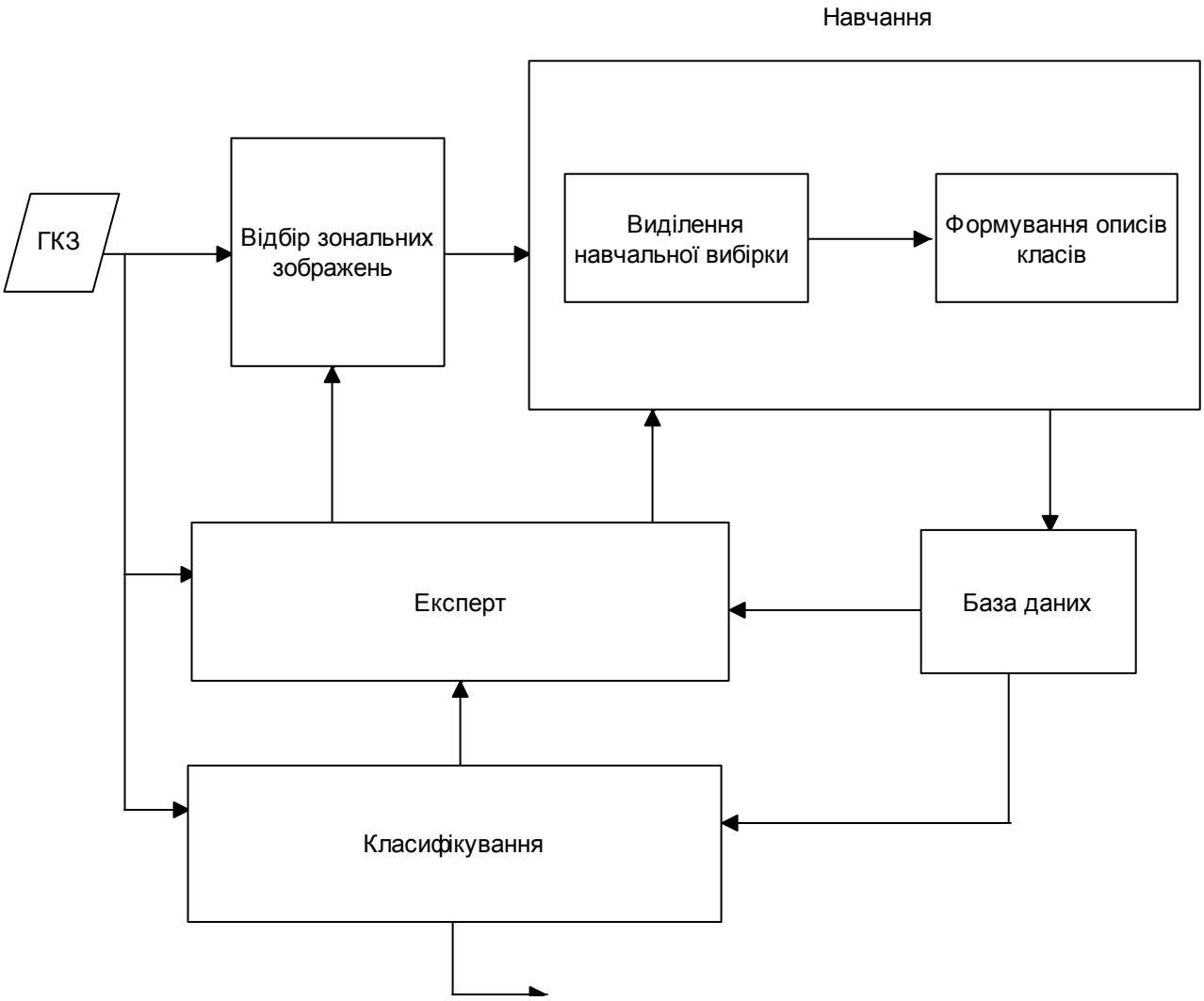


Рис. 3.1. Загальна схема організації виконання логіко-обчислювальних процедур при класифікуванні ГКЗ в умовах забруднення навчальних вибірок та кількісних обмежень

По завершенні класифікування ГКЗ проводяться оцінювання точності класифікування і перевіряння на задоволення встановленим (заданим) вимогам.

Розглянемо зміст логіко-обчислювальних процедур, що складають процес класифікування за пропонованим методом.

3.2.2. Відбір зональних зображень

При плануванні процедури відбору зональних зображень слід базуватись на тому, що, як правило, у складі ГКЗ частина зональних зображень характеризується значним рівнем шумів, а також доволі багато зональних зображень мають високий рівень схожості [74–80].

Тому процедура відбору зональних зображень включає два кроки. На першому кроці виявляються та відкидаються сильно зашумлені зональні зображення, а на другому кроці ті зображення, що залишились, аналізуються на предмет схожості. Якщо при аналізі виявлено кілька схожих зображень, то деякі з них відкидаються, й в результаті формується скороченна так звана репрезентативна комбінація зональних зображень.

Алгоритм зменшення числа зональних зображень наведено на рис. 3.2 у псевдокоді.

n: initial number of images S_i
K: desirable number of images
For $i := 1$ to n
For $j := 1$ to n
S is a vector of images S_i
Sort S
For $d := 2$ to n
For $l := 0$ to $n - d$
begin
write (“Поступово та попарно об’єднуючи d зображень між собою $S_i, S_j, ..., S_d$, обчислюємо міру Кульбака-Лейблера: $D_{KL}[(S_i, S_j, ...), S_d]$
readln ($D_{KL}[(S_i, S_j, ...), ..., S_d]$);

For $z := 1$ to $n - 1$ do
begin
if $D_{KL}[(S_i, S_j, ...), S_{df}] < D_{KL}[(S_i, S_j, ...), S_{dz}]$ then we unite current cluster $(S_i, S_j, ...)$ and image S_{df} into one cluster

else: we look for another nearest image for the current cluster $(S_i, S_j, ...)$
end.
end.
write (“we get a new number of clusters: K”);
end.
L: number of spectral bands for cluster C_i
function max(mas:array of functions W_i);
begin
$W_i := mas[1]$
for i:=1 to L do
begin
if $W_i < mas[i]$ then
$W_i := mas[i]$;
end;
max $:= W_i$;
end;
write (“we get a most informative spectral band for the cluster C_i”);
end.
write (“we get a new number of spectral bands, K”);
end.

Рис. 3.2. Псевдокод. Алгоритм зменшення числа спектральних каналів

Зазначимо, що під репрезентативною комбінацією зональних зображень розуміють таку, яка, з одного боку, забезпечує збереження інформаційного контенту, ще достатнього для вирішення завдань з класифікування, а з іншого боку – надає можливість суттєво зменшити обчислювальні витрати.

У найбільш простому випадку зональні зображення зі значним рівнем шумів можуть бути виявлені експертом методом візуального перегляду складових ГКЗ [81–87].

Рівень схожості зональних зображень можна кількісно оцінити за допомогою коефіцієнта їх кореляції. Такий підхід був залучений нами раніше при відборі зональних зображень при реалізації метода класифікування ГКЗ в умовах забруднених навчальних вибірок (п. 2.2.2.1).

Однак спосіб відбору зональних зображень шляхом візуального перегляду створює певне навантаження на експерта, а його оцінки носять суб'єктивний характер. Тому нами розроблено спосіб визначення репрезентативної комбінації зональних зображень при класифікуванні гіперспектрального космічного знімка, який мінімізує навантаження на експерта і дає можливість об'єктивізувати і прискорити процес вирішення задачі за рахунок зменшення вимог до необхідного обчислювального ресурсу.

Мета досягається шляхом автоматизації процедури оцінювання рівню шумів кожного з вихідних зональних зображень і відсіювання з подальшого аналізу зображень, які мають високий рівень зашумленості.

Операційна схема запропонованого способу наведена на рис. 3.3. Тут позначено: 1 – гіперспектральний аерокосмічний знімок, що складається з N зональних зображень; 2 – операція відбору зональних зображень; 3 – операція побудови функції розподілу величин сигналів пікселів для кожного відібраного зонального зображення; 4 – операція кластеризації відібраних зображень; 5 – операція встановлення у кожному кластері відповідного зображення з максимальною репрезентативністю; 6 – репрезентативна комбінація зональних зображень.

Згідно наведеної схеми спосіб реалізується шляхом послідовного виконання чотирьох операцій. Розглянемо детально їх зміст.

Операція відбору зональних зображень. ГКЗ складається з N зональних зображень, які мають однакове просторове розрізнення і одну кількість пікселів (наприклад, $K_i \times K_j$), але зображення можуть сильно відрізнятися між собою за шумовими властивостями. В роботі [88] зазначається, що зональні зображення зі складу одного ГКЗ можуть бути вражені шумами не тільки різного рівня, але й різного ймовірнісного розподілу, зокрема, Гауса, Пуасона та ін. Тому при оцінюванні шумів доцільно застосовувати параметричний підхід.

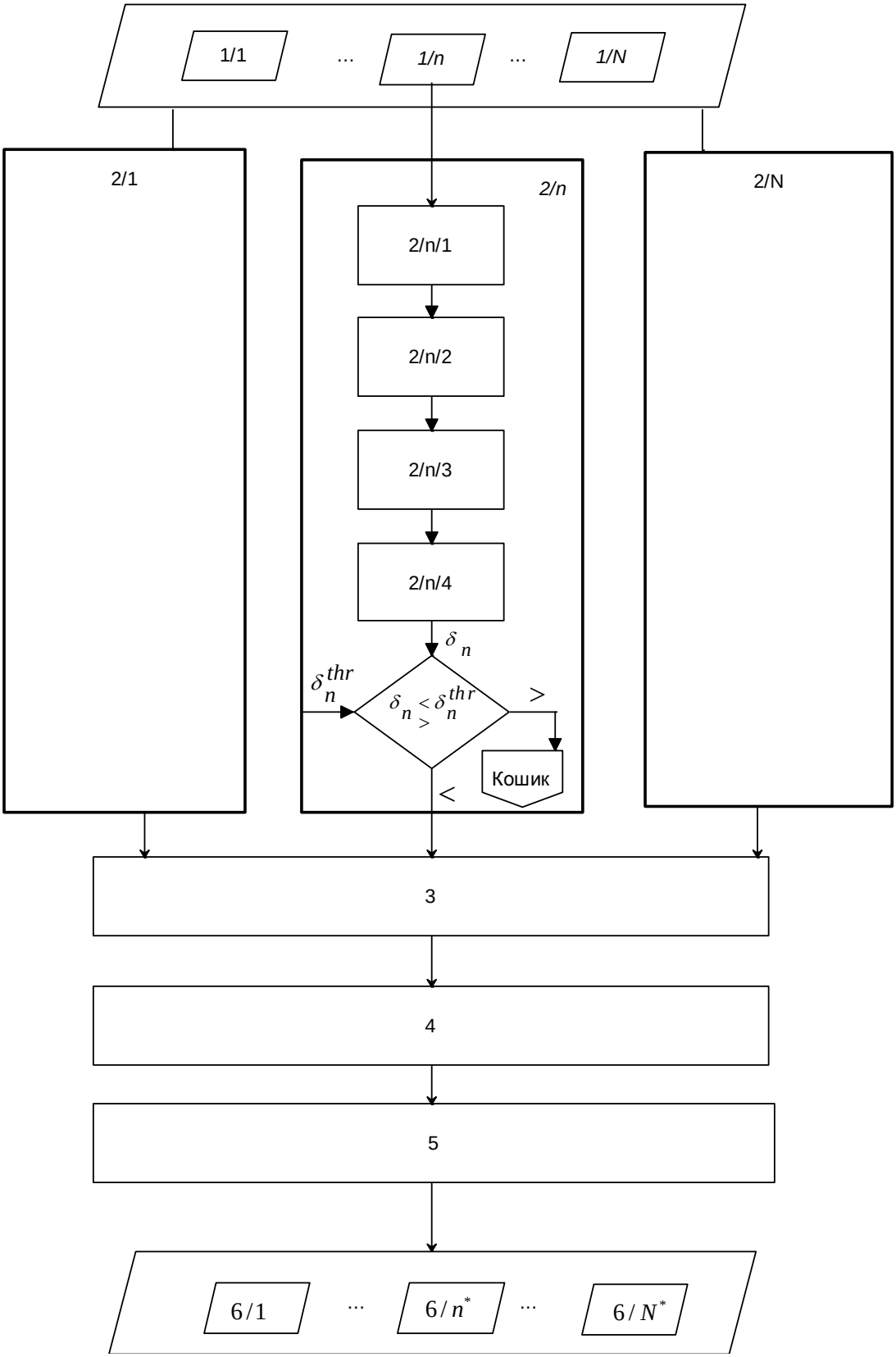


Рис. 3.3. Операційна схема способу визначення репрезентативної комбінації зональних зображень

Найбільш простою оцінкою може бути середньоквадратичне відхилення шуму σ_{noise} [89], його величина розраховується як

$$\sigma_{noise} = \sqrt{\frac{\sum_{i=1}^{K_i}\sum_{j=1}^{K_j}\Delta f^2(i,j)}{K_i \times K_j}}, \tag{3.4}$$

де $\Delta f(i,j)$ – величина шуму у *(i, j)* пікселі зображення.

Відбір зональних зображень здійснюється однаково, тому опишемо цю процедуру на прикладі одного будь-якого зонального зображення. Позначимо, що розподіл сигналів пікселів цього зображення описується як $G(i,j);\ i = 1,2,...,K_i;\ j = 1,2,...,K_j$.

Процедура відбору містить п’ять кроків:

1. Обчислюється медіана $median\{G\}$ на множині сигналів пікселів, що складають зональне зображення $G(i,j);\ i = 1,2,...,K_i;\ j = 1,2,...,K_j$. При обчисленні медіани застосовується ковзне вікно розміром 3x3 або 5x5 пікселів.

2. Формується матричний масив $MAD(i,j)$ різниць між цією медіаною і сигналами всіх пікселів зонального зображення:

$$MAD(i,j) = G(i,j) - median\{G\}. \tag{3.5}$$

3. Обчислюється медіана матричного масиву $MAD(i,j)$:

$$median\{MAD(i,j)\}. \tag{3.6}$$

4. Розраховується оцінка шуму в зображенні:

$$\sigma_{noise} = c \cdot median\{MAD(i,j)\}, \tag{3.7}$$

де c – постійний коефіцієнт. Його величина залежить від закону розподілу шумів і розраховується за певним алгоритмом [90]. Зокрема, при гаусовському характері шумів $c = 1.483$.

5. Розрахована на кроці 4 оцінка шуму, що присутній у зображенні, порівнюється з порогом σ_{noise}^{thr}. Якщо для даного зображення має місце $\sigma_{noise} \le \sigma_{noise}^{thr}$, тоді воно зберігається для послідуючого оброблення; зображення з

рівнем шумів $\sigma_{noise} > \sigma_{noise}^{thr}$ виключається з подальшого розгляду як сильно зашумлене.

В результаті застосування описаної процедури до всіх зображень зі складу гіперспектрального знімка відбираються лише такі зональні зображення, які характеризуються відносно невеликим рівнем шуму. Сумарна кількість N^* подібних зональних зображень зазвичай значно менше, ніж їх первинна кількість *N*. Так, результати експериментальних досліджень, проведених авторами на реальних зображеннях, одержаних сучасними бортовими гіперспектральними сенсорами EO-1 / Hyperion та AVIRIS, показали, що частка зашумлених, малоперспективних для подальшого опрацювання зональних зображень складає від 39 до 52% для знімків EO-1 та від 40 до 49% у випадку знімків AVIRIS.

Операція побудови функції розподілу величин сигналів пікселів для кожного відібраного зонального зображення. Загально кажучи, сигнали пікселів будь-якого зображення є випадковими величинами $\{w\}$- елементами простору Ω дійсних чисел. Виходячи з цього, для кожного відібраного зонального зображення будується функція розподілу величин сигналів його пікселів. Побудова функції розподілу здійснюється за способом, описаним в [70–90]. Нехай $G_r(i, j)-$ деяке відібране зональне зображення з номером *r*; функцію розподілу величин сигналів його пікселів будемо позначати як $p_r(w)$. Усього на даному операційному етапі будується N^* функцій розподілів по числу відібраних зональних зображень.

Операція кластеризації відібраних зображень. Здійснюють об’єднання відібраних зображень в окремі кластери по критерію близькості їх функцій розподілу. Для розрахунку ступеня близькості функцій розподілу зображень застосовується симетрична версія дивергенції Кульбака-Лейблєра [91]. Якщо ми маємо два зображення G_r і G_q з функціями розподілу сигналів їх пікселів

$p_r(w)$ і $p_q(w)$ відповідно, тоді симетрична версія дивергенції Кульбака-Лейблєра $D(G_r, G_q)$ має вигляд:

$$D(G_r, G_q) = \sum_{w \in \Omega} p_r \log \frac{p_r(w)}{p_q(w)} + \sum_{w \in \Omega} p_q \log \frac{p_q(w)}{p_r(w)} \quad . \tag{3.8}$$

На початку об'єднання зображень в кластери визначається (задається) число кластерів (позначимо це число як *CN*) і розраховується типовий об'єм *V* одного кластера (тобто кількість зональних зображень у кластері). Для цього слугує вираз: $V = N^* / CN$.

Кластери формуються послідовно, один за одним. Розглянемо, як формується перший кластер *Cluster1*. Вибір зонального зображення – засновника першого кластера здійснюється випадково на множині всіх N^* зональних зображень, що були відібрані за описаною вище процедурою. Припустимо, це буде деяке зображення G_r. Для зображення G_r.обчислюються дивергенції Кульбака-Лейблєра за формулою (3.8) на множині всіх зображень від $\Theta = 1$ до $\Theta = N^*$; $(\Theta \neq r)$, і відбирається зображення, яке має мінімальну відносно G_r.величину дивергенції. Припустимо, це деяке зображення G_q, тобто саме з ним мале місце:

$$D(G_r, G_q) = \min \{D(G_r, G_\Theta)\}, \quad \Theta = 1,2,..., N^*; \quad \Theta \neq r . \tag{3.9}$$

Тепер кластер *Cluster* складають два елементи - зображення G_r.та G_q.і, якщо для кластера задано об'єм *V* > 2, то потрібно додати до них чергове зображення, причому воно має бути найближчим для елементів цього кластера. Для обчислення відстані між двохелементним кластером (G_r, G_q) та черговим розглядуваним зображенням G_h застосовується формула [92]:

$$D[(G_h), (G_r, G_q)] = \alpha \cdot D(G_h, G_r) + \beta \cdot D(G_h, G_q) + \gamma \cdot D(G_r, G_q) \quad , \tag{3.10}$$

де *α, β, γ* – коефіцієнти, що розраховуються таким чином:

$$\alpha = \frac{n_r + n_h}{n_r + n_q + n_h}; \qquad \beta = \frac{n_q + n_h}{n_r + n_q + n_h}; \quad \gamma = \frac{-n_h}{n_r + n_q + n_h}. \tag{3.11}$$

У виразах (3.11) літерою *n* з нижнім індексом позначена поточна кількість зображень у відповідному кластері на момент розрахунку.

Вирази (3.11) для розрахунку величин коефіцієнтів обрані виходячи з того, щоб кластери, які формуються за формулою (3.10), розташовувалися максимально далеко один від одного, що у подальшому позитивно впливає на точність класифікування [92].

Застосовуючи вирази (3.10) та (3.11), послідовно обчислюють відстані між кластером (G_r, G_q) та іншими $N^* - 2$ зображеннями і серед них визначають таке, що характеризується мінімальною відстанню від (G_r, G_q). Цим зображенням даний кластер поповнюється, й знову застосовується описана процедура. Це продовжується, поки об’єм першого кластера досягне величини *V*.

Після цього починається формування наступного кластеру *Cluster2*. Зональне зображення – його засновник обирається, як і раніше, випадково на множині зональних зображень, що залишилися. Для обраного зображення в процесі обчислень за формулами (3.10) і (3.11) і відповідних порівнянь визначається найближчий за величиною дивергенції елемент, і вони вже разом представляють цей кластер. Далі за описаною вище процедурою визначається третій складовий елемент цього кластера, і т. д. По наповненні кластера *Cluster2* здійснюється перехід до формування наступного кластера *Cluster3* і т. д., поки не буде сформовано *CN* кластерів.

Операція встановлення у кожному кластері відповідного зображення з максимальною репрезентативністю. У кожному сформованому кластері встановлюється зображення, функція розподілу якого найкращим чином представлятиме у цілому функції розподілу всіх інших зображень цього кластера. Якщо деякий сформований кластер *Cluster* налічує у своєму складі *V*

певних зональних зображень, то для кожного *i*-го зображення $G_i \in Cluster$ розраховується його репрезентативність

$$R_i = \frac{1}{\sum\limits_{\substack{j=1 \\ j \neq i}}^{V} D\left(G_i, G_j\right)}, \tag{3.12}$$

де відповідні елементи-доданки розраховуються за виразом (3.8); *i=1,2,…,V*.

Шляхом порівняння величин репрезентативності R_i, розрахованих за допомогою виразу (3.12) для різних зображень (*i=1,2,…,V)* даного кластера, виявляється зображення, для якого величина R є максимальною. Це зображення, як таке, чия функція розподілу найкращим чином представляє у цілому функції розподілу всіх інших зображень даного кластера, додається до складу репрезентативної комбінації зональних зображень гіперспектрального космічного знімка.

Визначені подібним чином зображення з усіх *CN* кластерів разом утворюють репрезентативну комбінацію зональних зображень гіперспектрального аерокосмічного знімка.

Досягнення поставленої мети – зменшити вимоги до необхідного обчислювального ресурсу і прискорити процес вирішення задачі за рахунок застосування процедури оцінювання рівню шумів кожного з вихідних зональних зображень з відсіюванням з подальшого аналізу зображень, які мають високий рівень зашумленості, ілюструють дані, наведені в табл. 3.1. Тут наведено обчислювальні витрати, потрібні на визначення репрезентативної комбінації зональних зображень при класифікуванні гіперспектральних космічних знімків EO-1 / Hyperion та AVIRIS.

З табл. 3.1 видно, що введення операції відбору зашумлених зональних зображень у випадку знімків EO-1 / Hyperion дозволяє зменшити необхідні обчислювальні витрати в 2,78-4,64 рази, а у випадку знімків AVIRIS – в 2,92-3,64 рази.

Таблиця 3.1– Обчислювальні витрати на визначення репрезентативної комбінації зональних зображень при класифікуванні гіперспектральних аерокосмічних знімків

	EO-1 / Hyperion			AVIRIS		
	Без відбору зашумлених зональних зображень	З відбором зашумлених зональних зображень, відсіювання 39%	З відбором зашумлених зональних зображень, відсіювання 52%	Без відбору зашумлених зональних зображень	З відбором зашумлених зональних зображень, відсіювання 40%	З відбором зашумлених зональних зображень, відсіювання 49%
Обчислювальні витрати, TFlop	7,70	2,78	1,66	6,22	2,13	1,71

У загальному випадку графіки розподілу рівня необхідних обчислювальних витрат при реалізації запропонованого способу в залежності від частки зашумлених зональних зображень в гіперспектральному космічному знімку, що класифікується, наведено на рис. 3.3. Графіки показують, що запропонований спосіб дозволяє суттєво скоротити обчислювальні витрати на визначення репрезентативної комбінації зональних зображень при класифікуванні гіперспектральних космічних знімків.

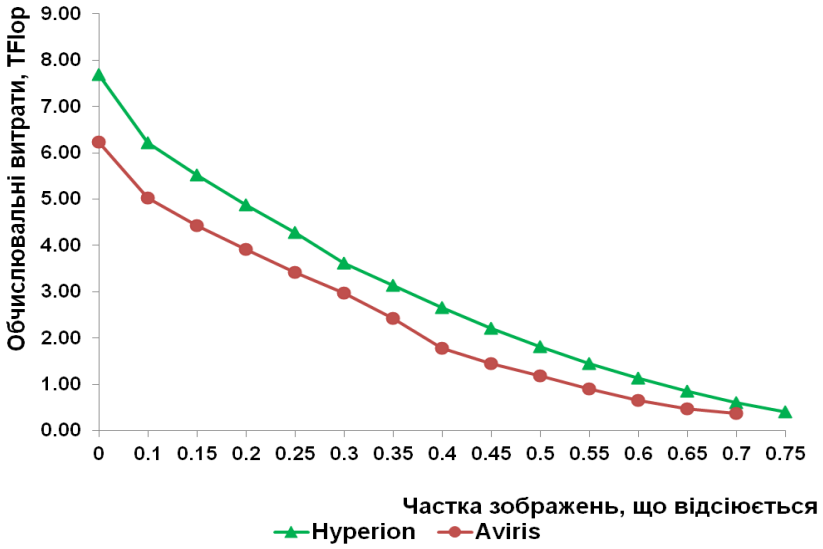


Рис. 3.4. Розподіл рівня необхідних обчислювальних витрат для запропонованого способу визначення репрезентативної комбінації зональних зображень в залежності від частки зашумлених зональних зображень в гіперспектральному космічному знімку

По завершенні процесу відбору зональних зображень робочим сигнальним спектральним простором стає простір меншої розмірності R^{K^*} $\left(K^* << K\right)$, та розмірність вектору сигналу піксела π_n відповідно понижується до величини K^*. Тобто тепер векторний сигнал будь-якого n-го піксела π_n матиме вигляд:

$$u_n = \left\{ u_{nk} \right\}_{k=1}^{K^*}, \quad u_n \in R^{K^*}. \tag{3.13}$$

Зауважимо, що видалення корельованих зональних зображень дозволяє не тільки зменшити об'єм подальших обчислень, але і сприяє більш строгому виконанню вимог щодо взаємної незалежності джерел даних, що комбінуються за правилом Демпстера.

3.2.3. Навчання

Навчальна вибірка формується експертом на основі візуального аналізу зображення, спираючись, при необхідності, на картографічну, тематичну та іншу доступну йому інформацію. Експерт відбирає компактні групи пікселів (ділянкові навчальні кластери) так, щоб кожна група могла розглядатися як представник (репрезент) відповідного класу із числа присутніх у кадрі. За результатами аналізу кожен піксел навчальної вибірки отримує позначку l, яка характеризує його класову приналежність (згідно із представленням експерта).

На матеріалі виділеної навчальної вибірки формуються описи класів, представлених на зображенні. В основу опису покладені міркування, аналогічні тим, що проводилися у розділі 2 під час опису методу класифікування при наявності забруднених навчальних вибірок.

Будь-який піксел ГКЗ описується сукупністю значень його сигналів на осях робочого спектрального простору R^{K^*} (спектральних осях). Природно припустити, що сукупність сигналів всіх пікселів – репрезентів одного класу в

навчальній вибірці буде займати певний інтервал на кожній із спектральних осей простору R^{K^*}. Будемо виходити із того, що положення цього класового інтервалу на спектральній осі визначається середнім значенням сигналів пікселів – репрезентів класу в цьому спектральному каналі. Якщо, наприклад, в k-му спектральному каналі для класів l та l+1 розраховані середні значення сигналів пікселів зазначених класів у вигляді $\overline{u}_{k,l}$ та $\overline{u}_{k,l+1}$, то вони й визначають положення інтервалів для цих двох класів (п. 2.2.6, рис. 2.5).

Як вже було наведено у розділі 2, класові інтервали межують один з одним на спектральній осі, але не перетинаються, тобто, положення точки $A_{l,l+1}$, що розділяє два вказаних класи на спектральній осі, визначається виходячи із співвідношення

$$\frac{a_l}{a_{l+1}} = \frac{\sigma_{k,l}}{\sigma_{k,l+1}},$$

де $\sigma_{k,l}$ і $\sigma_{k,l+1}$ – середньоквадратичне відхилення сигналів пикселів відповідних класів у k-ому спектральному каналі.

Аналогічно формуються границі та будуються інтервали для всіх інших класів, за винятком двох: класу з мінімальним середнім значенням і класу з максимальним середнім значенням сигналу пікселів у даному спектральному каналі. Для першого з названих класів не встановлюється нижня (ліва на рис. 2.5) гранична межа на спектральній осі, а для другого - верхня (права на рис. 2.5) гранична межа, тобто ці два інтервали є напіввідкритими.

Отже, в результаті клас, представлений на кожній спектральній осі простору R^{K^*} окремим інтервалом, положення якого визначається середнім значенням сигналів пікселів навчальної вибірки, відібраних експертом в якості репрезентів цього класу.

Після побудови класових інтервалів виконується фокусизація. Під фокусизацією класового інтервалу мається на увазі процедура створення списку

фокальних підмножин на множині пікселів, сигнали яких проектуються на цей інтервал, та обчислення для них базових ймовірностей.

Щоб пояснити зміст даної процедури, введемо ще одне визначення: класовий інтервал в *k*-ому спектральному каналі, положення якого визначається величиною $u_{k,l}$, будемо називати *власним інтервалом* класу *l*.

В умовах забруднених навчальних вибірок природно припустити, що в межах власного інтервалу будуть зустрічатися не тільки пікселі однойменного з цим інтервалом класу, але і пікселі інших класів. Тому будемо вважати, що в загальному випадку список фокальних підмножин в межах будь-якого класового інтервалу може включати:

- одноелементну підмножину із класу пікселів власного інтервалу;

- двоелементні підмножини, що породжуються невизначеністю класової приналежності деяких пікселів після формування навчальної вибірки.

Для кожної із фокальних підмножин необхідно розрахувати базову ймовірність. Покажемо, як ці розрахунки виконуються на наступному прикладі. Нехай у деякому власному інтервалі класу l_1 знаходяться, крім Q_1 пікселів, які породжують цей інтервал (тобто віднесених експертом на етапі навчання до однойменного класу), ще Q_2 пікселів, яким експертом раніше була надана мітка належності до класу l_2, і Q_3 пікселів, яким експертом була надана мітка належності до класу l_3 $(l_1, l_2, l_3 \in L)$. При цих умовах список фокальних підмножин на даному класовому інтервалі буде містити три фокальних підмножини: $\{l_1\}$, $\{l_1, l_2\}$, $\{l_1, l_3\}$. Базові ймовірності для цих фокальних підмножин розраховуються за формулами:

$$\left.\begin{aligned} m(\{l_1\}) &= \frac{Q_1}{Q_\Sigma} \\ m(\{l_1, l_2\}) &= \frac{Q_2}{Q_\Sigma} \\ m(\{l_1, l_3\}) &= \frac{Q_3}{Q_\Sigma} \end{aligned}\right\}, \qquad (3.14)$$

де $Q_\Sigma = Q_1 + Q_2 + Q_3$.

Аналогічно виконуються розрахунки при наявності будь-якого іншого списку фокальних підмножин.

Таким чином, за результатами навчання отримуємо базу даних (БД), що містить:

- опис (у вигляді номерів та границь) класових інтервалів у всіх K^* відібраних спектральних каналах;

- фокальні підмножини та їх маси для всіх класових інтервалів кожного з відібраних спектральних інтервалів.

На цьому процес навчання завершується і далі може виконуватися процедура класифікування.

3.2.4. Класифікування

Класифікування пікселів ГКЗ здійснюється послідовно та незалежно, тому достатньо розглянути процедуру класифікування одного піксела. Нехай даний піксел π_n характеризується векторним сигналом $\mathrm{u}_n = \{u_{nk}\}_{k=1}^{K^*}$. Покажемо, що, знаючи цей сигнал та використовуючи сформовану при навчанні БД, можна визначити, до якого із L класів даний піксел найбільш ймовірно належить.

Алгоритм, що реалізує метод класифікування на основі ТСДШ, складається із наступних кроків.

Крок 1. У кожному *k*-ому спектральному каналі $\left(k = 1,2,...K^{*}\right)$ позначається номер (клас) інтервалу, на який проектується відповідна компонента векторного сигналу u_{nk} піксела π_n.

Крок 2. З БД вилучається список фокальних підмножин та їх маси для зазначеного в кожному із спектральних каналів класового інтервалу.

Крок 3. Виконується об'єднання мас фокальних підмножин із вилученого списку із використанням комбінаційного правила Демпстера і отримуються оцінки мас цих підмножин із врахуванням всіх спектральних каналів.

Крок 4. Розраховуються за допомогою формули (3.2) пігністичні ймовірності *BetP* розглянутих фокальних підмножин.

Крок 5. Розраховані значення пігністичних ймовірностей фокальних підмножин проранжовуються, і за критерієм максимуму пігністичної ймовірності (3.3) визначається найбільш ймовірний клас піксела π_n.

Даний алгоритм застосовується послідовно до кожного піксела і результатом є відкласифіковане в цілому ГКЗ. З викладеного випливає, що обчислювальна складність процедур навчання та класифікування мінімальна, оскільки обидві реалізуються за допомогою простих логіко-арифметичних операцій. Ще одна важлива особливість: рішення щодо класової приналежності піксела приймається в автоматичному режимі, без залучення експерту.

3.3. Приклад застосування запропонованого методу класифікування

Розглянемо ілюстративний приклад застосування запропонованого в даному розділі методу класифікування. Припустимо, сигнал деякого піксела π_n у спектральному каналі S1 спроектувався на інтервал із трьома фокальними підмножинами: $\{l_1\}$, $\{l_1, l_2\}$, $\{l_1, l_3\}$, а в спектральному каналі S2 – на інтервал

із такими фокальними підмножинами: $\{l_2\}$, $\{l_1, l_2\}$, $\{l_2, l_3\}$. Маси вказаних фокальних підмножин, взяті з БД, наведені в табл. 3.2.

Таблиця 3.2 – Вихідні дані до прикладу

Спектральний канал S1	$m_1(\{l_1\}) = 0.56;$ $m_1(\{l_1, l_2\}) = 0.30;$ $m_1(\{l_1, l_3\}) = 0.14$
Спектральний канал S2	$m_2(\{l_2\}) = 0.65;$ $m_2(\{l_1, l_2\}) = 0.25;$ $m_2(\{l_2, l_3\}) = 0.10$

Треба визначити клас, до якого найбільш ймовірно належить піксел π_n.

Розв’язок задачі розпочинаємо із розрахунку об'єднаних мас фокальних підмножин, вилучених із БД, використовуючи для цього правило Демпстера. Всі можливі перетини фокальних елементів відобразимо за допомогою табл. 3.3. У клітках таблиці показані непусті та пусті перетини підмножин, а також їх перемножені ймовірності. Використовуючи ці дані, розраховуємо коефіцієнт конфліктності:

$$K = m_1(\{l_1\}) \cdot m_2(\{l_2\}) + m_1(\{l_1\}) \cdot m_2(\{l_2, l_3\}) + m_1(\{l_1, l_3\}) \cdot m_2(\{l_2\}) = 0{,}364 + 0{,}056 + 0{,}091 = 0{,}511,$$

а потім – об’єднані маси:

$$m_{12}(\{l_1\}) = 0{,}358; \quad m_{12}(\{l_2\}) = 0{,}460; \quad m_{12}(\{l_3\}) = 0{,}029; \quad m_{12}(\{l_1, l_2\}) = 0{,}153.$$

Таблиця 3.3 – До розрахунку об’єднаних мас фокальних підмножин

		Спектральний канал S2		
		$m_2(\{l_2\}) = 0{,}65$	$m_2(\{l_1, l_2\}) = 0{,}25$	$m_2(\{l_2, l_3\}) = 0{,}10$
Спектральний канал S1	$m_1(\{l_1\}) = 0{,}56$	0,364 $(\varnothing)$	0,14 $(\{l_1\})$	0,056 $(\varnothing)$
	$m_1(\{l_1, l_2\}) = 0{,}30$	0,195 $(\{l_2\})$	0,075 $(\{l_1, l_2\})$	0,030 $(\{l_2\})$
	$m_1(\{l_1, l_3\}) = 0{,}14$	0,091 $(\varnothing)$	0,035 $(\{l_1\})$	0,014 $(\{l_3\})$

Далі визначаємо пігністичні ймовірності (формула (3.2)):

$BetP(l_1) = 0{,}4345; \quad BetP(l_2) = 0{,}5365; \quad BetP(l_3) = 0{,}029.$

Зіставлення отриманих значень пігністичних імовірностей показує, що згідно критерію (3.3) найбільш імовірним для розглянутого піксела π_n буде клас l_2.

Завершуючи опис запропонованого методу класифікування на основі ТСДШ, необхідно вказати, що при наявності референсних (завіркових) даних повинна бути оцінена його точність. Для цього за результатами класифікування всіх пікселів ГКЗ будується матриця помилок, а потім на її основі обчислюються оцінки досягнутої точності. Якщо отримані оцінки задовольняють заданим вимогам, процедура класифікування вважається успішно виконаною. В іншому випадку можуть вживатися додаткові заходи, щоб підвищити точність класифікування (як-внесення змін у процедуру відбору зональних зображень, поповнення чи коригування навчальної вибірки, деталізація описів класів та ін).

ВИСНОВКИ ДО РОЗДІЛУ 3

1. У розділі запропоновано новий метод контрольованого класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та обмежень на кількість елементів у виборці. Наводяться обмеження, припущення та визначення, на яких побудовано метод. Пропонований метод враховує невизначеності, що можуть виникати при навчанні та класифікуванні, що досягається використанням математичного апарату та прийомів теорії свідчень Демпстера-Шейфера.

2. Детально описано організацію метода і основні процедури, з яких він складається, у тому числі процедуру відбору зональних зображень і схему класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень.

3. Запропоновано новий спосіб відбору зональних зображень і формування репрезентативної комбінації зональних зображень при класифікуванні гіперспектрального космічного знімка, який дозволяє об’єктивізувати і прискорити процес вирішення задачі за рахунок зменшення вимог до необхідного обчислювального ресурсу. Дана перевага досягається за рахунок автоматизації процедури оцінювання рівню шумів кожного з вихідних зональних зображень і відсіювання з подальшого аналізу зображень, які мають високий рівень зашумленості.

Показано, що введення операції відбору зашумлених зональних зображень у випадку знімків EO-1 / Hyperion дозволяє зменшити необхідні обчислювальні витрати в 2,78-4,64 рази, а у випадку знімків AVIRIS – в 2,92-3,64 рази.

4. Наведено приклад застосування запропонованого методу класифікування, який ілюструє простоту методу та полегшує його імплементацію.

РОЗДІЛ 4 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТИ РОЗРОБЛЕНИХ МЕТОДІВ КОНТРОЛЬОВАНОГО КЛАСИФІКУВАННЯ ГІПЕРСПЕКТРАЛЬНИХ КОСМІЧНИХ ЗОБРАЖЕНЬ

Розділ містить матеріали експериментальних досліджень ефективности розроблених методів контрольованого класифікування на основі використання реальних ГКЗ, отриманих апаратурою Hyperion супутника EO-1.

Розроблені і стисло описані архітектура та функціональні особливості спеціалізованого програмного забезпечення (СПЗ) для реалізації та дослідження розроблених методів контрольованого класифікування.

За допомогою зазначеного СПЗ проведено оцінку ефективності методу контрольованого класифікування ГКЗ в умовах забруднених навчальних вибірок та оцінку ефективності методу контрольованого класифікування ГКЗ в умовах забруднення навчальних вибірок та кількісних обмежень.

Наводяться результати оцінювання точності класифікування, яка досягається в реальних умовах. При цьому результати класифікування запропонованих методів порівнюються із результатами класифікування інших відомих методів, таких як метод SVM та алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками. Оцінка точності класифікування здійснюється за допомогою таких показників: повна точність класифікування, точність користувача, точність виробника, помилки пропуску класу, помилки ототожнення з класом. Результати експериментальних досліджень свідчать про досить високу точність запропонованих методів класифікування.

4.1. Спеціалізоване програмне забезпечення для реалізації та дослідження розроблених методів класифікування: архітектура і стислий опис

Для реалізації та експериментального дослідження ефективності запропонованих в дисертації нових методів контрольованого класифікування

ГКЗ в умовах навчальних вибірок обмеженої якості було розроблено СПЗ, яке далі йменується DST. СПЗ DST було реалізовано у середовищі пакету Visual Basic 6.0.

Архітектура СПЗ DST передбачає багатовіконний MDI інтерфейс; в якості його батьківського вікна служить головний модуль програми, здійснюючий загальне керування і забезпечуючий доступ до чотирьох розрахункових, двох допоміжних і двох інформаційних модулів (рис. 4.1).

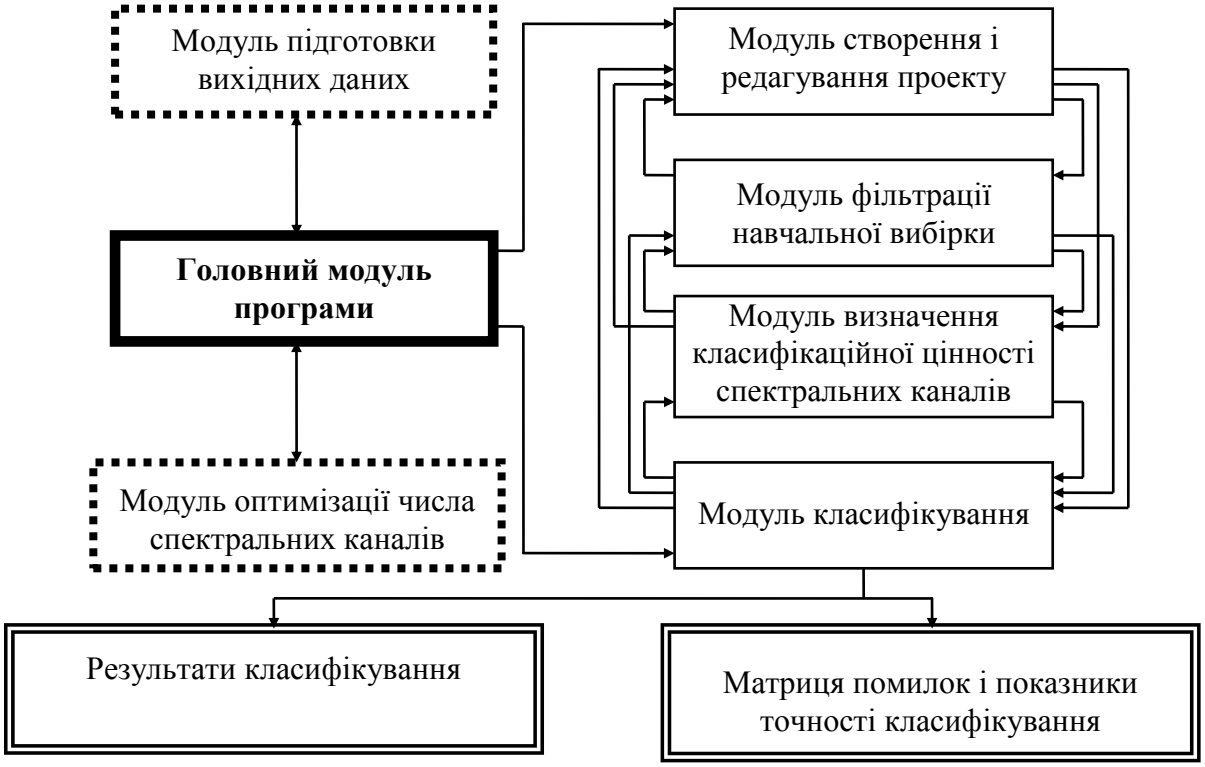


Рис. 4.1. Архітектура СПЗ DST. Суцільною жирною рамкою позначений головний модуль, суцільною тонкою рамкою - розрахункові, пунктирною рамкою - допоміжний, подвійною рамкою - інформаційні модулі

Кожен розрахунковий модуль програми підтримує відповідні логіко-обчислювальні процедури, необхідні для реалізації розроблених методів; на додаткові модулі покладено функції підготовки вихідного розрахункового зображення (його конвертування у внутрішній формат програми та відбір розрахункових спектральних каналів); інформаційні модулі забезпечують

виведення і збереження остаточних результатів класифікування. Зв'язок між розрахунковими модулями обмежується лише ступенем наявності вихідних даних (не менше двох спектральних зон і навчальних вибірок, кожна з яких визначає різні класи) і може здійснюватися з будь-якого місця програми.

На вхід програми DST надходять сигнали пікселів у представленні, яке має сприйматися системою оброблення зображень ENVI. Кінцевим результатом виконання програми DST є по-піксельне класифікування ГКЗ (або його деяких визначених ділянок), побудова матриці помилок і розрахунки показників точності класифікування. Результати класифікування представляються на екрані у вигляді відповідних растрових зображень і таблиць, а також можуть бути збережені в загальнодоступних файлах, список яких визначається користувачем на кінцевій стадії класифікування і знаходиться у відповідному робочому протоколі, який зберігається разом із отриманими результатами.

Розміщення всієї вихідної, розрахункової і проміжної інформації знаходиться в папках, які внесені в системний регістр Windows і визначаються користувачем в процесі роботи з програмою.

Процес виконання роботи відбувається в інтерактивному режимі, для загального випадку пропонується наступний порядок дій користувача:

Етап 1 - Підготовка вихідних даних і відбір розрахункових спектральних каналів;

Етап 2 - Створення робочого проекту у вигляді внутрішнього текстового файлу, який включає список розрахункових спектральних каналів і параметри навчальних вибірок виділених класів, і формування БЗ;

Етап 3 - Вирішення задачі класифікування, виходячи з конкретної постановки.

Стисло наведемо особливості програмного забезпечення означених етапів.

Етап 1. Підготовка даних і процедура відбору зональних зображень реалізовані в середовищі програми DST у вигляді двох допоміжних модулів -

підготовки вихідних даних і оптимізації числа спектральних каналів. Наявність модуля підготовки вихідних даних пояснюється тим, що для різних бортових сенсорів типи представлення піксельних сигналів, що конвертуються системою ENVI у її внутрішні формати, можуть відрізнятися один від одного і визначатися як byte, integer, single та ін. Цей факт створює значну незручність з точки зору створення компактного програмного коду, оскільки особливості VB6 (як мови високого рівня) вимагають його багаторазового дублювання для всіх можливих типів даних і створення відповідного числа умовних переходів в різних частинах програми при зчитуванні кожного типу з файлів системи ENVI.

Основний розрахунковий алгоритм модуля підготовки вихідних даних побудований таким чином, що після розпізнавання заголовка файлу вихідних даних у внутрішньому форматі системи ENVI та підтвердження на подальші дії виконується зчитування значень піксельних сигналів кожного зонального зображення з одночасним визначенням значень їх мінімуму, максимуму та середнього. Після цього, у разі наявності позитивного градієнта сигналів, вся отримана інформація записується для кожного спектрального каналу у відповідні файли довільного доступу, а їхні імена вносяться до єдиної бібліотеки каналів, яка у подальшому може багаторазово використовуватися.

Такий підхід дозволяє уникнути виконання цілого ряду проміжних обчислювальних операцій і відразу ж виключає канали з нульовими градієнтами, забезпечуючи, таким чином, виконання кроку 1 процедури відбору зональних зображень.

Модуль оптимізації числа спектральних каналів надає можливість зниження ступеня надмірності вихідного ГКЗ за рахунок скорочення числа зональних зображень зі значною взаємної кореляцією. Інтерфейс модуля (рис. 4.2) дозволяє початкове проведення візуального відбору явно непридатних спектральних зображень, визначення порогового значення коефіцієнта кореляції і мінімального числа необхідних спектральних каналів, а також -

запуск процедури їх подальшого відбору як в автоматичному, так і в інтерактивному режимах.

У лівій нижній частині інтерфейсу модуля показані параметри оптимізації відбору: значення коефіцієнта максимальної кореляції і мінімальне число відібраних каналів. При цьому при відборі зональних зображень після кожного виконання кроку 5б або 5в (див. п. 2.2.2.1) передбачається збереження зональних зображень в єдиних бібліотеках спектральних каналів, які можуть використовуватися при вирішенні задач класифікування з іншим числом каналів.

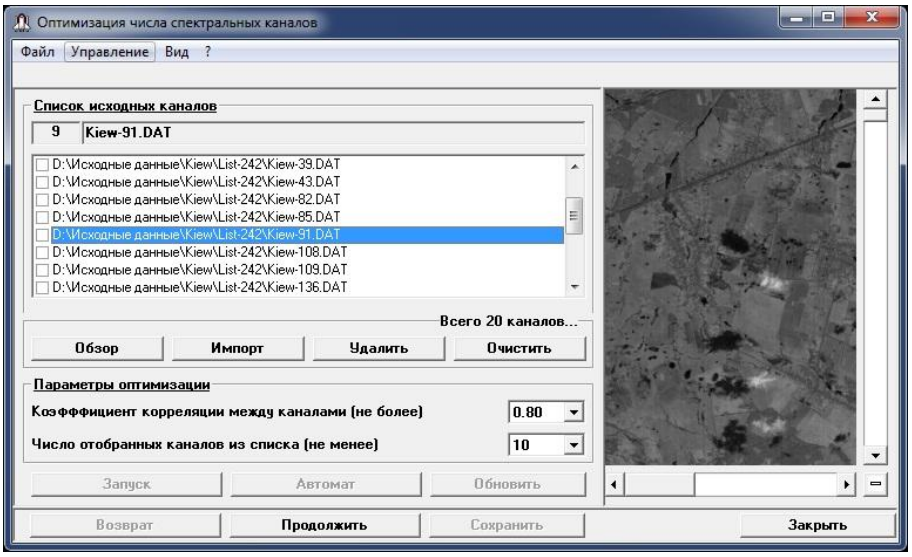


Рис. 4.2. Інтерфейс модуля оптимізації числа спектральних каналів

<u>Етап 2.</u> Під робочим проектом розуміється сукупність набору вихідних даних у вигляді списку розрахункових спектральних каналів (із зазначенням їх розташування на диску), властивостей пікселів навчальних вибірок (із зазначенням їх приналежності до виділених класів і координат на зображенні), а також - базу знань. Цілком очевидно, що створення останньої забезпечується мінімальним набором вихідних даних і виконується із застосуванням вищевказаних алгоритмів розбиття простору ознак, внутрішньоінтервальної оцінки для виділених класів та розрахунку класифікаційної цінності спектральних каналів. Такий підхід дозволяє багаторазове вирішення задачі

класифікування з залученням одних і тих же вихідних даних у різних постановках, що передбачено архітектурою програми. Більшість з перерахованих вище операцій по створенню (і зміні) робочого проекту виконуються у відповідному розрахунковому блоці модуля створення і редагування проекту. Можливості цього модуля (рис. 4.3) дозволяють здійснювати:

1. Створення та редагування списку розрахункових каналів вихідного зображення, включаючи його імпорт і експорт в середовище програмної системи ENVI.

2. Створення та редагування списку виділених класів, включаючи їх імпорт і експорт в середовище системи ENVI у вигляді сумісних областей інтересу (ROIs).

3. Оконтурювання границь навчальних вибірок, що представляють виділені класи у вигляді частин зображення довільних форм, обраних графічними інструментами модуля. Останні дозволяють визначати будь-яку комбінацію навчальних вибірок у вигляді областей інтересу ROIs, виключають дублювання приналежності одних і тих же їх пікселів до різних класів і можуть передавати їх в систему ENVI у вигляді сумісних файлів. Слід зауважити, що, крім можливостей інструментів модуля, навчальні вибірки можуть визначатися в допоміжному вікні підкладки, яка представляє собою графічне зображення ділянки класифікування у вигляді різного роду тематичних карт або схем.

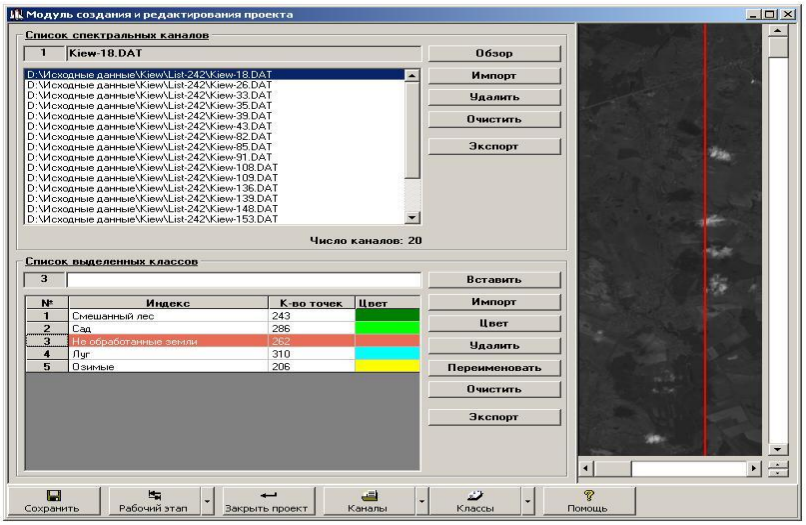


Рис. 4.3. Інтерфейс модуля створення і редагування проекту

4. Виконання процедур розбиття простору ознак, внутрішньоінтервальной оцінки для виділених класів та розрахунку класифікаційної цінності обраних спектральних каналів з їх внесенням (як бази знань) до складу робочого проекту.

5. Збереження робочого проекту у вигляді внутрішнього текстового файлу (“*.PRJ”), з можливістю його багаторазового подальшого використання, редагування і передавання в усі розрахункові модулі програми.

Модуль фільтрації навчальної вибірки (рис. 4.4) служить для підвищення чистоти навчальної вибірки, визначеної в режимі роботи з модулем створення і редагування проекту. В модулі використовується алгоритм, наведений в п. 2.2.2.5 (див. ще схему фільтрації навчальної вибірки на рис. 2.4).

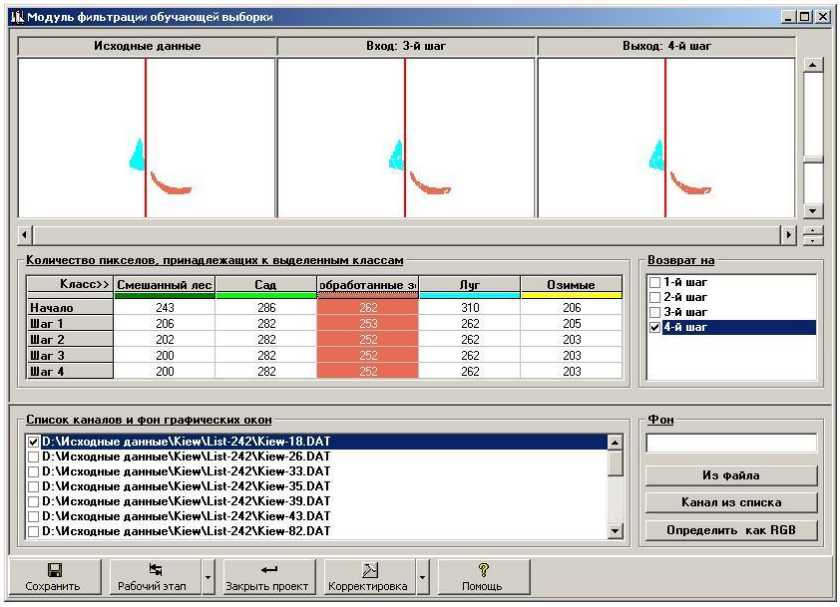


Рис. 4.4. Інтерфейс модуля фільтрації навчальної вибірки

Функціональні можливості даного модуля також передбачають:

- вивід результатів фільтрації на кожному кроці виконання розрахункового алгоритму в графічному або табличному вигляді;

- збереження результатів фільтрації після кожного кроку виконання розрахункового алгоритму у вигляді файлів, які можуть бути використані як самостійні робочі проекти програми DST;

- роботу в інтерактивному і автоматичному режимах, здійснення постійного контролю результатів фільтрації з видачею відповідних інформаційних та рекомендаційних повідомлень при виникненні непередбачених ситуацій (видалення всіх пікселів навчальної підвибірки того чи іншого класу і т. п.).

Модуль визначення класифікаційної цінності спектральних каналів (рис. 4.5) забезпечує обчислення значень функції *F* для кожного каналу і ранжування каналів за цими значеннями. Важливо, що модуль надає можливість візуального відображення всієї інформації, що міститься в БЗ (з урахуванням наявних на момент завантаження вихідних даних), проаналізувавши яку дослідник може прийняти остаточне рішення щодо проведення класифікування або внесення змін в активний проект.

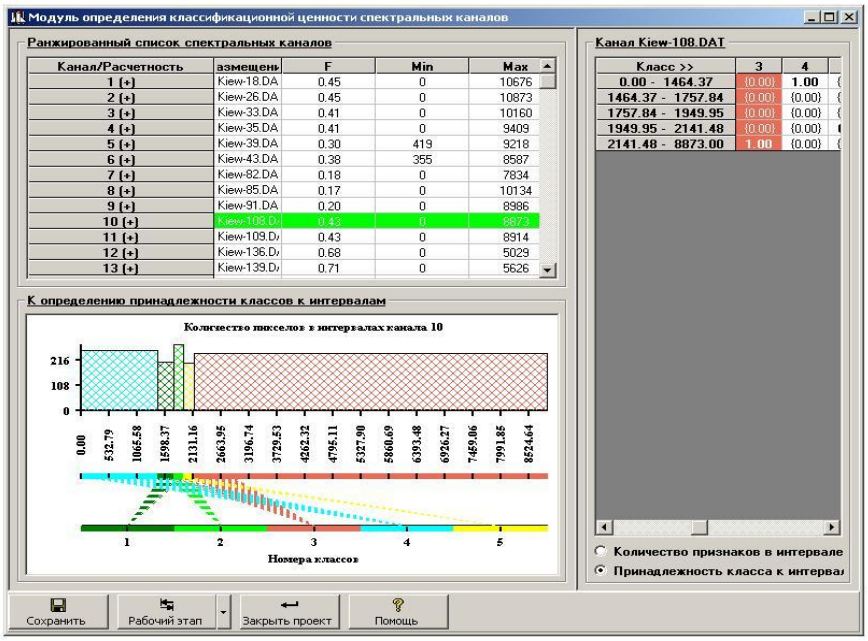


Рис. 4.5. Інтерфейс модуля визначення класифікаційної цінності спектральних каналів

<u>Етап 3.</u> Програмна реалізація етапу проводиться в модулі класифікування (рис. 4.6) і зводиться до остаточного визначення установочних параметрів у вирішальному правилі, на якому базується процедура класіфікування, а саме:

- визначення остаточного списку розрахункових спектральних каналів;
- визначення просторових границь ділянки (ділянок) в межах зображення, яка підлягає (які підлягають) класифікуванню;
- вибір форми відображення результатів по-піксельного класифікування і автоматичне завантаження в інформаційний модуль “Результати класифікування” (рис. 4.7);
- вибір референсних (завіркових) даних для побудови матриці помилок та їх автоматичне завантаження в інформаційний модуль “Матриця помилок і показники точності класифікування” (рис. 4.8);
- запуск процедури класифікування.

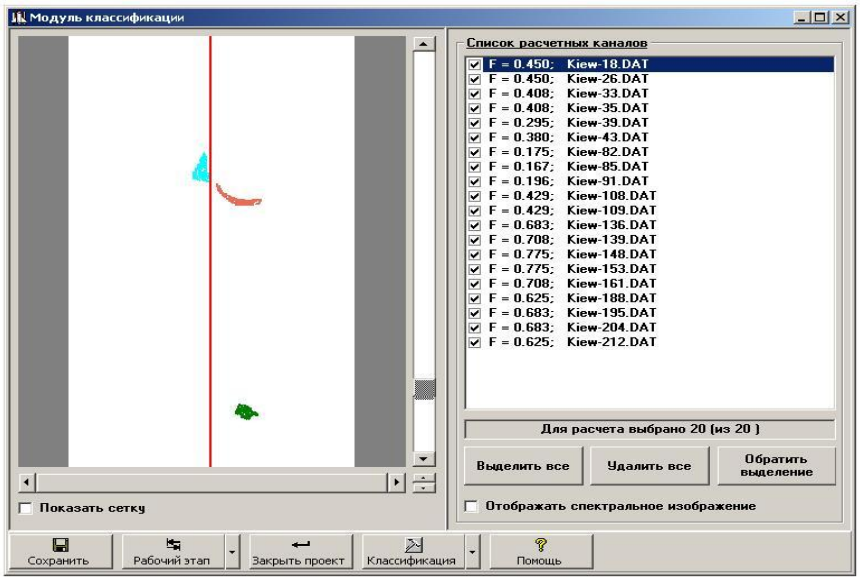


Рис. 4.6. Інтерфейс модуля класифікування

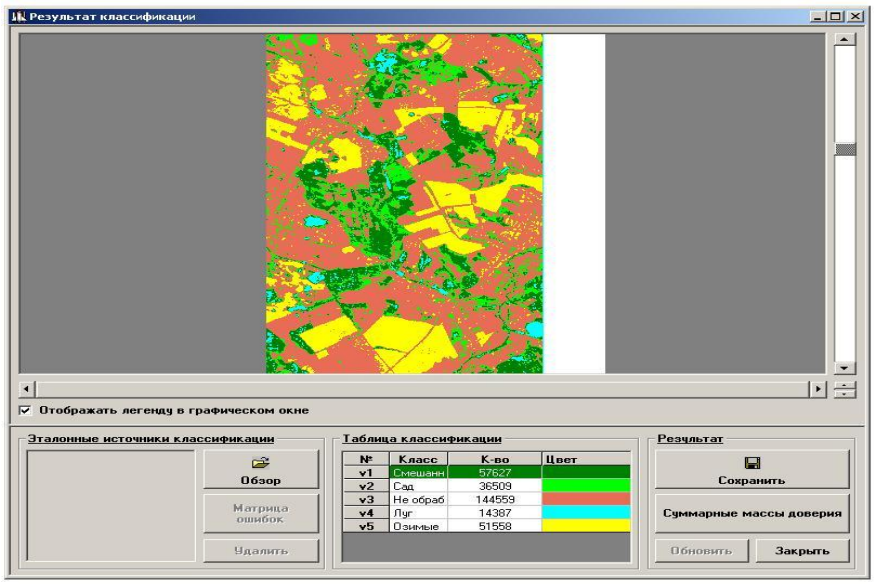
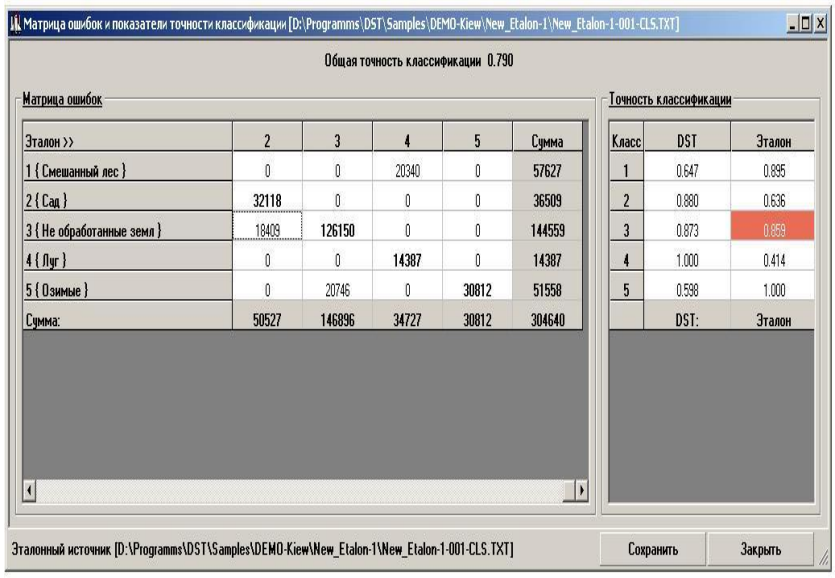


Рис. 4.7. Інтерфейс модуля “Результати класифікування”



Эталон >>	2	3	4	5	Сумма
1 { Смешанный лес }	0	0	20340	0	57627
2 { Сад }	32118	0	0	0	36509
3 { Не обработанные земл }	18409	126150	0	0	144559
4 { Луг }	0	0	14387	0	14387
5 { Озимые }	0	20746	0	30812	51558
Сумма:	50527	146896	34727	30812	304640

Класс	DST	Эталон
1	0.647	0.895
2	0.880	0.636
3	0.873	0.959
4	1.000	0.414
5	0.598	1.000
	DST:	Эталон

Рис. 4.8. Інтерфейс модуля “Матриця помилок і показники точності класифікування”

Виконання останнього пункту, після попереднього запиту програми про правильність вибору установочних параметрів, призводить до ініціалізації основної процедури класифікування (згідно пропонованого методу) і подальшому завантаженню отриманих результатів у відповідні інформаційні модулі програми.

Необхідно відзначити, що функціональні можливості останніх передбачають збереження (як окремо, так і спільно) отриманих результатів класифікування (з одночасним створенням відповідного протоколу) в папках, які визначаються користувачем-дослідником на цьому етапі роботи.

Розроблене і описане вище спеціалізоване програмне забезпечення DST дозволило поставити та провести детальні експериментальні дослідження ефективності запропонованих в дисертаціх двох нових методів контрольованого класифікування ГКЗ.

4.2. Оцінка ефективності методу контрольованого класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок

В дослідженні були поставлені та вирішувались такі задачі:

1. Оцінити точність запропонованого у розділі 2 методу контрольованого класифікування;

2. Дослідити вплив процедур фільтрації навчальної вибірки і відбору певних спектральних каналів шляхом ранжування за їх класифікаційною цінністю на точність класифікування.

У дослідженні використовувалося гіперспектральне зображення EO1H1810252013112110KF (рис. 4.9, а), отримане у весняний період (22.04.2013 р.) бортовою знімальною апаратурою Hyperion супутникової системи EO-1.

Знята сцена являє собою смугу території в межах Києво-Святошинського і Макарівського районів Київської області.

Після конвертації обраного ГКЗ програмною системою ENVI у внутрішній формат представлення даних і створення опису цього формату у вигляді заголовка був отриманий єдиний розрахунковий файл (з розширенням "*.BSQ"), в якому всі канали знімка були представлені у вигляді послідовності окремих зображень.

В процесі оброблення отриманого файлу засобами допоміжного модуля підготовки вихідних даних і виключення на цій стадії 87-ми зональних зображень з нульовими градієнтами піксельних сигналів і з вочевидь забрудненою структурою була створена бібліотека каналів, що включає в себе 155 внутрішніх файлів даних, записаних у внутрішньому представленні програми DST.

Після цього остаточний відбір розрахункових зональних зображень здійснювався з використанням допоміжного модуля оптимізації числа спектральних каналів. В результаті, після їх 4-х кратного попарно-лінійного

проріджування (за наведеним у п. 2.2.2.1 алгоритмом) було отримано список з 20-ти розрахункових спектральних каналів, для яких порогові коефіцієнти кореляції між суміжними діапазонами становили не більш 0,8.

Всі подальші результати і одержані оцінки базуються на використанні саме цих 20-ти одержаних зональних зображень.

На заснятій території домінують 5 основних класів площинних об'єктів, а саме: листяний ліс (1), сади (2), необроблені землі (3), луки (4) і поля (5).

Для кожного з цих класів в межах кадру ГКЗ за допомогою розрахункового модуля створення і редагування проекту у взаємодії з експертом виділялися навчальні підвибірки з відповідних пікселів.

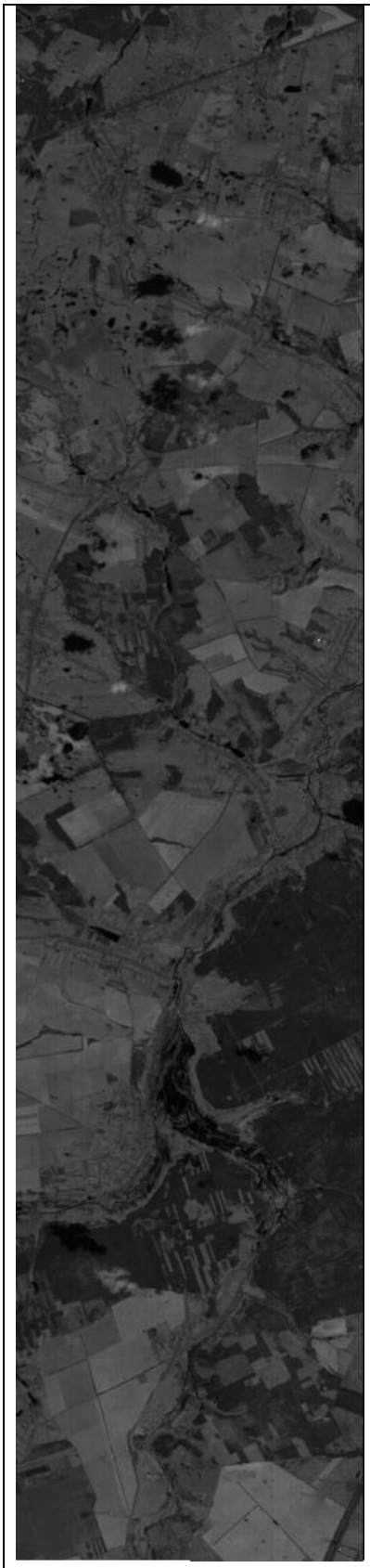
Результати класифікування обраного ГКЗ запропонованим у розділі 2 методом контрольованого класифікування представлені на рис. 4.9, б.

Для оцінювання точності класифікування використовувались тестові ділянки, які не перетинаються з ділянками з пікселів навчальної вибірки. Загальна точність (Overall Accuracy) класифікування тестових ділянок досліджуваним методом наведена в табл. 4.1.

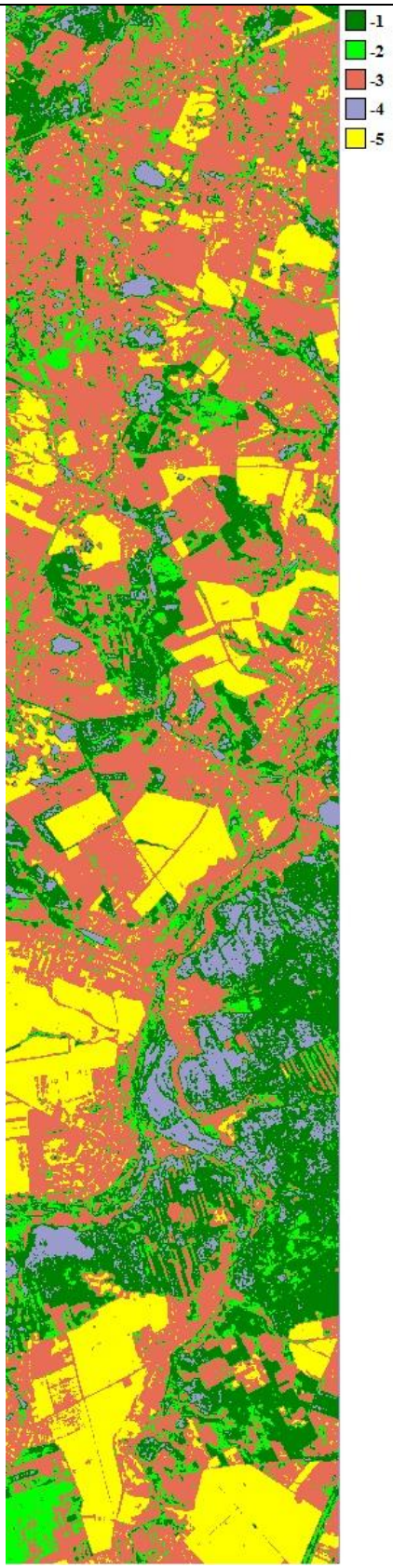
Для порівняння ті ж самі ділянки було відкласифіковано за допомогою ще двох алгоритмів – SVM і алгоритма класифікування об’єктів за їх спектрально-топологічними характеристиками.

Як вже зазначалось в розділі 1, метод SVM відомий своєю ефективністю в задачах класифікування гіперспектральних зображень.

Крім того, зручно, що алгоритм SVM реалізований у програмній системі обробки зображень ENVI.



а



б

Рис. 4.9: а – гіперспектральне зображення Hyperion EO1H1810252013112110KF (22.04.2013 р.); б – класифікаційна карта (просторовий розподіл класів)

Точностні результати класифікування представлено в табл. 4.1.

Таблиця 4.1– Точність класифікування ГКЗ різними методами

	Запропонований метод на основі ТСДШ	Метод SVM	Алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками
Точність (Overall Accuracy)	0,802	0,757	0,734

Порівняльний аналіз одержаних оцінок точності показує перевагу запропонованого методу класифікування перед іншими методами.

Застосування процедури фільтрації навчальної вибірки показало, що у середньому при цьому відсікається до 18% пікселів, а найбільші проблеми виникають з чистотою таких репрезентів, як луки (вихідна навчальна підвибірка зменшується на 33%) і листяний ліс (зменшення на 12%).

Проведене дослідження ефективності включення процедури відбору певних спектральних каналів шляхом ранжування за їх класифікаційною цінністю в загальну схему класифікування показало, що якщо класифікаційна цінність спектральних каналів становить $R \geq 0,7$, то для класифікування звичайно достатньо не більш 7-ми каналів (зональних зображень).

Подальше дослідження запропонованого методу було проведено за умовою, що на гіперспектральному зображенні EO1H1810252013112110KF розглядається 6 класів площинних об'єктів, а саме: листяний ліс (1), сади (2), необроблені землі (3), луки (4), злакові поля (5) і овочеві поля (6).

Для більшої впевненості у правильному визначенні виділених навчальних вибірок, піксели, що їм належать, пройшли процедуру фільтрації. При цьому, в силу різних обставин, пов’язаних із характером виділення навчальних вибірок, їх число після виконання процедури фільтрації зменшилося для виділених

класів, а незначний ступінь цього зменшення, в свою чергу, дає підставу про правильність визначення їх границь (табл. 4.2).

Таблиця 4.2 – Кількість пікселів, що належать до навчальних вибірок виділених класів до та після проведення коригування, а саме фільтрації

Клас	До коригування	Після коригування
1	231	225
2	255	250
3	248	245
4	350	341
5	303	303
6	335	335
Всього	1722	1699

Визначення класової належності пікселів вихідного зображення виконувалось за допомогою спеціального модуля класифікування об’єктів, функціональні можливості якого передбачають визначення пріорітетів (рангів) спектральних каналів у відповідності із класифікаційною цінністю R_k кожного зонального зображення k.

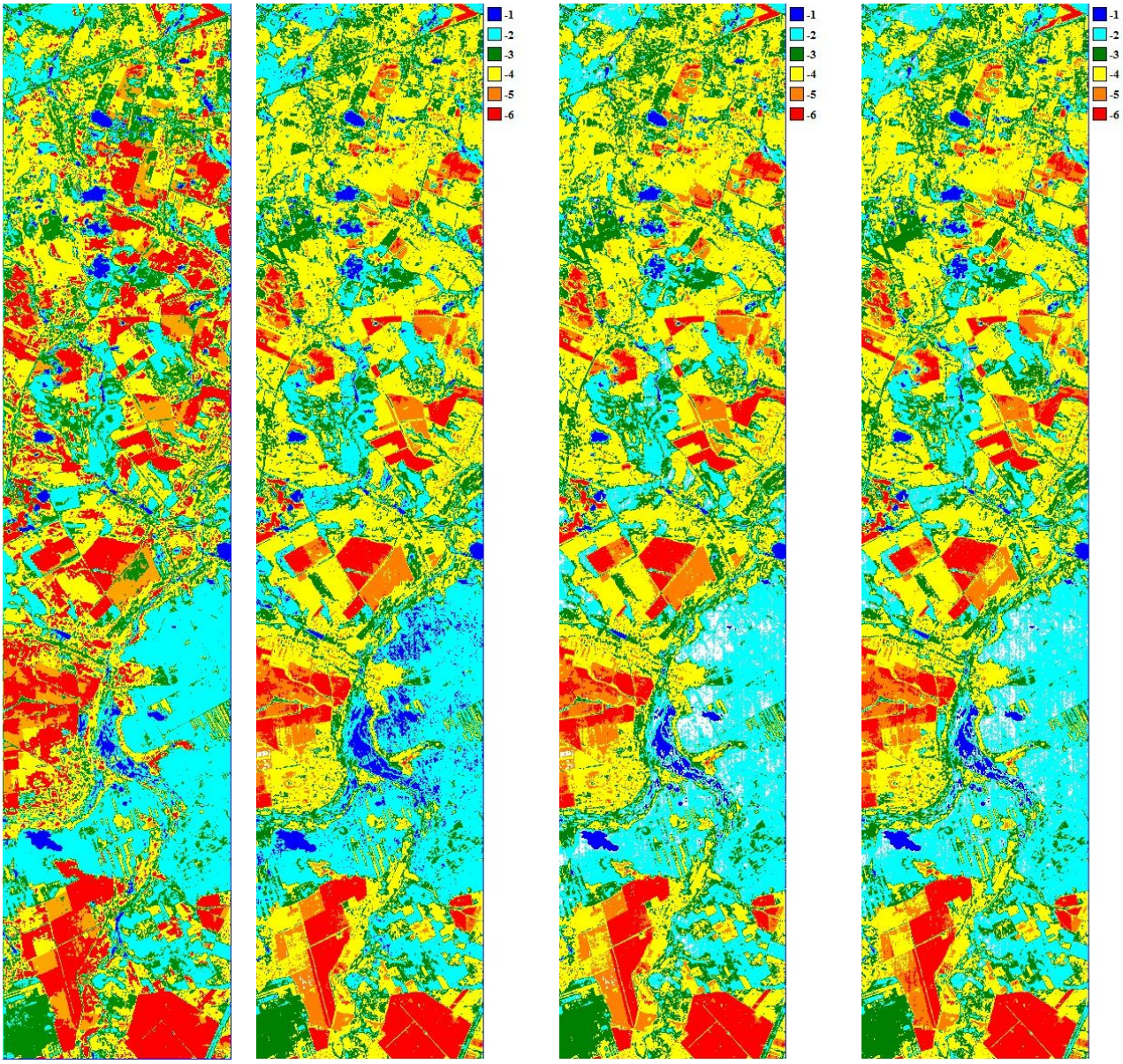
Слід зазначити, що виконання процедури класифікування проводилося для трьох варіантів розрахунку:

Варіант I – з використанням 20-ти відібраних каналів та нефільтрованих пікселів навчальних вибірок;

Варіант II – з використанням 20-ти відібраних каналів та фільтрованих пікселів навчальних вибірок;

Варіант III – класифікування з використанням 7-ми найбільш інформативних каналів, для яких класифікаційна цінність складає $R \geq 0{,}7$ та фільтрованих пікселів навчальних вибірок.

Паралельно із визначенням класової належності пікселів вихідного зображення проводилася побудова матриць помилок та розрахунок коефіцієнтів точності у порівнянні із еталонним класифікатором. У якості еталонного класифікатора використовувався метод SVM. Результати класифікування представлені на рис. 4.10 та в табл. 4.3.



a) – SVM/ENVI *b*) – варіант I *c*) – варіант II *d*) – варіант III

Рис. 4.10. Результати класифікування ГКЗ EO1H1810252013112110KF методом SVM / ENVI (a) і за допомогою запропонованого методу (програма DST, три варіанти розрахунку *(b, c, d)*

Таблиця 4.3 – Матриці помилок, отримані в результаті порівняння еталонних (SVM) та розрахункових (DST) класів для 3-х варіантів класифікування

		SVM						
		1	2	3	4	5	6	Сума
Варіант I								
DST	1	6366	5914	0	0	0	0	12280
	2	1547	59016	7760	1186	0	7	69516
	3	348	6788	34726	16820	0	719	59401
	4	0	568	16223	70919	3139	23614	114463
	5	0	12	841	857	11119	10824	23653
	6	0	1	38	80	1876	22831	24826
Сума		8261	72299	59588	89862	16134	57995	304139
Варіант II								
DST	1	6009	576	0	0	0	0	6585
	2	1383	54552	6858	1013	0	5	63811
	3	388	7384	36627	17455	0	675	62529
	4	1	621	15098	70196	2792	23074	111782
	5	0	14	933	1091	11448	11411	24897
	6	0	1	38	80	1883	22830	24832
Сума		7781	63148	59554	89835	16123	57995	294436
Варіант III								
DST	1	6268	559	0	0	0	0	6827
	2	1686	60722	7906	646	0	14	70974
	3	0	3191	40526	17620	3	567	61907
	4	0	184	10644	71062	4398	23872	110160
	5	0	2	274	484	10390	11088	22238
	6	0	0	25	75	1366	22893	24359
Сума		7954	64658	59375	89887	16157	58434	296465

Базуючись на даних матриці помилок (табл. 4.3), обчислимо коефіцієнти точності для кожного з зазначених трьох варіантів.

1) Коефіцієнти точності для І-го варіанта класифікування:

$$n = 304139;\ \sum_{i=1}^{k} n_{ii} = 204977;\ OA = \frac{\sum_{i=1}^{K} n_{ii}}{n} = \frac{204977}{304139} = 0{,}674 -$$ повна точність класифікування;

Точності виробника:

$$PA_1 = \frac{n_{11}}{n_{+1}} = \frac{6366}{8261} = 0{,}771;$$

$$PA_2 = \frac{n_{22}}{n_{+2}} = \frac{59016}{72299} = 0{,}816;$$

$$PA_3 = \frac{n_{33}}{n_{+3}} = \frac{34726}{59588} = 0{,}583;$$

$$PA_4 = \frac{n_{44}}{n_{+4}} = \frac{70919}{89862} = 0{,}789;$$

$$PA_5 = \frac{n_{55}}{n_{+5}} = \frac{11119}{16134} = 0{,}689;$$

$$PA_6 = \frac{n_{66}}{n_{+6}} = \frac{22831}{57995} = 0{,}394;$$

Точності користувача:

$$UA_1 = \frac{n_{11}}{n_{1+}} = \frac{6366}{12280} = 0{,}518;$$

$$UA_2 = \frac{n_{22}}{n_{2+}} = \frac{59016}{69516} = 0{,}849;$$

$$UA_3 = \frac{n_{33}}{n_{3+}} = \frac{34726}{59401} = 0{,}585;$$

$$UA_4 = \frac{n_{44}}{n_{4+}} = \frac{70919}{114463} = 0{,}62;$$

$$UA_5 = \frac{n_{55}}{n_{5+}} = \frac{11119}{23653} = 0{,}47;$$

$$UA_6 = \frac{n_{66}}{n_{6+}} = \frac{22831}{24826} = 0{,}92;$$

Капа-індекс: $$KI = \frac{n\sum_{i=1}^{K} n_{ii} - \sum_{i=1}^{K} n_{i+} n_{+i}}{n^2 - \sum_{i=1}^{K} n_{i+} n_{+i}} = 0{,}5795261 \approx 0{,}6.$$

2) Коефіцієнти точності для ІІ-го варіанта класифікування:

$$n = 294436;\ \sum_{i=1}^{k} n_{ii} = 201662;\ OA = \frac{\sum_{i=1}^{K} n_{ii}}{n} = \frac{201662}{294436} = 0{,}685 -$$ повна точність класифікування;

Точності виробника:

$$PA_1 = \frac{n_{11}}{n_{+1}} = \frac{6009}{7781} = 0{,}772;$$

Точності користувача:

$$UA_1 = \frac{n_{11}}{n_{1+}} = \frac{6009}{6585} = 0{,}913;$$

$$PA_2 = \frac{n_{22}}{n_{+2}} = \frac{54552}{63148} = 0{,}864; \qquad UA_2 = \frac{n_{22}}{n_{2+}} = \frac{54552}{63811} = 0{,}855;$$

$$PA_3 = \frac{n_{33}}{n_{+3}} = \frac{36627}{59554} = 0{,}615; \qquad UA_3 = \frac{n_{33}}{n_{3+}} = \frac{36627}{62529} = 0{,}586;$$

$$PA_4 = \frac{n_{44}}{n_{+4}} = \frac{70196}{89835} = 0{,}781; \qquad UA_4 = \frac{n_{44}}{n_{4+}} = \frac{70196}{111782} = 0{,}628;$$

$$PA_5 = \frac{n_{55}}{n_{+5}} = \frac{11448}{16123} = 0{,}71; \qquad UA_5 = \frac{n_{55}}{n_{5+}} = \frac{11448}{24897} = 0{,}46;$$

$$PA_6 = \frac{n_{66}}{n_{+6}} = \frac{22830}{57995} = 0{,}394; \qquad UA_6 = \frac{n_{66}}{n_{6+}} = \frac{22830}{24832} = 0{,}919;$$

Капа-індекс: $KI = \dfrac{n\sum_{i=1}^{K} n_{ii} - \sum_{i=1}^{K} n_{i+}n_{+i}}{n^2 - \sum_{i=1}^{K} n_{i+}n_{+i}} = 0{,}5923255 \approx 0{,}6.$

3) Коефіцієнти точності для III-го варіанта класифікування:

$n = 296465;\ \sum_{i=1}^{k} n_{ii} = 211861;\ OA = \dfrac{\sum_{i=1}^{K} n_{ii}}{n} = \dfrac{211861}{296465} = 0{,}715 -$ повна точність класифікування;

Точності виробника: Точності користувача:

$$PA_1 = \frac{n_{11}}{n_{+1}} = \frac{6268}{7954} = 0{,}788; \qquad UA_1 = \frac{n_{11}}{n_{1+}} = \frac{6268}{6827} = 0{,}918;$$

$$PA_2 = \frac{n_{22}}{n_{+2}} = \frac{60722}{64658} = 0{,}939; \qquad UA_2 = \frac{n_{22}}{n_{2+}} = \frac{60722}{70974} = 0{,}856;$$

$$PA_3 = \frac{n_{33}}{n_{+3}} = \frac{40526}{59375} = 0{,}683; \qquad UA_3 = \frac{n_{33}}{n_{3+}} = \frac{40526}{61907} = 0{,}655;$$

$$PA_4 = \frac{n_{44}}{n_{+4}} = \frac{71062}{89887} = 0{,}791; \qquad UA_4 = \frac{n_{44}}{n_{4+}} = \frac{71062}{110160} = 0{,}645;$$

$$PA_5 = \frac{n_{55}}{n_{+5}} = \frac{10390}{16157} = 0{,}643; \qquad UA_5 = \frac{n_{55}}{n_{5+}} = \frac{10390}{22238} = 0{,}467;$$

$$PA_6 = \frac{n_{66}}{n_{+6}} = \frac{22893}{58434} = 0{,}392; \qquad UA_6 = \frac{n_{66}}{n_{6+}} = \frac{22893}{24359} = 0{,}94;$$

Капа-індекс: $KI = \dfrac{n\sum_{i=1}^{K} n_{ii} - \sum_{i=1}^{K} n_{i+}n_{+i}}{n^2 - \sum_{i=1}^{K} n_{i+}n_{+i}} = 0{,}6305355 \approx 0{,}6.$

З отриманих розрахунків бачимо, що повна точність класифікування приймає найбільше значення при проведенні III-го варіанту класифікування, тобто класифікування із використанням 7-ми найбільш інформативних каналів та фільтрованих пікселів навчальних вибірок.

Таким чином, детальне проведення оцінки точності запропонованого методу контрольованого класифікування ГКЗ при наявності необмежених за об’ємом вибірок, підтверджує, що запропонований метод дає більш точні результати, ніж відомі методи класифікування.

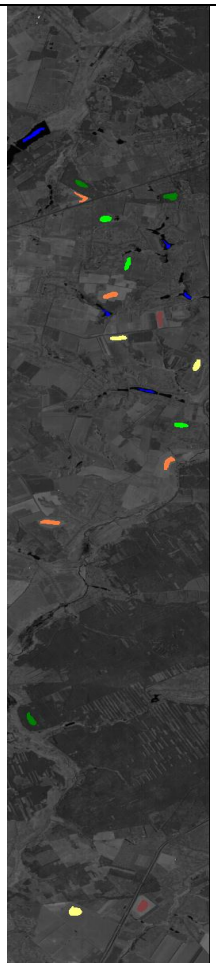
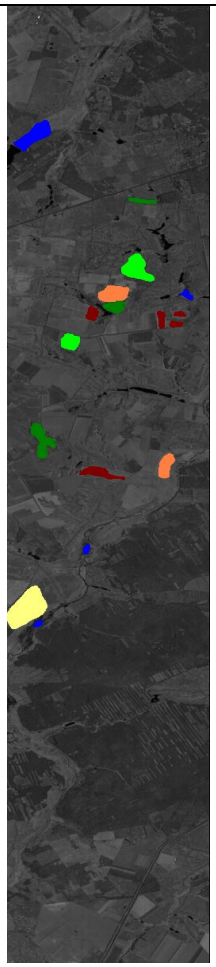
4.3. Оцінка ефективності методу контрольованого класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень

Для оцінки ефективності запропонованого у розділі 3 методу контрольованого класифікування ГКЗ, орієнтованого на умови забруднених навчальних вибірок, обмежених за об’ємом, було використано два зображення: гіперспектральний знімок Hyperion EO1H1810252014211110KF, отриманий 30 липня 2014 р., та гіперспектральний знімок Hyperion EO1H1810252013157110K3-Refint, отриманий 6 червня 2013 р.

Спочатку проведемо оцінювання ефективності розробленого методу класифікування за допомогою гіперспектрального знімку EO1H1810252014211110KF. На цьому знімку, представленому в табл. 4.4 у RGB-форматі, - смуга території на границі Макарівського, Києво-Святошинського та Фастівського районів Київської області. На території виділені наступні класи об'єктів: вода (1); хвойний ліс (2); листяний ліс (3)

(включаючи сади та лісопосадки); злакові поля (4), овочеві поля (5), прибрані поля (6).

Таблиця 4.4 – Класи об’єктів, навчальні підвибірки та контрольні ділянки

Класи об’єктів	**Навчальні підвибірки**		**Контрольні ділянки**		**Легенда**
Вода	335		1394		
Сосновий ліс	397		1471		
Листяний ліс	344		1541		
Злакові поля	525		1349		
Овочеві поля	440		2059		
Прибрані поля	315		1347		
Усього пікселів	2356		9161		

Спочатку експертом за допомогою розробленого СПЗ зі складу ГКЗ були видалені сильно зашумлені, а також подібні між собою зональні зображення (застосовувався критерій: коефіцієнт кореляції r_{thr} між сусідніми за спектром зображеннями не повинен перевищувати величини 0,8). Після відбору для подальшої роботи було залишено 17 зональних зображень.

Вихідна навчальна вибірка формувалася із використанням GoogleEarth, при цьому була проведена актуалізація гуглівської графічної інформації

шляхом польового виходу у район відзнятої території. Розташування підвибірок, використаних при навчанні, а також кількість пікселів у кожній з них показані в табл. 4.4 (ліве зображення).

В дослідженні також були задіяні контрольні ділянки, які не перетинаються із пікселами навчальних підвибірок. Розташування контрольних ділянок та кількість пікселів в них показані у правій частині табл. 4.4.

Точностні можливості запропонованого методу на основі ТСДШ (нижче у таблицях та на графіках цей метод позначається як DST) оцінювались у порівнянні з методом SVM. При навчанні ступінь забруднення вихідної навчальної вибірки моделювалася введенням до складу навчальних підвибірок класів забруднюючих пікселів, сигнали яких формувалися комп'ютерним генератором шуму. Всього було розглянуто 8 варіантів забруднення навчальної вибірки: 0, 10, 20, 30, 35, 40, 45 і 50% (включаючи випадок вихідної вибірки - без зашумлення).

Для кожного варіанту забрудненої навчальної вибірки ГКЗ класифікувалося окремо методом DST та методом SVM. Точність класифікування оцінювалася із використанням одних і тих же контрольних ділянок (див. табл. 4.4). У табл. Д.1 (Додаток1) наведена кількість пікселів у навчальних вибірках виділених класів для різних варіантів забруднення. У кожному випадку будувалася своя матриця помилок і розраховувалися показники точності класифікування (Табл. Д.2–Д.10, Д.12-Д.14) (Додатки 1, 2, 3). Дані для побудови графіків зміни точності класифікування пікселів еталонних ділянок в залежності від ступеня забруднення навчальних вибірок кожного виділеного класу наведені у табл. Д.11 (Додаток 3). Також у Додатку 4 наведений розподіл помилок пропуску для кожного із розглянутих класів при 30, 35, 40, 45% (рис. Д.1–Д.4) та розподіл помилок ототожнення із класом при забрудненні 40, 45% (рис. Д.5 і Д.6).

Наведемо основні із отриманих результатів. У табл. 4.5 і 4.6 представлені матриці помилок класифікування ГКЗ методом DST, отримані відповідно на

вихідній (незабрудненій) навчальній вибірці і на 50% забрудненій навчальній вибірці. Розрахунки показали, що метод DST дав середню за всіма класами загальну точність (Overall Accuracy) 0,859 при навчанні на незабрудненій вибірці і 0,629 - на 50% забрудненій вибірці. У випадку методу SVM були отримані відповідно такі точності класифікування: 0,778 і 0,475. Тобто точність методу DST виявилася на 8-16% вище у порівнянні із методом SVM. Більш детальну інформацію дають графіки точності класифікування кожного із методів в залежності від ступеня забруднення навчальної вибірки, наведені на рис. 4.11.

Таблиця 4.5 – Матриця помилок методу DST при вихідній навчальній вибірці (при відсутності забруднення)

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку
DST	Вода	1062	0	0	0	0	0	1062
	Сосновий ліс	162	1260	27	23	14	6	1492
	Листяний ліс	77	133	1372	18	3	26	1629
	Злакові поля	17	5	8	1175	140	74	1419
	Овочеві поля	76	73	134	102	1902	143	2430
	Прибрані поля	0	0	0	31	0	1098	1129
	Сума елементів стовпчика	1394	1471	1541	1349	2059	1347	9161
	Точність	0,762	0,857	0,890	0,871	0,924	0,815	

Таблиця 4.6 – Матриця помилок методу DST при 50%-му забрудненні навчальної вибірки

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку
DST	Вода	1217	853	27	12	93	4	2206
	Сосновий ліс	71	564	101	20	331	13	1100
	Листяний ліс	17	10	882	9	3	13	934
	Злакові поля	30	3	111	914	552	2	1612
	Овочеві поля	39	32	112	15	874	5	1077
	Прибрані поля	20	9	308	379	206	1310	2232
	Сума елементів стовпчика	1394	1471	1541	1349	2059	1347	9161
	Точність	0,873	0,383	0,572	0,678	0,424	0,973	

Точності класифікування для 8 варіантів забруднення навчальної вибірки: 0, 10, 20, 30, 35, 40, 45 і 50% (включаючи випадок вихідної вибірки - без зашумлення) для методу DST та методу SVM наведені на рис.4.11.

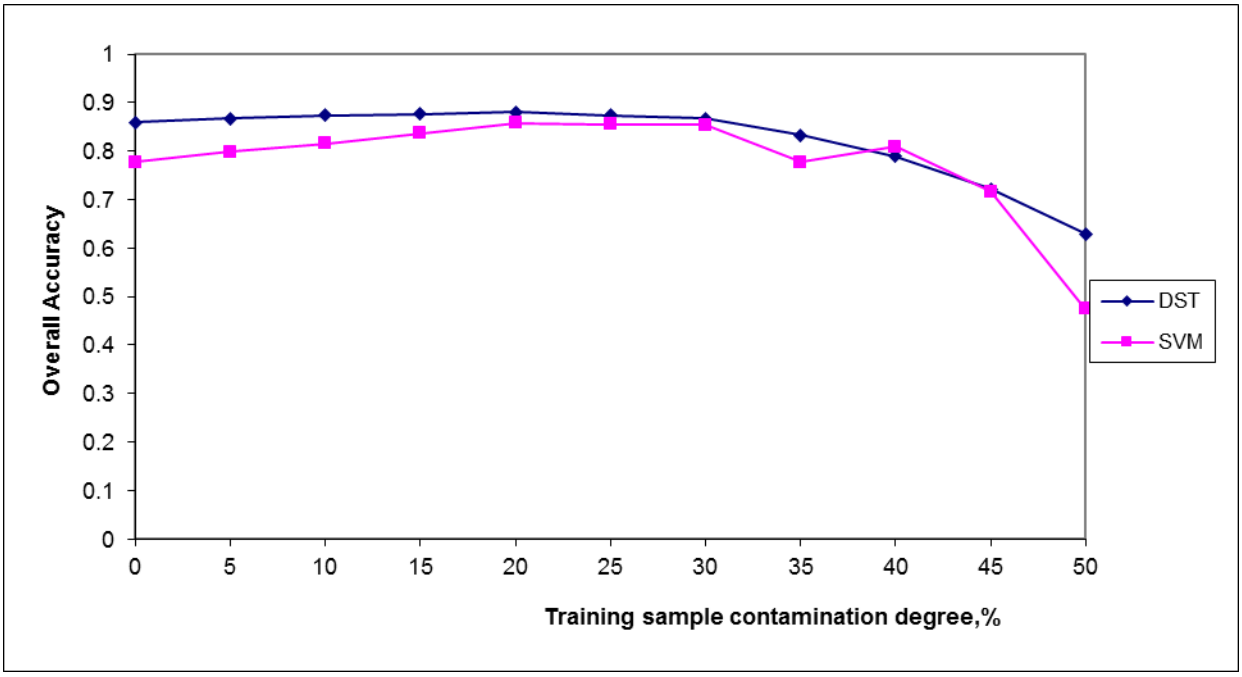


Рис. 4.11. Точність класифікування у залежності від ступеня забруднення навчальної вибірки

Для кожного класу були також розраховані помилки пропуску класу (Error of Omission) і помилки ототожнення із класом (Error of Commission) (Додаток). На рис. 4.12 показано розподіл помилок пропуску для кожного із розглянутих класів для випадку вихідної (без забруднення) навчальної вибірки. Нумерація класів відповідає наведеній у табл. 4.4.

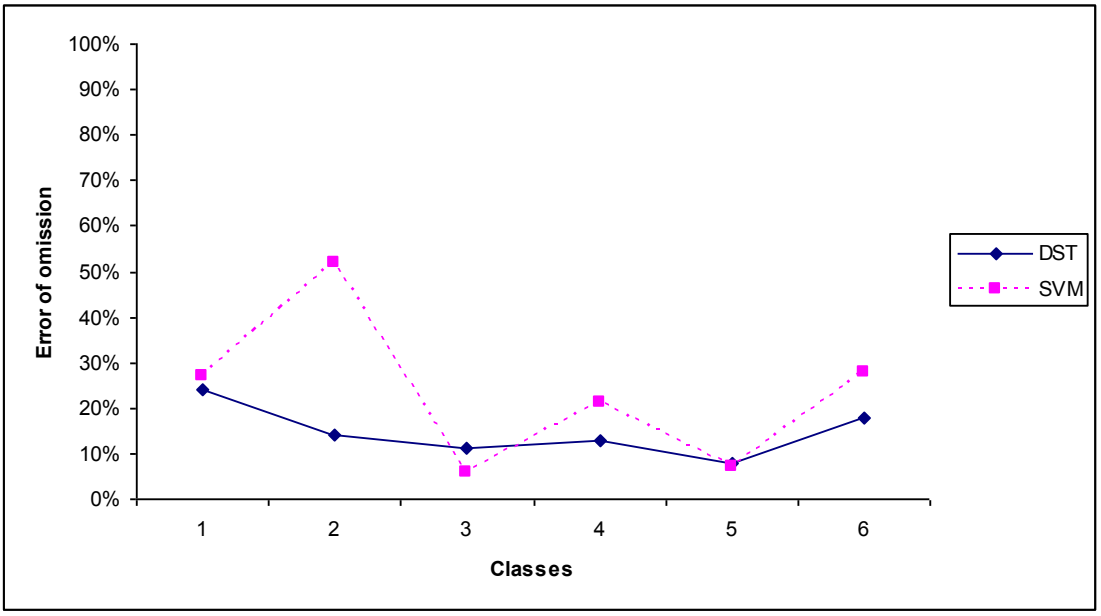


Рис. 4.12. Розподіл помилок пропуску для кожного із розглянутих класів (випадок вихідної навчальної вибірки - забруднення 0%)

На графіках рис. 4.12 та 4.13 наведені величина помилки пропуску класу та величина помилки ототожнення із класом для кожного із розглянутих класів у випадку вихідної навчальної вибірки для запропонованого методу класифікування DST і методу SVM. З наведених графіків видно, що запропонований метод DST дає значно меншу помилку пропуску класу та помилку ототожнення із класом, ніж метод SVM.

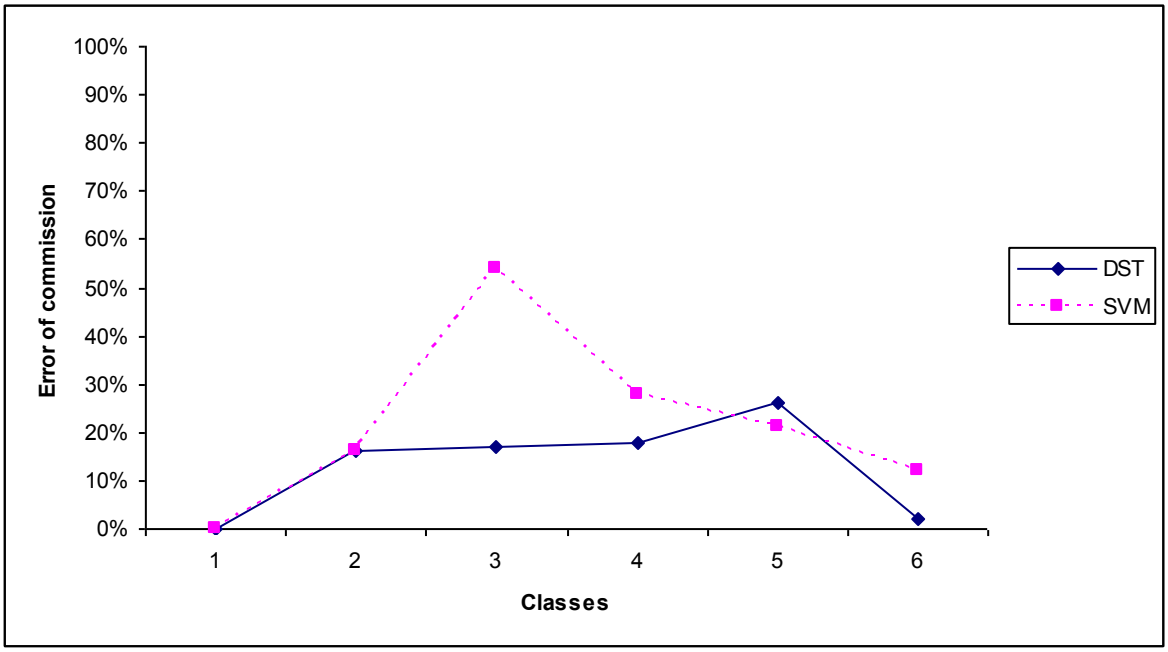


Рис. 4.13. Розподіл помилок ототожнення із класом (випадок вихідної навчальної вибірки - забруднення 0%)

Усереднені за класами результати розрахунків для випадку вихідної навчальної вибірки (без забруднення) наведено у табл. 4.7. Їхне порівняння показує, що і за цим показником метод класифікування DST виявляється на 9% більш ефективним, ніж метод SVM.

Таблиця 4.7 – Усереднені помилки пропуску та ототожнення

Метод класифікування	Усереднена помилка пропуску класу (Average Omission Error)	Усереднена помилка ототожнення із класом (Average Commission Error)
DST	15%	13%
SVM	24%	22%

Тепер розглянемо графіки, де показана залежність величини помилки пропуску классу від розглянутих класів при забрудненні навчальної вибірки на 10% та 20% (рис. 4.14 та 4.15).

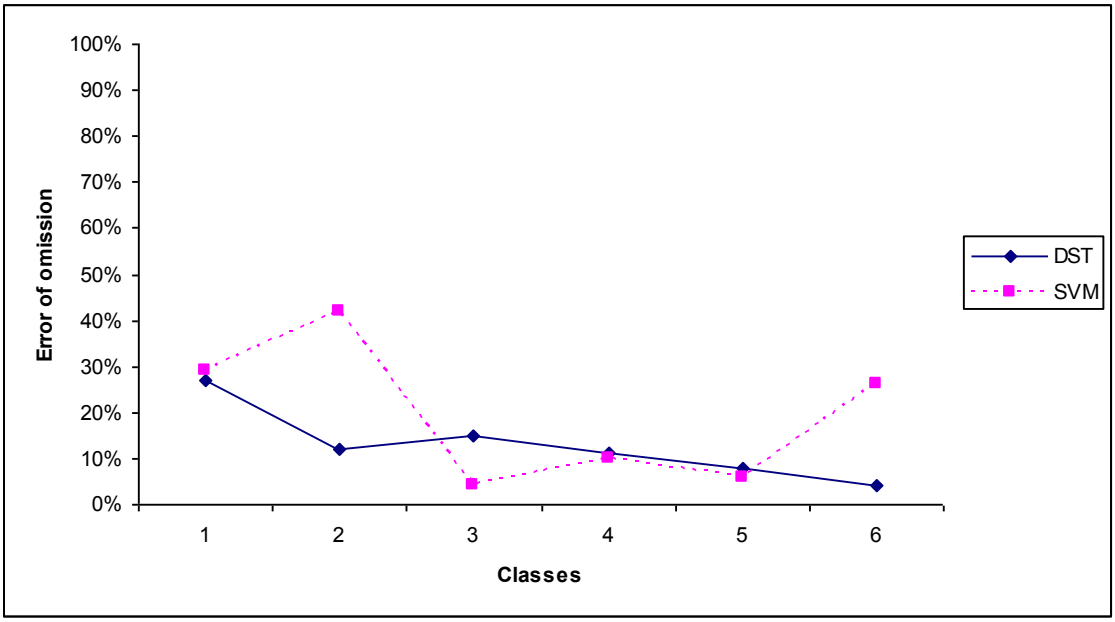


Рис.4.14. Розподіл помилок пропуску для кожного із розглянутих класів при забрудненні 10%

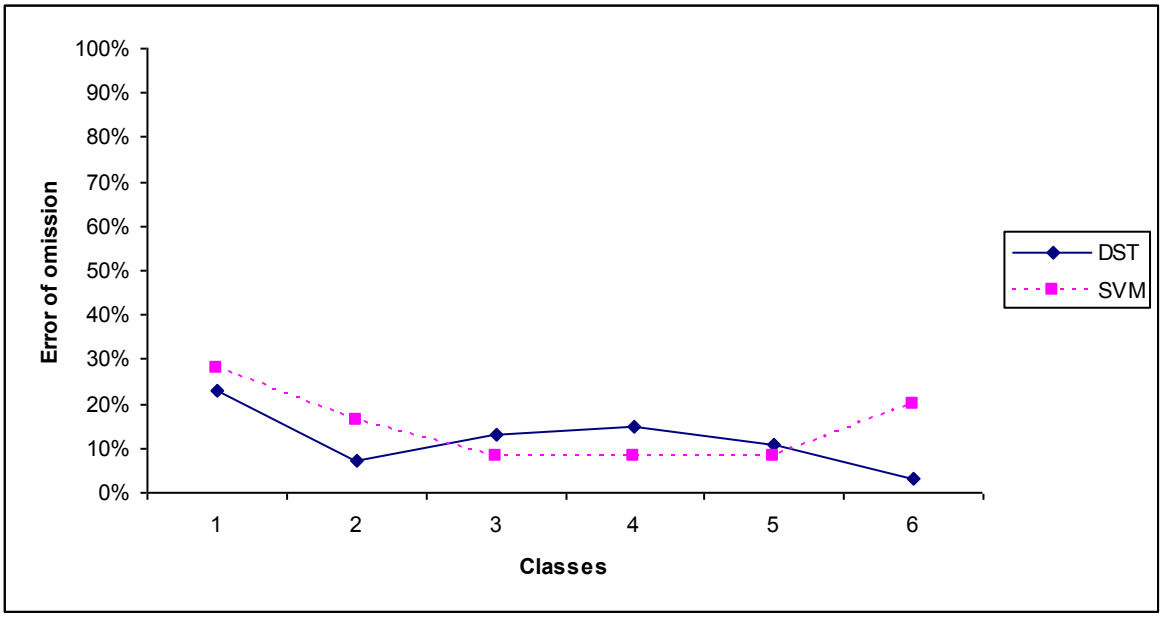


Рис.4.15. Розподіл помилок пропуску для кожного із розглянутих класів при забрудненні 20%

З наведених графіків можна зробити висновок, що метод DST дає значно меншу помилку пропуску класу, ніж метод SVM при забрудненні 10% та 20%.

Тепер наведемо графіки розподілу помилок ототожнення із класом при забрудненні навчальної вибірки на 10%, 20%, 30% та 35% (рис. 4.16 - 4.19).

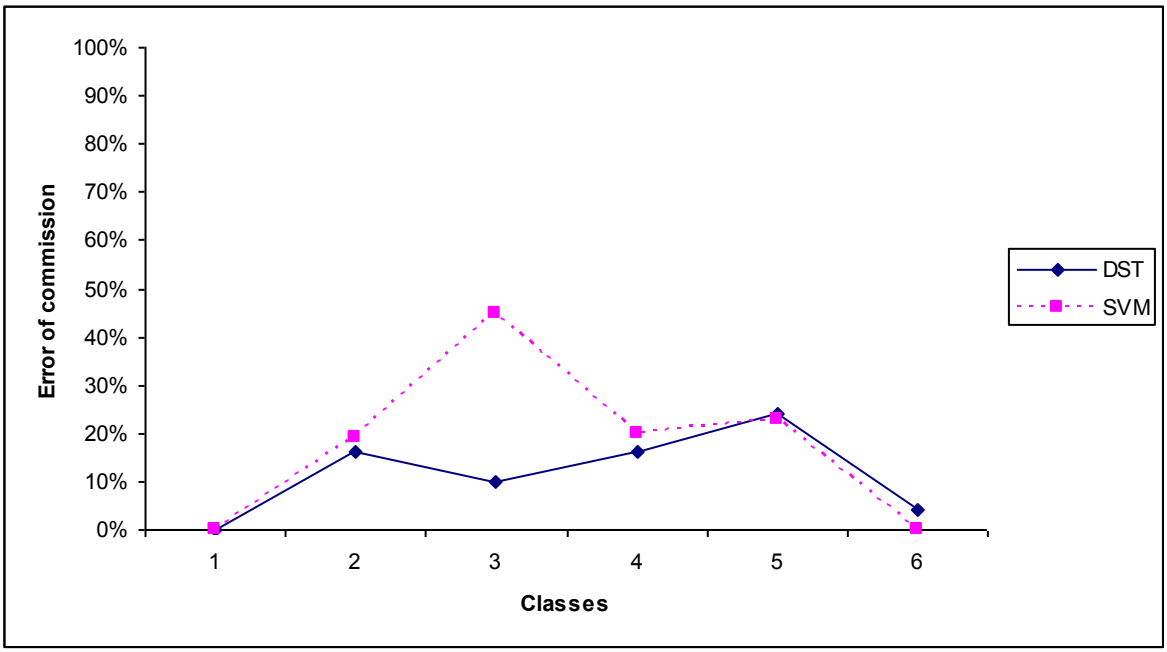


Рис. 4.16. Розподіл помилок ототожнення із класом при забрудненні 10%

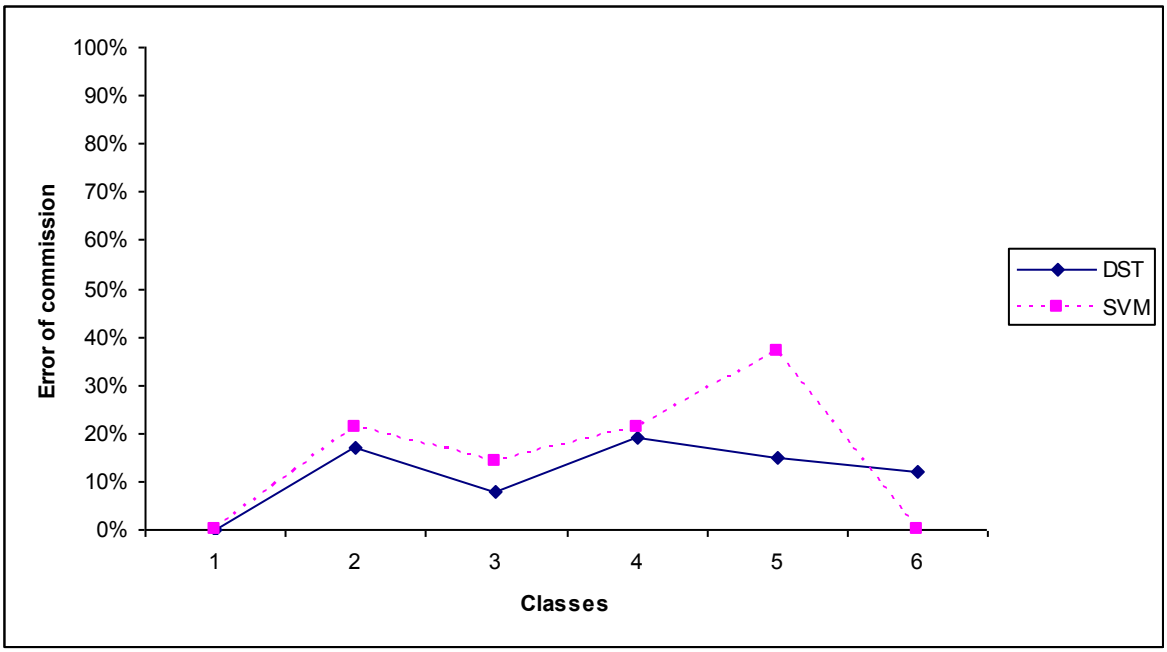


Рис. 4.17. Розподіл помилок ототожнення із класом при забрудненні 20%

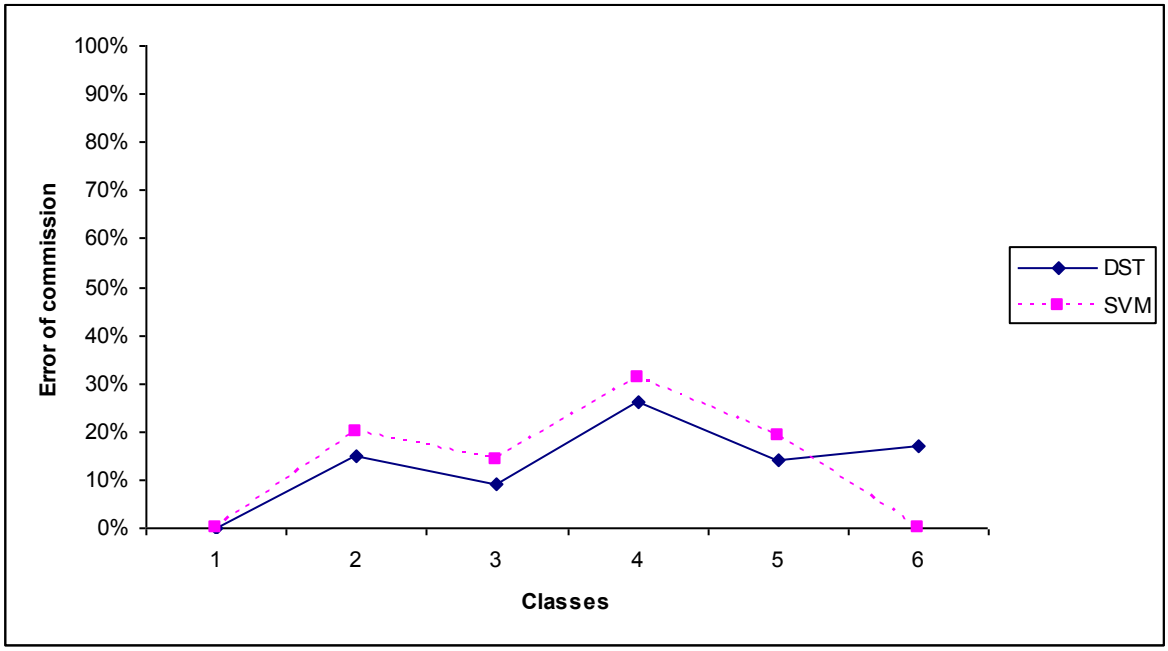


Рис. 4.18. Розподіл помилок ототожнення із класом при забрудненні 30%

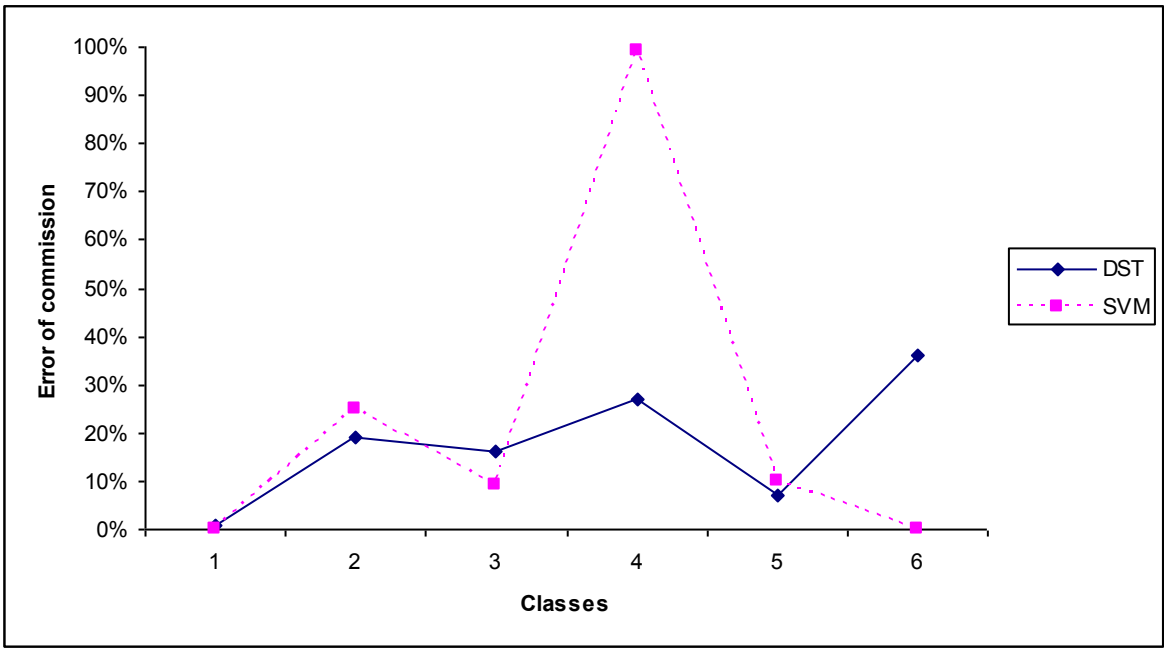


Рис. 4.19. Розподіл помилок ототожнення із класом при забрудненні 35%

Із наведених графіків можна видно, що метод DST дає значно меншу помилку ототожнення із класом, ніж метод SVM.

На графіках рис. 4.20 і 4.21 наведена величина помилки пропуску класу та величина помилки ототожнення із класом для кожного з розглянутих класів при забрудненні навчальної вибірки на 50 % для запропонованого методу класифікування DST і методу SVM.

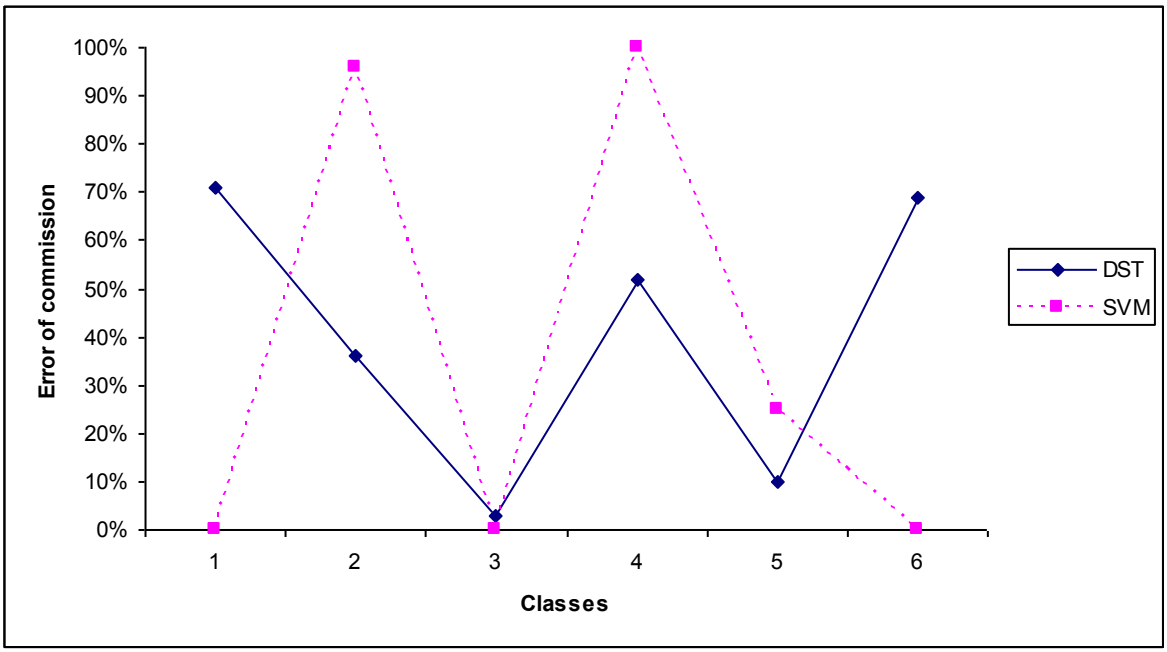


Рис. 4.20. Розподіл помилок пропуску для кожного із розглянутих класів при забрудненні 50%

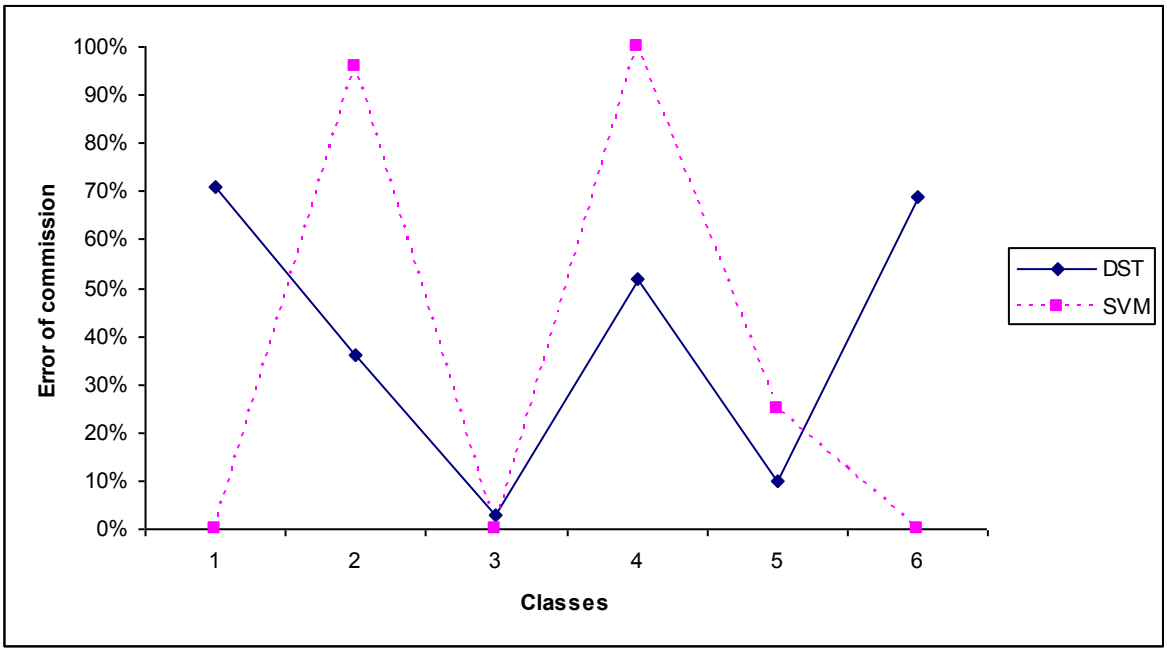


Рис. 4.21. Розподіл помилок ототожнення із класом при забрудненні 50%

З наведених тут графіків видно, що запропонований метод DST дає значно менші помилку пропуску класу та помилку ототожнення із класом, ніж метод SVM.

Подальше дослідження ефективності розробленого і описаного в розділі 3 методу класифікування ГКЗ для малих за кількістю складових пікселів та значно забруднених навчальних вибірок проводилася з використанням ще одного гіперспектрального знімку EO1H1810252013157110K3-Refint, отриманого апаратурою Hyperion супутника EO-1 6 червня 2013 р. (рис. 4.22). На цьому знімку, представленому в табл. 4.8 у RGB-форматі, - ділянки Іванківського та Вишгородського районів Київської області. Також на знімку представлена дельта річки Тетерів при впаданні у Київське водосховище.

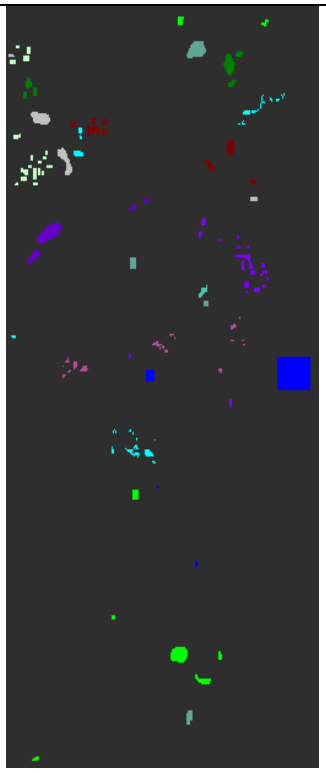
За описаною вище процедурою методу DST на знятій території було виділено і прокласифіковано наступні класи об'єктів: вода (1); листяний ліс (2); перелоги (3); трава (4), піщані грунти (5), грунти (6); хвойний підлісок (7); водно-болотні угіддя 1 (8); водно-болотні угіддя 2 (9); водно-болотні угіддя 3 (10); хвойний ліс (11); мішаний ліс (12) (див. табл. 4.8). Також даний знімок був оброблений із залученням алгоритму класифікування об’єктів за їх спектрально-топологічними характеристиками.

При цьому, для повноти дослідження і одержання більш репрезентативних оцінок [98–111], процедура класифікування в обох випадках повторювалась при різних способах відбору інформативних спектральних каналів:

- з використанням функції кореляції (Correlation split/merge);
- на основі дивергенції Кульбака-Лейблера (KL selection);
- з використанням кількісної міри інформативності набору спектральних каналів на основі дивергенції Кульбака-Лейблера між спектральними розподілами оптичних сигналів об’єктів і фонів (C selection);
- з використанням функції похибок (Erf selection).

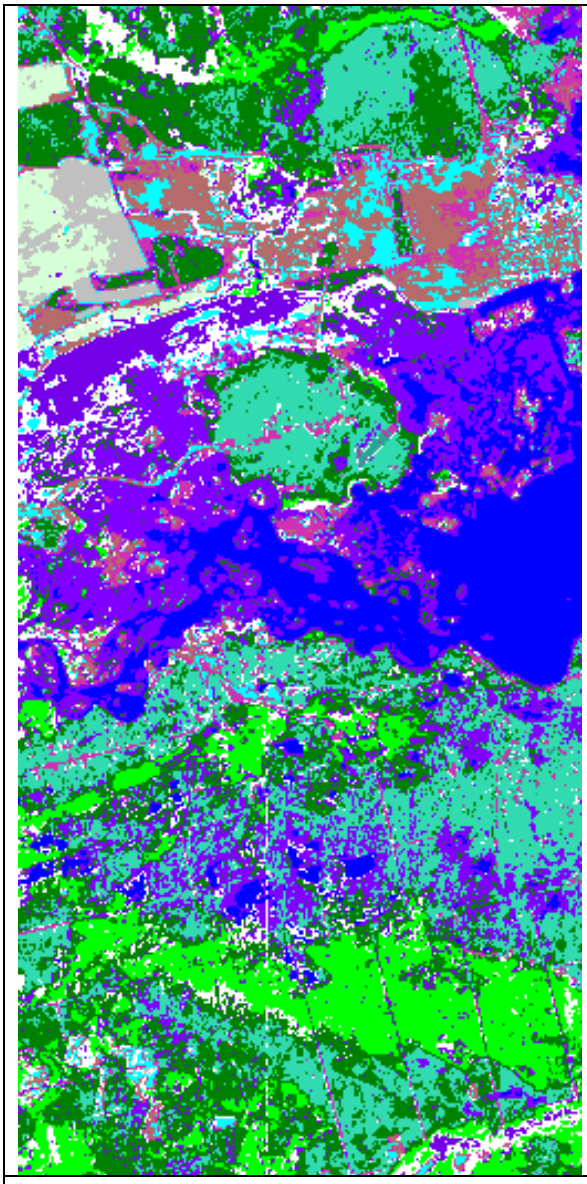
Рис. 4.22. Знімок Hyperion EO1H1810252013157110K3-Refint (06.06.2013 р.)

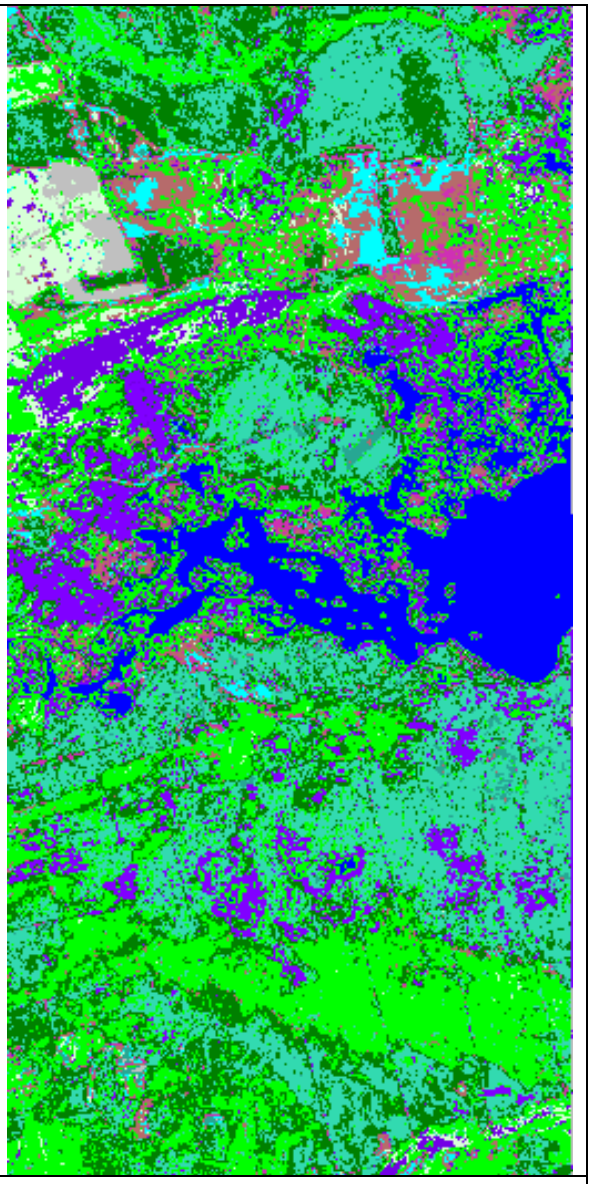
Таблиця 4.8 – Класи об’єктів та навчальні підвибірки

Класи об’єктів	**Число пікселів для**		**Розміщення навчальних підвибірок виділених класів**	**Легенда**
	контрольних ділянок	**перевіряємих ділянок**		
Вода	1332	688		
Листяний ліс	582	290		
Перелоги	433	212		
Трава	422	204		
Піщані грунти	404	234		
Грунт	142	201		
Хвойний підлісок	0	30		
Водно-болотні угіддя1	430	244		
Водно-болотні угіддя2	76	201		
Водно-болотні угіддя3	77	118		
Хвойний ліс	461	237		
Мішаний ліс	441	200		

Одержані по результатах класифікування зображення EO1H1810252013157110K3-Refint, розподіли класів за методом DST та алгоритмом класифікування об’єктів за їх спектрально-топологічними характеристиками при перелічених способах відбору спектральних каналів наведені на рис. 4.23-4.26.

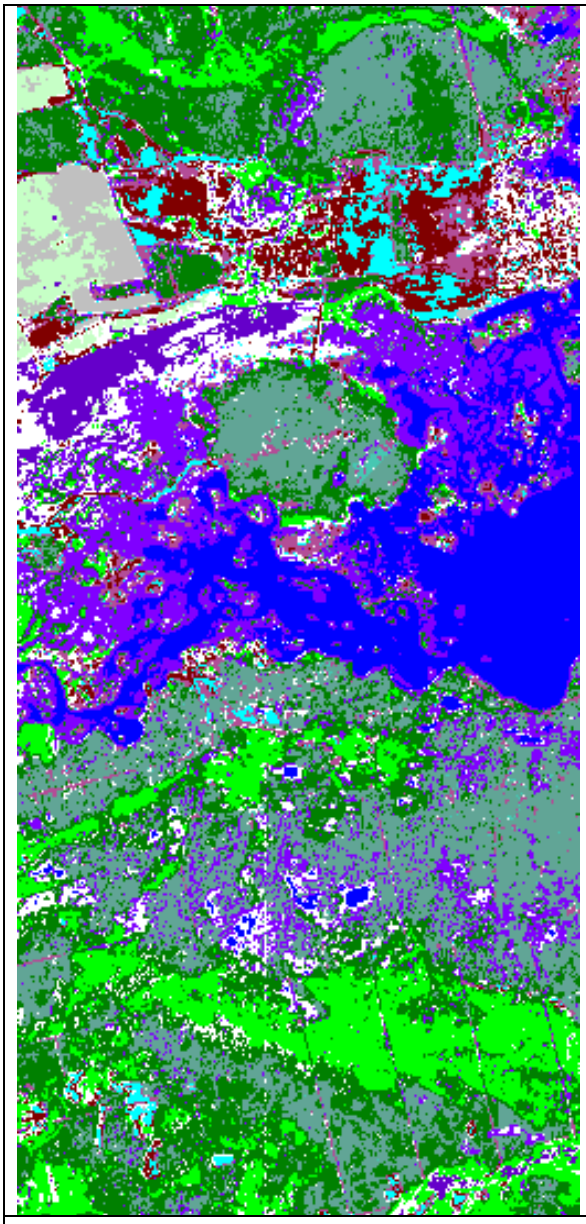
Загальна точність кожного методу відображена в табл. 4.9.

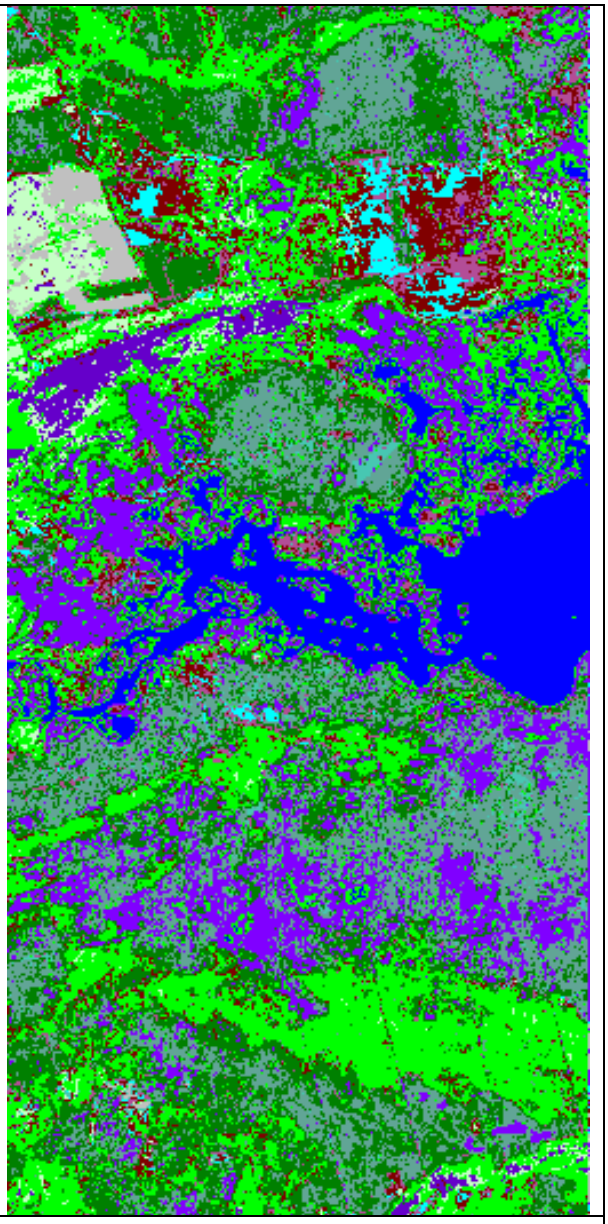




DST	Алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками

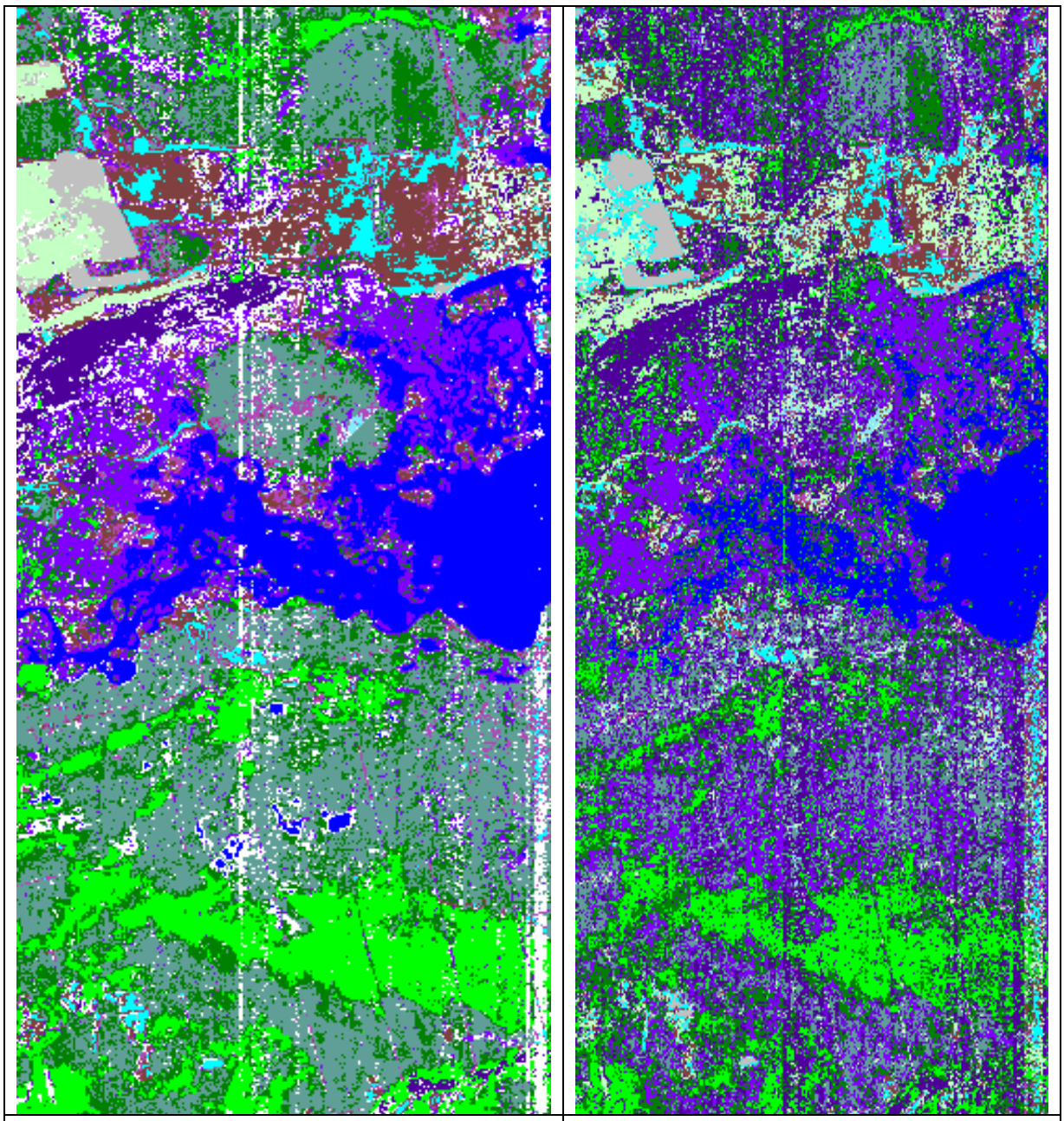
Рис. 4.23. Результат класифікування: (в дужках вказані розрахункові спектральні канали)
Correlation split/merge (12:34; 37:74; 87:96; 100:115; 134:163; 191:217)





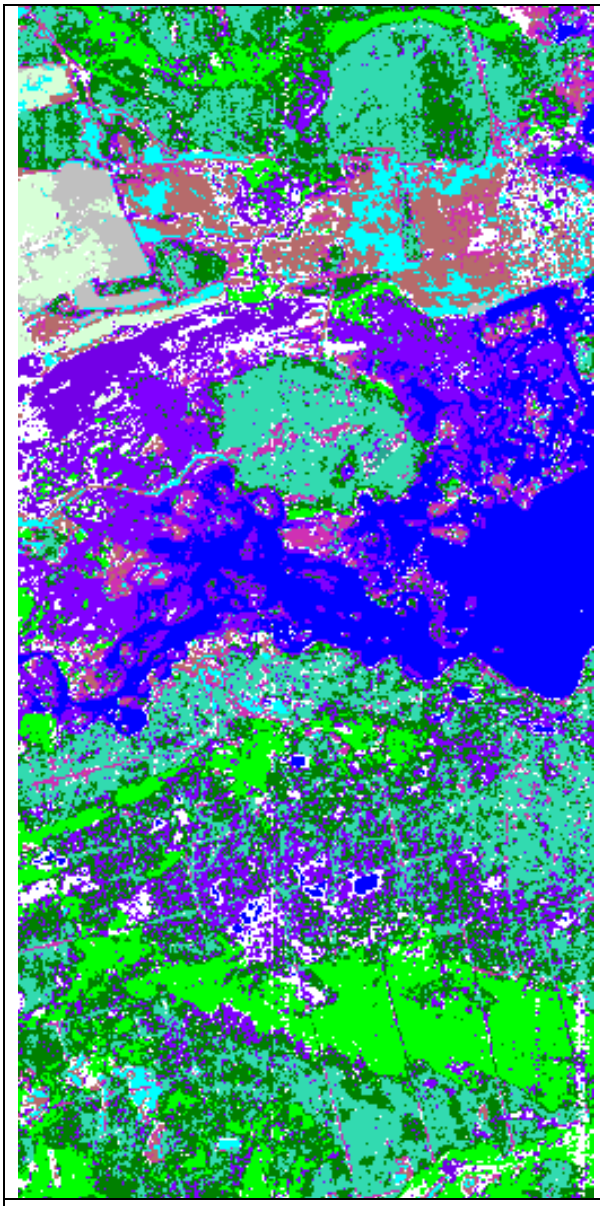
DST	Алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками

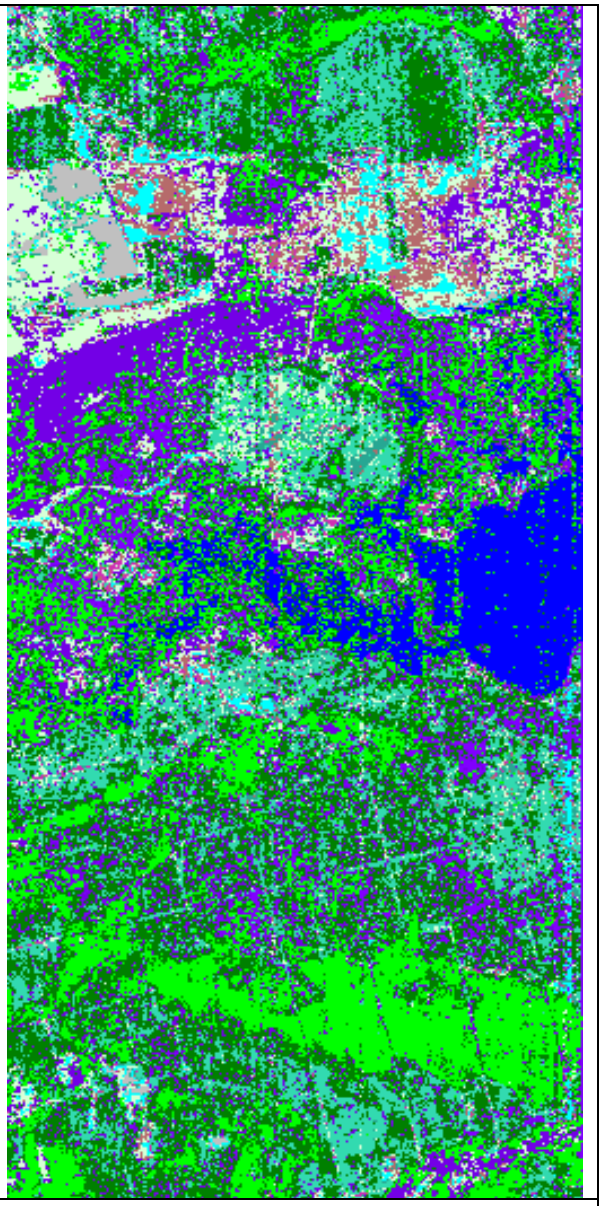
Рис. 4.24. KL selection (23, 45, 72, 66, 103, 108, 145, 157, 199)



DST	Алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками

Рис. 4.25. C selection (12:13, 40, 43:45, 130, 132, 181, 182:184, 185:186, 189:190)





DST	Алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками

Рис. 4.26. Erf selection (20, 33, 40, 130, 153, 181, 183, 211)

Таблица 4.9 – Точність класифікування

Image classification algorithm	Спосіб відбору інформативних спектральних каналів				
	Radiometric calibration and surface reflectance (172 bands total)	Correlation split/merge (12:34, 37:74, 87:96, 100:115, 134:163, 191:217)	KL selection (23, 45, 72, 66, 103, 108, 145, 157, 199)	C selection (12:13, 40, 43:45, 130, 132, 181, 182:184, 185:186, 189:190)	Erf selection (20, 33, 40, 130, 153, 181, 183, 211)
DST	0,900	0,905	0,888	0,883	0,888
Алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками	0,817	0,772	0,794	0,685	0,735

Порівняння одержаних в проведеному експериментальному дослідженні оцінок показує, що запропонований в розділі 3 метод має хоча й невелику, але стабільну перевагу в точності класифікування перед топологічним алгоритмом.

ВИСНОВКИ ДО РОЗДІЛУ 4

1. Описано програмну реалізацію запропонованих методів класифікування гіперспектральних зображень в умовах навчальних вибірок обмеженої якості. Розроблено спеціалізоване програмне забезпечення DST у середовищі пакету Visual Basic 6.0.

2. Проведено оцінку ефективності запропонованих методів на реальних гіперспектральних зображеннях, отриманих апаратурою Hyperion супутника EO-1. Досліджено вплив процедур фільтрації навчальної вибірки і відбору інформативних спектральних каналів шляхом ранжування за їх класифікаційною цінністю на точність класифікування.

3. Проведено оцінку ефективності розробленого методу класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок з використанням гіперспектрального знімка від 30 липня 2014 р. EO1H1810252014211110KF, отриманого апаратурою Hyperion супутника EO-1. Результати оцінювання даного методу показали, що його загальна точність класифікування (Overall Accuracy) складає 0,802 і перевищує точність відомого методу SVM (складає 0,757) та алгоритму класифікування об’єктів за їх спектрально-топологічними характеристиками (складає 0,734).

4. Проведено експериментальну оцінку ефективності запропонованого методу DST класифікування гіперспектральних космічних зображень в умовах забруднених малих навчальних вибірок, для чого були використані гіперспектральні знімки, отримані апаратурою Hyperion. Було оцінено можливості запропонованого методу DST у порівнянні з методом SVM. Дослідження показало, що метод DST дав середню за всіма класами загальну точність 0,859 при навчанні на незабрудненій вибірці і 0,629 - на 50% забрудненій вибірці. У випадку методу SVM були отримані відповідно такі точності класифікування: 0,778 і 0,475. Тобто точність методу DST виявилася на 8-16% вище у порівнянні із методом SVM. Встановлено, що усереднена

помилка пропуску та усереднена помилка ототожнення із класом для методу DST становлять 15% та 13% відповідно, а для методу SVM: 24% та 22% відповідно, тобто запропонований метод DST в умовах забруднених малих навчальних вибірок показує перевагу перед методом SVM.

5. Порівняльна оцінка точності запропонованого метода DST та алгоритма класифікування об’єктів за їх спектрально-топологічними характеристиками була здійснена з урахуванням способу відбору спектральних каналів. Для методу DST при відборі спектральних каналів із використанням функції кореляції загальна точність методу DST склала 0,905; при відборі на основі дивергенції Кульбака-Лейблера - 0,888; при відборі з використанням кількісної міри інформативності на основі дивергенції Кульбака-Лейблера між спектральними розподілами оптичних сигналів об’єктів і фонів - 0,883; при відборі з використанням функції похибок або функції Лапласа - 0,888. В свою чергу, для алгоритму класифікування об’єктів за їх спектрально-топологічними характеристиками значення загальної точності класифікування при різних способах відбору спектральних каналів складало відповідно: 0,772, 0,794, 0,685, 0,735. Тобто, загальна точність класифікування методом DST значно перевищує загальну точність класифікування за допомогою алгоритма класифікування об’єктів за їх спектрально-топологічними характеристиками.

ЗАГАЛЬНІ ВИСНОВКИ ПО РОБОТІ

1. Розглянуто місце і роль процедури класифікування космічних зображень при розв’язанні тематичних задач дистанційного зондування Землі. Проведено стислий огляд сучасних методів і алгоритмів класифікування гіперспектральних космічних зображень. Показано, що сучасні алгоритми контрольованого класифікування гіперспектральних космічних зображень досить чутливі до характеристик навчальних вибірок. Більш конкретно, точність відомих алгоритмів суттєво залежить від об’ємів навчальних вибірок та рівня їх забрудненості.

2. Сформульовано завдання розробити та обґрунтувати нові ефективні методи контрольованого класифікування гіперспектральних космічних зображень в умовах, коли навчальна вибірка є забрудненою, в тому числі сильно, а її об’єм є дуже обмеженим. Наведено критерії, за якими передбачається оцінювати ефективність нових методів контрольованого класифікування. Математичною платформою для розроблення нових методів контрольованого класифікування обрано апарат та положення теорії свідчень Демпстера-Шейфера. Вибір пов'язаний з тим, що математичний апарат ТСДШ дає можливість в умовах забрудненості та неповноти даних працювати зі складними багатоальтернативними гіпотезами, пропонує просте правило для комбінування даних від різних джерел (у тому числі зображень), а також має деякі інші переваги перед традиційним теоретико-ймовірнісним підходом.

3. Розроблено новий метод класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок. Введено модель гіперспектрального космічного зображення та припущення. Детально розглянуто організацію і зміст методу класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок. Запропонований метод відрізняється від відомих розв’язків однойменних задач введенням процедури оцінювання класифікаційної цінності спектральних

каналів за допомогою спеціальної емпіричної функції і особливим способом розбиття спектрального ознакового простору, що, у сукупності, дозволяє застосовувати лише найбільш інформативні спектральні канали і, таким чином, суттєво скоротити розмірність ознакового простору при одночасному підвищенні точності класифікації. Детально описано процедуру відбору зональних зображень для виключення зашумлених зображень із використанням коефіцієнту кореляції між сусідніми зональними зображеннями, наведено процедури фільтрації навчальної вибірки, розбиття ознакового простору.

4. Запропоновано новий метод контрольованого класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та обмежень на кількість елементів у виборці. Наводяться обмеження, припущення та визначення, на яких побудовано метод. Пропонований метод враховує невизначеності, які можуть виникати при навчанні та класифікуванні. Детально описано організацію метода і основні процедури, з яких він складається, у тому числі процедуру відбору зональних зображень і схему класифікування гіперспектральних космічних зображень в умовах забруднення навчальних вибірок та кількісних обмежень.

5. Запропоновано новий спосіб відбору зональних зображень і формування репрезентативної комбінації зональних зображень при класифікуванні гіперспектрального аерокосмічного знімка, який дозволяє об’єктивізувати і прискорити процес вирішення задачі за рахунок зменшення вимог до необхідного обчислювального ресурсу. Дана перевага досягається за рахунок автоматизації процедури оцінювання рівню шумів кожного з вихідних зональних зображень і відсіювання з подальшого аналізу зображень, які мають високий рівень зашумленості.

Показано, що введення операції відбору зашумлених зональних зображень у випадку знімків EO-1 / Hyperion дозволяє зменшити необхідні обчислювальні витрати в 2,78-4,64 рази, а у випадку знімків AVIRIS – в 2,92-3,64 рази.

6. Проведено експериментальну оцінку ефективності обох запропонованих в дисертаційній роботі методів контрольованого класифікування на реальних гіперспектральних космічних зображеннях, отриманих апаратурою Hyperion супутника EO-1, на базі спеціально розробленого програмного забезпечення DST.

7. Проведено оцінку ефективності розробленого методу класифікування гіперспектральних космічних зображень в умовах забруднених навчальних вибірок з використанням гіперспектрального знімка EO1H1810252014211110KF, отриманого 30 липня 2014 р. бортовою апаратурою Hyperion супутника EO-1. Результати оцінювання даного методу показали, що його загальна точність (Overall Accuracy) складає 0,802 і перевищує загальну точність відомого методу SVM (склала 0,757) та алгоритму класифікування об’єктів за їх спектрально-топологічними характеристиками (склала 0,734).

8. Проведено оцінку ефективності розробленого методу DST класифікування гіперспектральних космічних зображень Hyperion в умовах забруднених малих навчальних вибірок. Дослідження показало, що метод DST дав середню за всіма класами загальну точність 0,859 при навчанні на незабрудненій вибірці і 0,629 на 50% забрудненій вибірці. У випадку методу SVM були отримані відповідно такі точності класифікування: 0,778 і 0,475. Тобто точність методу DST в умовах забруднених малих навчальних вибірок є на 8-16% вище у порівнянні із методом SVM. Встановлено, що усереднена помилка пропуску та усереднена помилка ототожнення із класом для методу DST становлять 15% та 13% відповідно, а для методу SVM - 24% та 22% відповідно, тобто запропонований метод DST в умовах забруднених малих навчальних вибірок показує перевагу перед методом SVM.

9. Здійснено порівняльну оцінку точності запропонованого метода DST та алгоритма класифікування об’єктів за їх спектрально-топологічними характеристиками з урахуванням способу відбору спектральних каналів. Для методу DST при відборі спектральних каналів із використанням функції

кореляції загальна точність методу DST склала 0,905; при відборі на основі дивергенції Кульбака-Лейблера - 0,888; при відборі з використанням кількісної міри інформативності на основі дивергенції Кульбака-Лейблера між спектральними розподілами оптичних сигналів об’єктів і фонів - 0,883; при відборі з використанням функції похибок або функції Лапласа - 0,888. Для алгоритму класифікування об’єктів за їх спектрально-топологічними характеристиками значення загальної точності класифікування при різних способах відбору спектральних каналів складало відповідно: 0,772, 0,794, 0,685, 0,735. Тобто, точність запропонованого метода DST в умовах забруднених малих навчальних вибірок є на 11-17% вище у порівнянні із тою, що забезпечує алгоритм класифікування об’єктів за їх спектрально-топологічними характеристиками.

10. Таким чином, результати проведених наукових досліджень свідчать, що розроблено і обгрунтовано нові ефективні методи контрольованого класифікування гіперспектральних космічних зображень, які здатні забезпечити підвищення точності класифікування в умовах забруднення та малої кількості складових елементів, тобто мета дисертаційної роботи *досягнута.*

11. Подальші дослідження мають бути спрямовані на оптимізацію обчислювальних витрат на відбір найбільш інформативних зональних зображень, а також на пошук методів попереднього перетворення ознакового простору, які забезпечать можливість краще розрізняти класи об’єктів у просторі спектральних ознак.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Альперт С.І. Сучасні критерії оцінки точності класифікації аерокосмічних зображень / С.І. Альперт // Математичні машини і системи. – 2013. – № 4. – С. 187–197.

2. Альперт С. І. Оцінка якості класифікації аерокосмічних зображень на основі матриці помилок та коефіцієнтів точності / С. І. Альперт // Математичні машини і системи. – 2014. – № 1. – С. 101–107.

3. Альперт С.І. Аналіз методів відбору найбільш інформативних спектральних каналів для вивчення об’єктів дистанційного зондування / С. І. Альперт // Математичні машини і системи. – 2014. – № 2. – С. 71–77.

4. Попов М.О. Метод класифікування гіперспектральних космічних зображень з використанням комбінаційного правила Демпстера / М.О. Попов., С.І. Альперт, В.М. Подорван // Космічна наука і технологія. – 2015. – Т. 21. – № 1. С.25–37.

5.Попов М.А. Метод классификации космических изображений с использованием подхода Демпстера Шейфера / М.А. Попов, С.И. Альперт, В.Н. Подорван // Исследование Земли из космоса (РФ). – 2016. – № 5. – С. 26–37.

6. Альперт С.І. Методи селекції інформативних зональних зображень при класифікації гіперспектральних супутникових зображень / С.І. Альперт // Математичні машини і системи. – 2015. – № 2. – С. 40 – 48.

7. Альперт С.І. Оцінка точності класифікації космічних зображень на основі теорії Демпстера-Шафера / С.І. Альперт // Збірник праць XI-ої Міжнародної молодіжної науково-практичної конференції “Історія розвитку науки, техніки та освіти” за темою “Розбудова дослідницького університету”, (Київ, 25 квітня 2013 р.). – Київ, 2013. – С. 242–245.

8. Альперт С.І. Сучасні методи оцінки точності класифікації/ С.І. Альперт // Геоінформатика – теоретичні і прикладні аспекти: матеріали 12-ї Міжнародної конференції, (Київ, 13–16 травня 2013). – К.: Наук. думка, 2013. – С. 1–4.

9. Альперт С.І. Огляд різних методів відбору найбільш інформативних спектральних діапазонів / С.І. Альперт // Геоінформатика – теоретичні та прикладні аспекти: матеріали 13-ої Міжнародної конференції, (Київ, 12–15 травня 2014 р.). – К.: Наук. думка, 2014. – С. 1–5.

10. Альперт С.І. Огляд оцінок точності та неточності класифікації земляного покриву та їх властивостей / С.І. Альперт // Тези доповідей IV–ої Всеукраїнської конференції "Аерокосмічні спостереження в інтересах сталого розвитку та безпеки GEO-UA", (Київ, 26–30 травня 2014 р.). – Київ, 2014. – С. 6 –8.

11. Альперт С.І. Методи відбору найбільш інформативних спектральних каналів/ С.І. Альперт //Збірник праць VII-ої Міжнародної науково-технічної конференції студентів та аспірантів "Перспективи розвитку інформаційно-телекомунікаційних технологій та систем", (Київ, 21–24 квітня 2015).– Київ, 2015. – С. 335–337.

12. Альперт С.І. Нова модифікована критеріальна функція інформативності спектральних каналів / С.І. Альперт // Геоінформатика – теоретичні та прикладні аспекти: матеріали 14–ої Міжнародної конференції, (Київ, 11–14 травня 2015). – К.: Наук. думка, 2015. – С. 1–6.

13. Попов М.А. Метод автоматической классификации гиперспектральных космических изображений на основе теории свидетельств Демпстера–Шейфера/ М.А. Попов, В.Н. Подорван, С.И. Альперт // Тези доповідей XIV-ої Української конференції з космічних досліджень (Київ, 8–12 вересня 2014). – Київ, 2014. – С. 152.

14. Національний стандарт України "Дистанційне зондування Землі з космосу" ДСТУ 4220-2003 "Терміни та визначення понять". – К.: Держспоживстандарт України, 2003. – 24 с.

15. Національний стандарт України "Дистанційне зондування Землі з космосу" ДСТУ 4758:2007 "Оброблення даних. Терміни та визначення понять" – К.: Держспоживстандарт України. 2007. – 21 с.

16. Кочуб Е.В. Анализ методов обработки материалов дистанционного зондирования Земли / Е.В. Кочуб, А.А. Топаз // Вестник Полоцкого государственного университета, 2012. – 132 с.

17. Марченко Н.А. Методика классификации объектов по данным дистанционного зондирования Земли / Н.А. Марченко, А.К. Авраменко // Системный анализ, управление и информационные технологии. Научное периодическое издание – Харьков: Харьковский политехнический институт. – 2008. – Вип. №26. – С. 145–149.

18. Кузнецов А.В. Сравнение алгоритмов управляемой поэлементной классификации гиперспектральных изображений / А.В. Кузнецов, В.В. Мясников // Компьютерная оптика. – Том 38. – №3. – 2014. – С. 494 – 502.

19. Congalton R.A. Review of Assessing the Accuracy of Classifications of Remotely Sensed Data / R.A. Congalton – Remote Sensing of Environment, 1991. – N 37. – P. 35–46.

20. Gopal S. Theory and Methods for Accuracy Assessment of Thematic Maps Using Fuzzy Sets/ S. Gopal, C. Woodcock. – Photogrammetric Engineering and Remote Sensing. – Februrary 1994. –Vol. 60. – №2. – P. 181–188.

21. Story M. Accuracy assessment: A user’s perspective / M. Story, R.G. Congalton – Photogrammetric Engineering and Remote Sensing. – 1986. – Vol.52. – P. 397–399.

22. Hardin P.J. Statistical significance and normalized confusion matrices / P.J. Hardin, J.M. Shumway// Photogrammetric Engineering and Remote Sensing. – 1997. – Vol. 63. – P. 735–740.

23. Hinton J.C. GIS and remote sensing integration for environmental applications / J.C. Hinton // Int. J. GIS 10. – 1996. – P. 877–890.

24. Congalton R.G. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques / R.G. Congalton, R.G. Oderwald, R.A. Mead // Photogrammetric Engineering and Remote Sensing. – 1983. – Vol. 1. - P. 1671–1678.

25. Лялько В.И. Использование данных спутниковой съемки для изучения природоресурсных проблем / В.И. Лялько, А.Д. Федоровский, М.А. Попов и др. // Космічні дослідження в Україні 2002–2004. – Київ, 2004. – С. 2–14.

26. Rosenfield G.H. A coefficient of agreement as a measure of thematic classification accuracy/ G.H. Rosenfield, K. Fitzpartrick-Lins // Photogrammetric Engineering and Remote Sensing. – 1986. – Vol. 52, P. 223–227.

27. Hegarat-Mascle S. Application of Dempster-Shafer Evidence Theory to Unsupervised Classification in Multisource Remote Sensing/ S. Hegarat-Mascle, I Bloch, D.Vidal–Madjar – IEEE Transactions on Geoscience and Remote Sensing. – July 1997. – Vol. 35. – №4. – P. 1018–1031.

28. Лурье И.К. Теория и практика цифровой обработки изображений / И.К. Лурье, А.Г. Косиков. – М.: Научный мир, 2003. – 356 с.

29. Shafer G. A Mathematical Theory of Evidence / G. Shafer – Princeton, NJ: Princeton University Press, 1976. – P.875–883.

30. Janssen L.L.F. Accuracy assessment of satellite derived land-cover data: A review / L.L.F. Janssen, F.J.M. van der Wel // Photogrammetric Engineering and Remote Sensing. –1994. – Vol. 60. – P. 419 – 426.

31. Irons J.R. The next Landsat satellite—the Landsat Data Continuity Mission / J.R. Irons, J.L. Dwyer, J.A. Barsi // Remote Sensing of Environment. – 2012. – Vol. 122. – P. 7–21.

32. Попов М.О. Сучасні погляди на інтерпретацію даних аерокосмічного дистанційного зондування Землі / М.О. Попов // Космічна наука і технологія. – 2002. - Т. 8. – № 2/3. - С. 110–115.

33. Congalton R.G. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices / R.G. Congalton, K. Green // CRC Press: Taylor & Francis Group, 1999. P. 130–137.

34. Cochran W.G. Sampling Techniques / W.G. Cochran // New York: John Wiley and Sons, 1977. – P. 421–428.

35. Abidi M.A. Data Fusion in Robotics and Machine Intelligence/ M.A. Abidi, R.C. Gonzalez – New York: Academic, 1992. – P. 562–569.

36. Чандра А.М. Дистанционное зондирование и геоинформационные системы. / А.М. Чандра, С.К. Гош // М: Техносфера, 2008. – 312 с.

37. Кельберт М.Я. Об оценивании параметра по загрязненной выборке / М.Я. Кельберт, А.И. Мерлушкин // Научно-техническая информация. Сер. 2. Информационные процессы и системы. – М.: ВИНИТИ РАН, 1994. – С. 20–25.

38. Zhu X. Bridging Local and Global Data Cleansing: Identifying Class Noise in Large, Distributed Data Datasets / X. Zhu, X. Wu, Q. Chen // Data Mining and Knowledge Discovery. – 2006. – Vol.12. P. 275–308. doi: 10.1007/s10618-005-0012-8.

39. Ruwet C. Impact of contamination on training and test error rates in statistical clustering/ C. Ruwet, G. Haesbroeck // Communications in Statistics - Simulation and Computation, Taylor & Francis: STM, Behavioural Science and Public Health Titles. – 2011. 40 (3). – P. 394 – 411. https://hal.archives-ouvertes.fr/hal-00666340

40. Yan D. Classification under Data Contamination with Application to Remote Sensing Image Mis-registration / D. Yan, P. Gong, A. Chen, L. Zhong // Journal: Computing Research Repository – CoRR. – 2011. – Vol. abs / 1101.3.

41. Попов М. Методы оптимизации числа спектральных каналов в задачах обработки и анализа данных дистанционного зондирования Земли/ М. Попов, С. Станкевич // Современные проблемы дистанционного зондирования земли из космоса. М.: ИКИ РАН, 2006. – Т. 2. – №1. – С. 61–63.

42. Beynon M.J. The Dempster-Shafer theory of evidence: an alternative approach to multicriteria decision modeling / M.J. Beynon, B. Curry, P. Morgan // Omega: Int. Journal of Management Science, 2000. – Vol. 28. – No. 1. – P. 37–50.

43. Уткин Л.В. Анализ риска и принятие решений при неполной информации/ Л.В. Уткин // СПб.: Наука, 2007. – 404 с.

44. Taroun A. Dempster-Shafer Theory of Evidence: Potential usage for decision making and risk analysis in construction project management / A. Taroun, J.B. Yang

// The Built & Human Environment Review. – 2011. – Vol. 4. – Special Issue 1. – P. 155–166.

45. Трифонова Т.А. Геоинформационые системы и дистанционное зондирование в экологических исследованиях / Т.А.Трифонова, Н.В. Мищенко, А.Н. Краснощеков – М.: Академпроект, 2005. – 349 с.

46. Попов М. Підхід до класифікації стану лісових ресурсів по багатоспектральних космічних зображеннях на основі принципу злиття даних та використання елементів ГІС-технологій / М. Попов, О. Рябоконенко, О. Петроченко // Ученые записки Таврического национального университета им. В.И. Вернадского, сер. География. – 2003. – Т. 16 (55). – №2. – С. 142–150.

47. Попов М. Шляхи отримання космічної інформації в інтересах національної безпеки та оборони / М. Попов // Наука і оборона. – 2002. – №2. – С. 38–50.

48. Bandos T.V. Classification of Hyperspectral Images with Regulized Linear Disctiminant Analysis / T.V. Bandos, L. Bruzzone, G. Camps-Valls // IEEE Transactions on Geoscience and Remote Sensing. – 2009. –Vol. 47. – № 3. – P. 862 – 873.

49. Бурштинська Х.В. Аерокосмічні знімальні системи / Х.В. Бурштинська, С.А. Станкевич // Львів: Львівська політехніка, 2013. – 316 с.

50. Bongasser M. Hyperspectral Remote Sensing: Principles and Applications / M. Borengasser, W.S. Hungate, R. Watkins // Boca Raton: CRC Press, 2008. – 119 p.

51. Багатоспектральні методи дистанційного зондування Землі в задачах природокористування / За ред. В.І. Лялька та М.О. Попова. – К.: Наук. думка, 2006. – 360 с.

52. Imaging Spectrometry: Basic Principles and Prospective Applications / Ed. by F.D. Van Der Meer, S.M. De Jong // Dodrecht: Kluwer Academic Publishers, 2001. – 404 p.

53. Козодеров В.В. Обработка и интерпретация данных гиперспектральных аэрокосмических измерений для дистанционной диагностики природно-

техногенных объектов / В.В. Козодеров, Т.В. Кондранин, О.Ю. Казанцев и др. // Исследование Земли из космоса (РФ). – 2009. – № 2. С. 36–54.

54. Hyperspectral Data Exploitation: Theory and Applications / Ed. by Ch.–I Chang // John Willey & Sons, 2007. – 430 p.

55. Varshey P.K. Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data / P.K. Varshey, M.K. Arora // Berlin-Heidelberg: Springer–Verlag, 2004. – 322 p.

56. Попов М.О. Гіперспектральна аерокосмічна інформація у виявленні та спостереженні об'єктів / М.О. Попов, С.А. Станкевич, В.Д. Молдован // Наука і оборона. – 2006. – № 3. – С. 25-35.

57. Chang C. –I. Hyperspectral Data Processing: Algorithm Design and Analysis / C. I. Chang // Hoboken, NJ: John Willey & Sons, 2013. – 1164 p.

58. Претт У. Цифровая обработка изображений. Кн. 1. / У. Претт // М.: Мир, 1982. –382 с.

59. Аковецкий В. И. Дешифрирование снимков. / Аковецкий В. И. // М.: Недра, 1983. – 320 с.

60. Gong P. Integrated Analysis of Spatial Data from Multiple Sources: Using Evidential Reasoning and Artificial Neural Network Techniques for Geological Mapping / P. Gong // Photogrammetric Engineering and Remote Sensing. – 1996. Vol. 62. – No. 5. – P. 513 – 523.

61. Mertikas P. Exemplifying the Theory of Evidence in Remote Sensing Image Classification / P. Mertikas, M.E. Zervakis // Int. Journal of Remote Sensing. – 2001. – Vol. 22, No. 6. – P.1081 –1095.

62. Lein J.K. Applying evidential reasoning methods to agricultural land cover classification / J.K. Lein // Int. Journal of Remote Sensing. – 2003. – Vol. 24, No. 21. – P. 4161– 4180.

63. Popov M.A. A Dempster-Shafer evidence theory-based approach to object classification on multispectral / hyperspectral images // M.A. Popov, M.V.

Topolnitskiy // Proceedings of the 10th International Conference IEEE on Digital Technologies (DT'2014). – Žilina, Slovakia, 9 –11 July 2014. – P. 296 –300.

64. Попов М.О. Классификация объектов на многоспектральных / гиперспектральных аэрокосмических изображениях на основе теории свидетельств Демпстера-Шейфера // М.О. Попов, М.В. Топольницький // Математичні машини і системи, 2014. – № 1. – С. 58–69.

65. Congalton G. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. 2nd Edition / G. Congalton, K. Green // Boca Raton, FL: CRC Press, 2009. – 183 p.

66. Дюличева Ю.Ю. О задачах фильтрации обучающих данных / Ю.Ю. Дюличева // Штучний інтелект, 2006. – № 2. – С. 65–71.

67. Dash M. Feature selection for classification / M. Dash, H. Liu // Intelligent Data Analysis, 1997. – No. 1. – P. 131–156.

68. Tso B. Classification Methods for Remotely Sensed Data / B. Tso // London: Tailor & Francis, 2001. – 332 p.

69. Станкевич С.А. Алгоритм статистичної класифікації об’єктів дистанційного спостереження за їх спектрально-топологічними характеристиками / С.А. Станкевич // Науковий вісник Національного гірничого університету. – 2006.– № 7.– С. 38–40.

70. Вентцель Е.С. Теория вероятностей. / Е.С. Вентцель // М.: Высш. шк., 1999. – 576 с.

71. Еремеев В. Современные гиперспектральные сенсоры и методы обработки гиперспектральных данных / В. Еремеев, И. Мордвинцев, Н. Платонов// Исследование Земли из космоса (РФ). – 2003. – № 6. – С. 80–90.

72. Лялько В.І. Багатоспектральні методи ДЗЗ в задачах природокористування / В.І. Лялько, М.О. Попов, О.Д. Федоровський та ін. // К.: Наук. думка, 2006. – 360 с.

73. Кронберг П. Дистанционное изучение Земли: Основы и методы дистанционных исследований в геологии / П. Кронберг // М.: Мир, 1988. – 343 с.

74. Козлов В. Избранные труды по теории планирования эксперимента и обратным задачам оптического зондирования / В. Козлов // С.-Пб.: СпбУ, 2000. –500 с.

75. Ферлаг Т. И. Дешифрирование многозональных аэрокосмических снимков. / Т. И. Ферлаг. – М.: Наука, 1988. – 650 с.

76. Гарбук С. Космические системы дистанционного зондирования Земли / С. Гарбук, В. Гершензон // М.: Изд–во А и Б, 1997. – 296 с.

77. Smets P. Constructing the pignistic probability function in a context of uncertainty / P. Smets, M. Henrion, R.D.Shachter, L.N. Kanal, J.F. Lemmer // Uncertainty in Artificial Intelligence. – North Holland, Amsterdam, 1990. – Vol. 5. – P. 29 – 40.

78. Lukin V. Application of filtering efficiency prediction to hyperspectral data pre-processing / V. Lukin, S.S. Krivenko, O.S. Rubel et al. // Український журнал дистанційного зондування Землі № 7, 2015. – P. 4–11.

79. Lukin V. Approaches to Automatic Data Processing in Hyperspectral Remote Sensing / V. Lukin, S. Abramov, N. Ponomarenko, S. Krivenko et al. // Telecommunications and Radio Engineering. – 2014, 73(13). – P. 1125–1139.

80. Li Q. Denoising of Hyperspectral Images Employing Two-Phase Matrix Decomposition / Q. Li, H. Li, Zh. Lu, Q. Lu, and W. Li // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. – 2014. – Vol. 7. – № 9. – P. 3742– 3754.

81. Method for Real Target Detection Based on Reduced Complexity Hyperspectral Processing / US Patent US 2010/0329512 A1 of 30 Dec. 2010 // G06K 9/00. Inventors: Y.Y. Nam, S.J. Hong, W.D. Cho, K.S. Park.

82. Arzuaga-Cruz E. Unsupervised Feature Extraction and Band Subset Selection Techniques Based on Relative Entropy Criteria for Hyperspectral Data Analysis / E.

Arzuaga-Cruz, L.O. Jimenez-Rodriguez, M. Vélez-Reyes // in: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery IX, S.S. Shen, P.E. Lewis (Eds), Proceedings of SPIE, Vol. 5093. 2003. P. 462-473.

83. Popov M.A. Method of Hyperspectral Satellite Image Classification under Contaminated Training Samples Based on Dempster-Shafer's Paradigm / M.A. Popov, S.I. Alpert, V.N. Podorvan, M.V. Topolnytskyi, S.I. Mieshkov // Central European Researchers Journal. – 2015. – Vol. 1. – №1. – P. 86-97. http://ceres-journal.eu/iss150101

84. Fisher K. System and Method for Progressive Band Selection for Hyperspectral Images / Patent US 8,406,469B2. March 26, 2013. G06K 9/00; G01J 3/40. Inventor: K. Fisher.

85. Datta A. Band elimination of hyperspectral imagery using partitioned band image correlation and capacitory discrimination / A. Datta, S. Ghosh, A. Ghosh // International Journal of Remote Sensing. – 2014. – Vol. 35. – № 2. – P. 554–577.

86. Айвазян С.А. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. Под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.

87. Minet J. Influence of band selection and target estimation error on the performance of the matched filter in hyperspectral imaging / J. Minet, J. Taboury, F. Goudail, M. Péalat, N. Roux, J. Lonnoy, Y. Ferrec // Applied Optics. Vol. 50. No. 22. Aug. 2011. P. 4276-4285.

88. Karami A. Band-Specific Shearlet-Based Hyperspectral Image Noise Reduction / A. Karami, R. Heylen, and P. Scheunders // IEEE Transactions on Geoscience and Remote Sensing. – 2015. – Vol. 53. – №9. – P. 1–13.

89. Ponomarenko N.N. Visual quality analysis for images degraded by different types of noise / N.N. Ponomarenko, V.V. Lukin, O.I. Ieremeyev, K. Egiazarian, J. Astola // Proceedings of SPIE Symposium on Electronic Imaging. – 2013. – Vol. 8665. – 12 p.

90. Donoho D.L. De-noising by soft-thresholding / D.L. Donoho // IEEE Transactions on Information Theory. – 1995. – Vol. 41. – № 3. – P. 613–627.

91. Кульбак С. Теория информации и статистика / С. Кульбак // М.: Наука, 1967. – 408 с.

92. Уиллиамс У.Т. Методы иерархической классификации / У.Т. Уиллиамс, Дж. Н. Ланс // Статистические методы для ЭВМ. Под ред. К. Энслейна, Э. Рэнстона, Г.С. Уилфа. – М.: Наука, 1986. С. 269-300.

93. Кононов В. Сравнительная оценка информативности цифровых аэрокосмических изображений высокого и низкого разрешения / В. Кононов, С. Станкевич // Ученые записки Таврического национального университета им. В.И. Вернадского. – 2004. – Т. 17. – №2. – С.88–95.

94. Эбелинг В. Физика процессов эволюции / В. Эбелинг, А. Энгель, Р. Файстель // М.: Эдиториал УРСС, 2001. – 328 с.

95. Zhang L. Representation, independence and combination of evidence in the Dempster-Shafer Theory of Evidence / L. Zhang // Advances in the Dempster-Shafer Theory of Evidence. Ed. by R.R. Yager, J. Kacprzyk, M. Ftdrizzi. – New York: John Wiley & Sons. – 1994. – P. 51–69.

96. Dubois D. Fuzzy Sets and Systems: Theoty and Applications / D. Dubois, H. Prade // Academic Press, New York, 1980. – P. 645 – 651.

97. Feller W. An introduction to probability theory and its applications. Vol. I. 3rd Ed. / W. Feller // New York: John Wiley and Sons. – 1968. – 509 p.

98. Popov M.A. Multispectral / hyperspectral satellite image classification algorithm based on the Dempster-Shafer evidence theory / M.A. Popov, M.V. Topolnytskyi // Journal of Information, Control and Management Systems. – 2014. – Vol. 12, No. 2. – P. 165 –176.

99. Лялько В.И. Состояние и перспективы развития аэрокосмических исследований в Украине / В.И. Лялько // Космічна наука і технологія. – 2002. – № 1. – С. 7–14.

100. Barnett J.A. Calculating Dempster–Shafer plausibility/ J.A. Barnett – IEEE Trans. Pattern Anal. Machine Intell., June 1991. – Vol.13. – P. 599 – 602.

101. Renyi A. Probability theory / A. Renyi. // Amsterdam, North – Holland Pub. Co, 1970. – 670 p.

102. Шовенгердт Р.А. Дистанционное зондирование. Модели и методы обработки изображений / Р.А. Шовенгердт // М.: Техносфера, 2010. – 560 с.

103. Ball G.H. ISODATA, A Novel Method of Data Analysis and Pattern Classification: Technical Report / G.H. Ball, D.J. Hall // Menlo Park, California: Stanford Research Institute. – April 1965. – 79 p.

104. MacQueen J.B. Some Methods for classification and Analysis of Multivariate Observations / J.B. MacQueen // Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. – 1967. – P. 281–297.

105. Романов А.А. Эффективность непараметрических классификаторов в условиях лимитированной обучающей выборки / А.А. Романов, К.А. Рубанов // Journal of Siberian Federal University: Engineering & Technologies. – 2012. 5. – P. 495–506.

106. Dempster A. Upper and lower probabilities induced by a multivalued mapping/A. Dempster // Annals of Mathematical Statistics. – 1967. 38. – P. 325–339.

107. Багатоспектральні методи ДЗЗ в задачах природокористування / За ред. акад. НАН України В.І. Лялька і д. т. н. М.О. Попова // К.: Наук. думка, 2006. - 360 с.

108. Зайцев О.В. Модель інтеграції даних від різних інформаційних джерел на основі теорії свідоцтв / О.В. Зайцев, Т.В. Савченко, С.І. Глухов // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ. – 2013. – Вип. №43 – С.142–145.

109. Зайцев О.В. Модель комплексування інформації отриманої від різних джерел на основі теорії свідчень Демпстера-Шейфера / О.В. Зайцев // Системи управління, навігації та зв’язку. Випуск 1 (33). Наукове періодичне видання – Полтава: Полтавський національний технічний університет ім. Юрія Кондратюка. – 2015. – № 1 (33). – С. 100–102.

110. Резник А.М. Развивающиеся системы / А.М. Резник // Інженерія програмного забезпечення. – 2012. – №4. – С. 5–13.

111. Reznik A.M. Dynamic Neural Cortical Ensembles Model / A.M. Reznik, D.A. Dziuba // Optical Memory & Neural Networks. – 2012. – №3. – Vol. 22. – P. 159 – 165.

112. Попов М.О. Оптимізація кластерної процедури визначення репрезентативної комбінації зональних зображень при класифікуванні гіперспектральних аерокосмічних знімків / М.О. Попов, С.А. Станкевич, С.П. Ковальчук, М.В. Топольницький, В.М. Подорван, В.В. Пилипчук, С.І. Альперт // Abstracts of 16th Ukrainian Conference on Space Research (Odesa, 2016, August, 22–27). – Kyiv: SSAU, 2016. – P. 219.

ДОДАТКИ

Додаток 1

Розрахункові дані для оцінки точності запропонованого методу контрольованого класифікування космічних зображень в умовах малих забруднених навчальних вибірок на прикладі гіперспектрального знімку EO1H1810252014211110KF, отриманого 220-канальною апаратурою Hyperion супутника EO-1 30 липня 2014 р.

Таблиця Д.1– Кількість пікселів у навчальних вибірках виділених класів для різних варіантів (ступенів) забруднення гіперспектрального знімку EO1H1810252014211110KF

Клас	**Ступень забруднення пікселів навчальних підвибірок**							
	0%	**10%**	**20%**	**30%**	**35%**	**40%**	**45%**	**50%**
Вода	335	338	344	340	364	359	341	381
Сосновий ліс	397	397	398	385	390	408	411	391
Листяний ліс	344	353	358	383	369	373	368	364
Злакові поля	525	511	486	468	470	458	458	437
Овочеві поля	440	434	426	425	408	416	415	426
Прибрані поля	315	323	344	355	355	342	363	357
Всього	2356	2356	2356	2356	2356	2356	2356	2356

Таблиця Д.2 – Загальна точність класифікування пікселів еталонних ділянок методами DST та SVM при різному ступені забруднення τ підвибірок виділених класів

Класи	**τ = 0%**		**τ =10%**		**τ =20%**		**τ =30%**		**τ =35%**		**τ =40%**		**τ =45%**		**τ =50%**	
	DST	**SVM**	**DST**	**SVM**	**DST**	**SVM**	**DST**	**SVM**	**DST**	**SVM**	**DST**	**SVM**	**DST**	**SVM**	**DST**	**SVM**
Вода	0,762	0,730	0,735	0,712	0,765	0,717	0,792	0,717	0,814	0,732	0,827	0,758	0,836	0,748	0,873	0,812
Сосновий ліс	0,857	0,477	0,878	0,853	0,929	0,840	0,933	0,835	0,939	0,920	0,861	0,932	0,649	0,947	0,383	0,920
Листяний ліс	0,890	0,994	0,853	0,956	0,869	0,921	0,831	0,934	0,858	0,936	0,756	0,886	0,694	0,947	0,572	0,000
Злакові поля	0,871	0,790	0,888	0,901	0,851	0,917	0,864	0,936	0,779	0,951	0,735	0,953	0,716	0,961	0,678	0,963
Овочеві поля	0,924	0,928	0,915	0,939	0,892	0,917	0,834	0,847	0,695	0,427	0,654	0,572	0,559	0,133	0,424	0,034
Прибрані поля	0,815	0,722	0,961	0,742	0,967	0,805	0,968	0,857	0,970	0,846	0,967	0,854	0,971	0,810	0,973	0,368
Загальна точність	0,859	0,778	0,854	0,816	0,880	0,858	0,867	0,854	0,833	0,777	0,789	0,809	0,772	0,716	0,629	0,475

Додаток 2

Матриці помилок та показники точності класифікування виділення класів при різному ступені забруднення навчальних підвибірок

Таблиця Д.3 – Варіант 1: забруднення 0%

а) метод класифікування DST; загальна точність 0,859

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1062	0	0	0	0	0	1062
	Сосновий ліс	162	1260	27	23	14	6	1492
	Листяний ліс	77	133	1372	18	3	26	1629
	Злакові поля	17	5	8	1175	140	74	1419
	Овочеві поля	76	73	134	102	1902	143	2430
	Прибрані поля	0	0	0	31	0	1098	1129
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,762	0,857	0,890	0,871	0,924	0,815	

б) метод класифікування SVM; загальна точність 0,778

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
SVM	Вода	1018	0	0	0	0	0	1018
	Сосновий ліс	196	702	25	9	0	3	935
	Листяний ліс	89	651	1455	30	4	52	2281
	Злакові поля	11	0	0	1066	144	219	1440
	Овочеві поля	80	118	61	78	1911	101	2349
	Прибрані поля	0	0	0	166	0	972	1138
	Сума елементів стовпця	1394	1471	1541	1349	2059	1347	9161
	Точність	0,730	0,477	0,994	0,790	0,928	0,722	

Таблиця Д.4 – Варіант 2: забруднення 10%

а) метод класифікування DST; загальна точність 0,854

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1024	0	0	0	0	0	1024
	Сосновий ліс	186	1292	26	19	0	7	1530
	Листяний ліс	63	42	1315	21	5	20	1466
	Злакові поля	14	3	22	1198	168	8	1413
	Овочеві поля	104	132	156	86	1884	17	2379
	Прибрані поля	3	2	22	25	2	1295	1349
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,735	0,878	0,853	0,888	0,915	0,961	

б) метод класифікування SVM; загальна точність 0,816

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
SVM	Вода	993	0	0	0	0	0	993
	Сосновий ліс	222	857	25	9	0	5	1118
	Листяний ліс	91	514	1473	31	4	46	2159
	Злакові поля	9	0	0	1216	121	138	1484
	Овочеві поля	79	100	43	93	1934	158	2407
	Прибрані поля	0	0	0	0	0	1000	1000
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,712	0,853	0,956	0,901	0,939	0,742	

Таблиця Д.5 – Варіант 3: забруднення 20%

а) метод класифікування DST; загальна точність 0,880

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1067	0	0	0	0	0	1067
	Сосновий ліс	172	1366	45	23	1	9	1616
	Листяний ліс	54	27	1339	20	4	23	1467
	Злакові поля	13	3	22	1148	210	4	1400
	Овочеві поля	81	73	82	59	1837	9	2141
	Прибрані поля	7	2	53	99	7	1302	1470
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,765	0,929	0,869	0,851	0,892	0,967	

б) метод класифікування SVM; загальна точність 0,858

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
SVM	Вода	999	0	0	0	0	0	999
	Сосновий ліс	241	1235	41	15	0	6	1538
	Листяний ліс	59	93	1420	23	3	43	1641
	Злакові поля	12	0	0	1237	167	98	1514
	Овочеві поля	83	143	80	74	1889	116	2385
	Прибрані поля	0	0	0	0	0	1084	1084
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,717	0,840	0,921	0,917	0,917	0,805	

Таблиця Д.6 – Варіант 4: забруднення 30%

а) метод класифікування DST; загальна точність 0,867

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1104	2	0	0	0	0	1106
	Сосновий ліс	134	1372	50	23	3	10	1592
	Листяний ліс	58	31	1280	16	6	20	1411
	Злакові поля	21	3	11	1166	303	6	1510
	Овочеві поля	69	60	101	53	1718	7	2008
	Прибрані поля	8	3	99	91	29	1304	1534
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,792	0,933	0,831	0,864	0,834	0,968	

б) метод класифікування SVM; загальна точність 0,854

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
SVM	Вода	999	0	0	0	0	0	999
	Сосновий ліс	239	1228	41	14	0	5	1527
	Листяний ліс	54	85	1439	25	3	53	1659
	Злакові поля	22	3	0	1262	312	78	1677
	Овочеві поля	80	155	61	48	1744	56	2144
	Прибрані поля	0	0	0	0	0	1155	1155
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,717	0,835	0,934	0,936	0,847	0,857	

Таблиця Д.7– Варіант 5: забруднення 35%

а) метод класифікування DST; загальна точність 0,833

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1135	15	0	0	1	0	1151
	Сосновий ліс	118	1381	56	27	72	11	1665
	Листяний ліс	63	41	1322	13	101	23	1563
	Злакові поля	18	0	5	1051	345	1	1420
	Овочеві поля	47	29	13	44	1432	5	1570
	Прибрані поля	13	5	145	214	108	1307	1792
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,814	0,939	0,858	0,779	0,695	0,970	

б) метод класифікування SVM; загальна точність 0,777

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
SVM	Вода	1021	0	0	0	0	0	1021
	Сосновий ліс	257	1353	76	18	3	14	1721
	Листяний ліс	32	22	1442	25	3	50	1574
	Злакові поля	44	11	0	1283	1174	106	2618
	Овочеві поля	40	85	23	23	879	37	1087
	Прибрані поля	0	0	0	0	0	1140	1140
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,732	0,920	0,936	0,951	0,427	0,846	

Таблиця Д.8 – Варіант 6: забруднення 40%

а) метод класифікування DST; загальна точність 0,789

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1153	136	3	2	2	0	1296
	Сосновий ліс	108	1267	67	42	146	13	1643
	Листяний ліс	46	20	1165	14	4	21	1270
	Злакові поля	34	7	143	992	539	8	1723
	Овочеві поля	45	38	69	74	1347	2	1575
	Прибрані поля	8	3	94	225	21	1303	1654
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,827	0,861	0,756	0,735	0,654	0,967	

б) метод класифікування SVM; загальна точність 0,809

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
SVM	Вода	1057	0	0	0	0	0	1057
	Сосновий ліс	235	1371	130	18	6	17	1777
	Листяний ліс	15	11	1366	24	0	43	1459
	Злакові поля	46	12	4	1286	876	127	2351
	Овочеві поля	41	77	41	21	1177	9	1366
	Прибрані поля	0	0	0	0	0	1151	1151
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,758	0,932	0,886	0,953	0,572	0,854	

Таблиця Д.9 – Варіант 7: забруднення 45%

а) метод класифікування DST; загальна точність 0,772

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1165	449	4	8	2	0	1628
	Сосновий ліс	122	955	86	23	335	13	1534
	Листяний ліс	24	15	1070	10	1	15	1135
	Злакові поля	24	3	63	966	446	2	1504
	Овочеві поля	41	42	124	26	1150	9	1392
	Прибрані поля	18	7	194	316	125	1308	1968
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,836	0,649	0,694	0,716	0,559	0,971	

б) метод класифікування SVM; загальна точність 0,716

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
SVM	Вода	1043	0	0	0	0	0	1043
	Сосновий ліс	218	1393	59	20	16	13	1719
	Листяний ліс	48	17	1460	25	4	61	1615
	Злакові поля	61	21	7	1297	1765	178	3329
	Овочеві поля	24	40	15	7	274	4	364
	Прибрані поля	0	0	0	0	0	1091	1091
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,748	0,947	0,947	0,961	0,133	0,810	

Таблиця Д.10 –Варіант 8: забруднення 50%

а) метод класифікування DST; загальна точність 0,629

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сума елементів рядку n_{i+}
DST	Вода	1217	853	27	12	93	4	2206
	Сосновий ліс	71	564	101	20	331	13	1100
	Листяний ліс	17	10	882	9	3	13	934
	Злакові поля	30	3	111	914	552	2	1612
	Овочеві поля	39	32	112	15	874	5	1077
	Прибрані поля	20	9	308	379	206	1310	2232
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,873	0,383	0,572	0,678	0,424	0,973	

б) метод класифікування SVM; загальна точність 0,475

		Завіркові дані						
	Класи	Вода	Сосновий ліс	Листяний ліс	Злакові поля	Овочеві поля	Прибрані поля	Сумма элементов строки n_{i+}
SVM	Вода	1132	0	0	0	0	0	1132
	Сосновий ліс	162	1353	1188	34	4	24	2765
	Листяний ліс	0	0	0	0	0	0	0
	Злакові поля	64	27	23	1299	1986	793	4192
	Овочеві поля	36	91	330	16	69	34	576
	Прибрані поля	0	0	0	0	0	496	496
	Сума елементів стовпця n_{+j}	1394	1471	1541	1349	2059	1347	9161
	Точність	0,812	0,920	0,000	0,963	0,034	0,368	

Додаток 3

Значення показників точності класифікування для методу DST та методу SVM в залежності від ступеня забруднення навчальних підвибірок кожного виділеного класу

Таблиця Д.11 – Дані для побудови графіків зміни точності класифікування пікселів еталонних ділянок в залежності від ступеня забруднення навчальних підвибірок кожного виділеного класу

Варіант	Ступінь Забруднення	Метод DST							Метод SVM						
		Класи							Класи						
		1	2	3	4	5	6	Загальна	1	2	3	4	5	6	Загальна
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**
1	0%	0,762	0,857	0,890	0,871	0,924	0,815	0,859	0,730	0,477	0,994	0,790	0,928	0,722	0,778
2	10%	0,735	0,878	0,853	0,888	0,915	0,961	0,854	0,712	0,853	0,956	0,901	0,939	0,742	0,816
3	20%	0,765	0,929	0,869	0,851	0,892	0,967	0,880	0,717	0,840	0,921	0,917	0,917	0,805	0,858
4	30%	0,792	0,933	0,831	0,864	0,834	0,968	0,867	0,717	0,835	0,934	0,936	0,847	0,857	0,854
5	35%	0,814	0,939	0,858	0,779	0,695	0,970	0,833	0,732	0,920	0,936	0,951	0,427	0,846	0,777
6	40%	0,827	0,861	0,756	0,735	0,654	0,967	0,789	0,758	0,932	0,886	0,953	0,572	0,854	0,809
7	45%	0,836	0,649	0,694	0,716	0,559	0,971	0,772	0,748	0,947	0,947	0,961	0,133	0,810	0,716
8	50%	0,873	0,383	0,572	0,678	0,424	0,973	0,629	0,812	0,920	0,000	0,963	0,034	0,368	0,475

Таблиця Д.12 – Зміна значень капа-коефіцієнтів для методу DST та методу SVM в залежності від ступеня забруднення навчальних підвибірок кожного виділеного класу

Варіант	Ступінь забруднення	Метод DST	Метод SVM
A	**B**	**C**	**D**
1	0%	0,829	0,731
2	10%	0,848	0,781
3	20%	0,855	0,829
4	30%	0,84	0,824
5	35%	0,799	0,734
6	40%	0,747	0,771
7	45%	0,667	0,663
8	50%	0,558	0,377

Таблиця Д.13 – Зміна значень точності користувача в залежності від ступеня забруднення навчальних підвибірок кожного виділеного класу

Ступінь забруднення	Метод DST						Метод SVM					
	Класи						Класи					
	1	2	3	4	5	6	1	2	3	4	5	6
0%	1	0,845	0,842	0,828	0,783	0,973	1	0,751	0,638	0,74	0,814	0,854
10%	1	0,844	0,897	0,848	0,792	0,96	1	0,767	0,682	0,819	0,803	1
20%	1	0,845	0,913	0,82	0,858	0,886	1	0,803	0,865	0,817	0,792	1
30%	0,998	0,862	0,907	0,772	0,856	0,85	1	0,804	0,867	0,753	0,813	1
35%	0,99	0,829	0,846	0,74	0,912	0,729	1	0,786	0,916	0,49	0,809	1
40%	0,89	0,771	0,917	0,576	0,855	0,788	1	0,772	0,936	0,547	0,862	1
45%	0,716	0,623	0,943	0,642	0,826	0,665	1	0,81	0,904	0,39	0,753	1
50%	0,552	0,513	0,944	0,567	0,812	0,587	1	0,489	0,000	0,31	0,12	1

Таблиця Д.14 – Зміна значень часткових капа-коефіцієнтів для методу DST та методу SVM в залежності від ступеня забруднення навчальних підвибірок кожного виділеного класу

Ступінь забруднення	Метод DST						Метод SVM					
	Класи						Класи					
	1	2	3	4	5	6	1	2	3	4	5	6
0%	1	0,815	0,81	0,798	0,72	0,968	1	0,703	0,565	0,695	0,759	0,829
10%	1	0,815	0,876	0,822	0,732	0,953	1	0,722	0,618	0,788	0,747	1
20%	1	0,816	0,895	0,789	0,817	0,866	1	0,765	0,838	0,785	0,732	1
30%	1	0,835	0,888	0,733	0,814	0,824	1	0,767	0,841	0,71	0,759	1
35%	1	0,797	0,815	0,695	0,887	0,683	1	0,745	0,899	0,402	0,753	1
40%	1	0,727	0,901	0,502	0,813	0,751	1	0,728	0,923	0,469	0,822	1
45%	1	0,55	0,931	0,581	0,776	0,607	1	0,774	0,885	0,284	0,681	1
50%	0	0,42	0,933	0,492	0,757	0,516	1	0,392	0,000	0,191	0,135	1

Додаток 4

Розподіл помилок пропуску та ототожнення для кожного із розглянутих класів при різному ступені забруднення навчальних вибірок

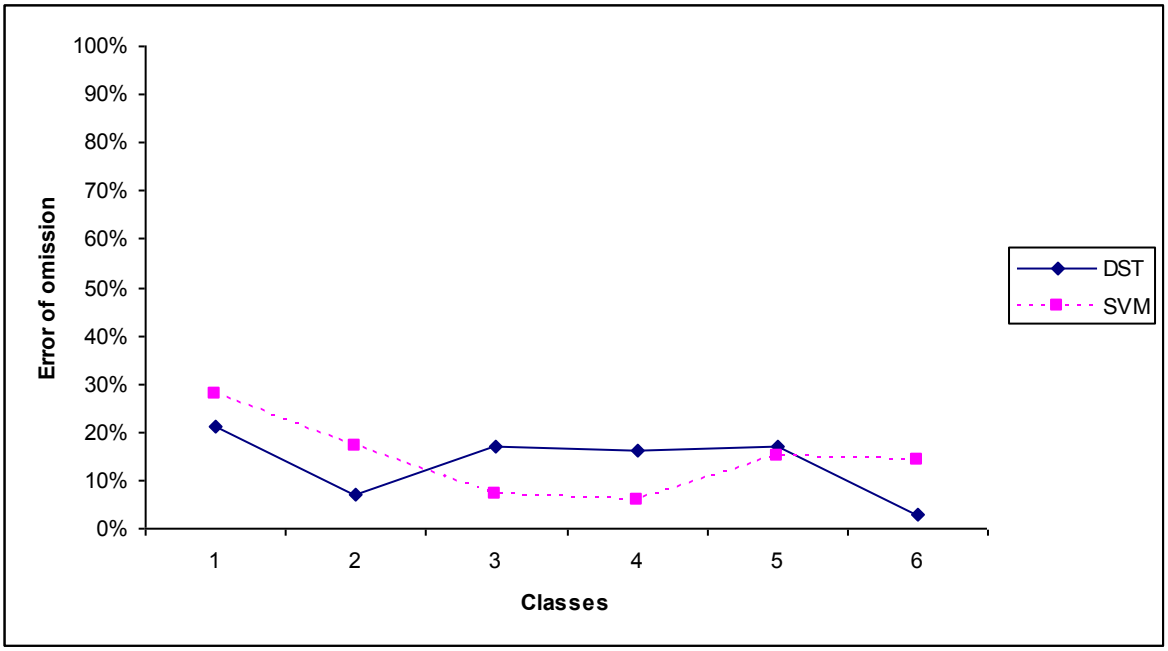


Рис. Д.1. Розподіл помилок пропуску для кожного із розглянутих класів при забрудненні 30%

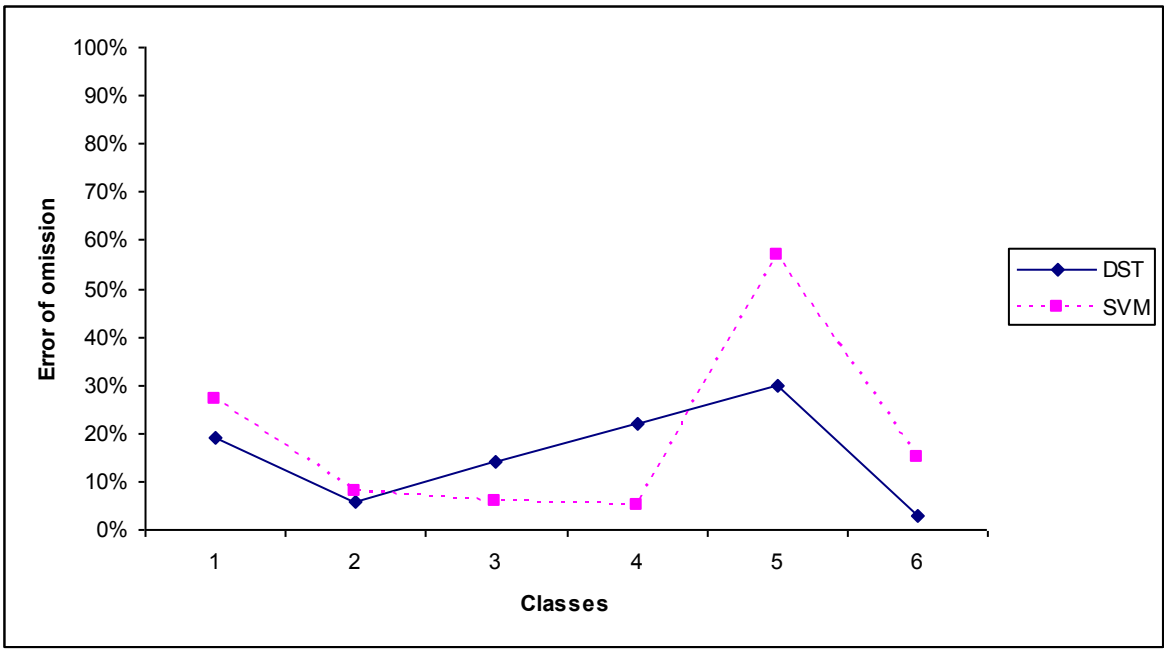


Рис. Д.2. Розподіл помилок пропуску для кожного із розглянутих класів при забрудненні 35%

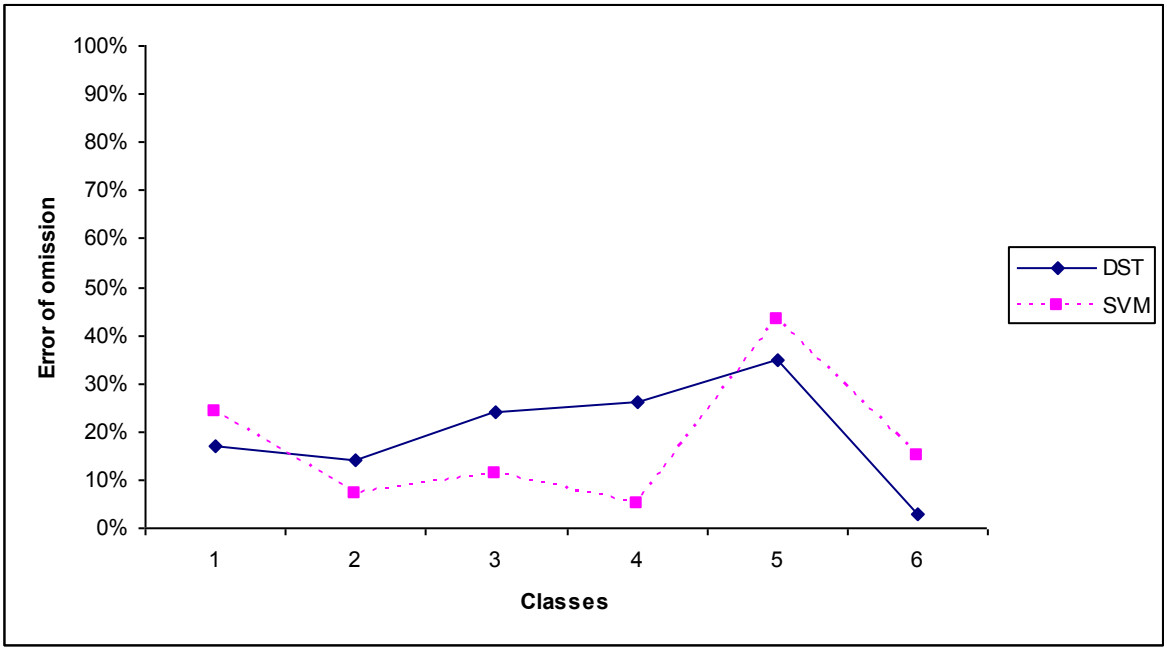


Рис. Д.3. Розподіл помилок пропуску для кожного із розглянутих класів при забрудненні 40%

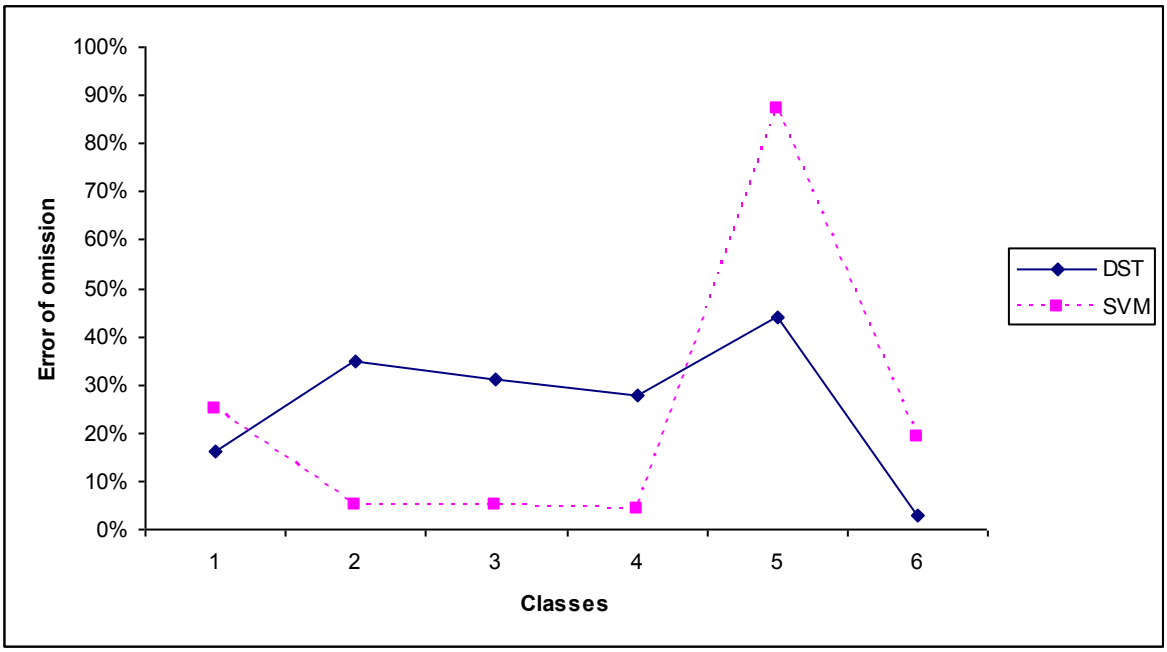


Рис. Д.4. Розподіл помилок пропуску для кожного із розглянутих класів при забрудненні 45%

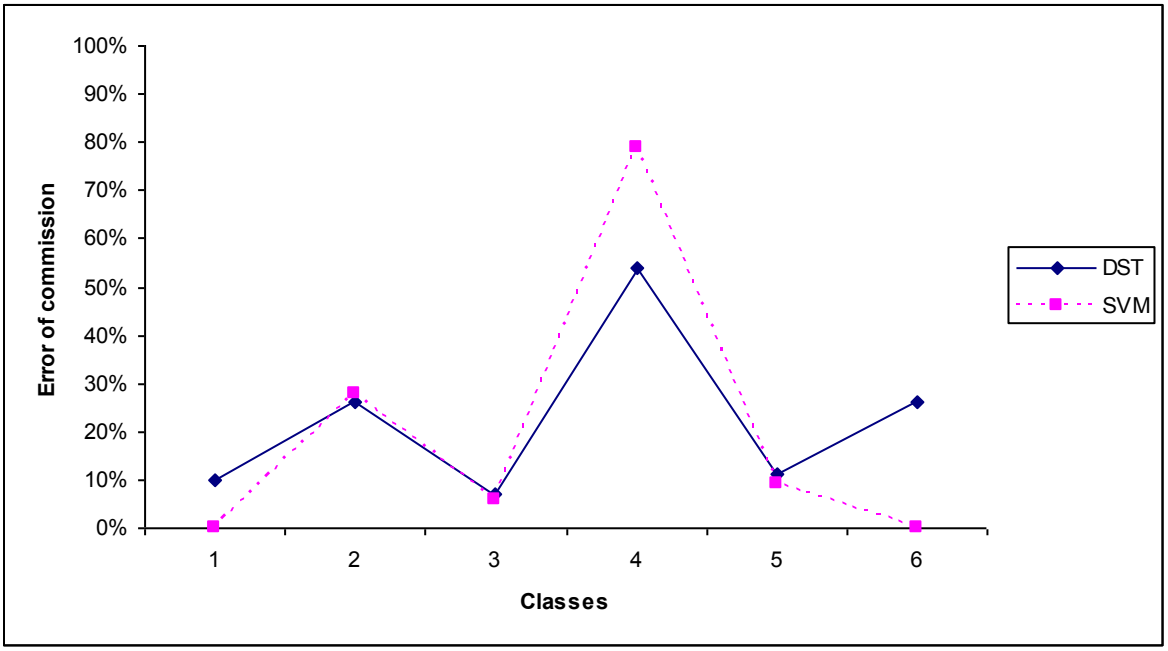


Рис. Д.5. Розподіл помилок ототожнення із класом при забрудненні 40%

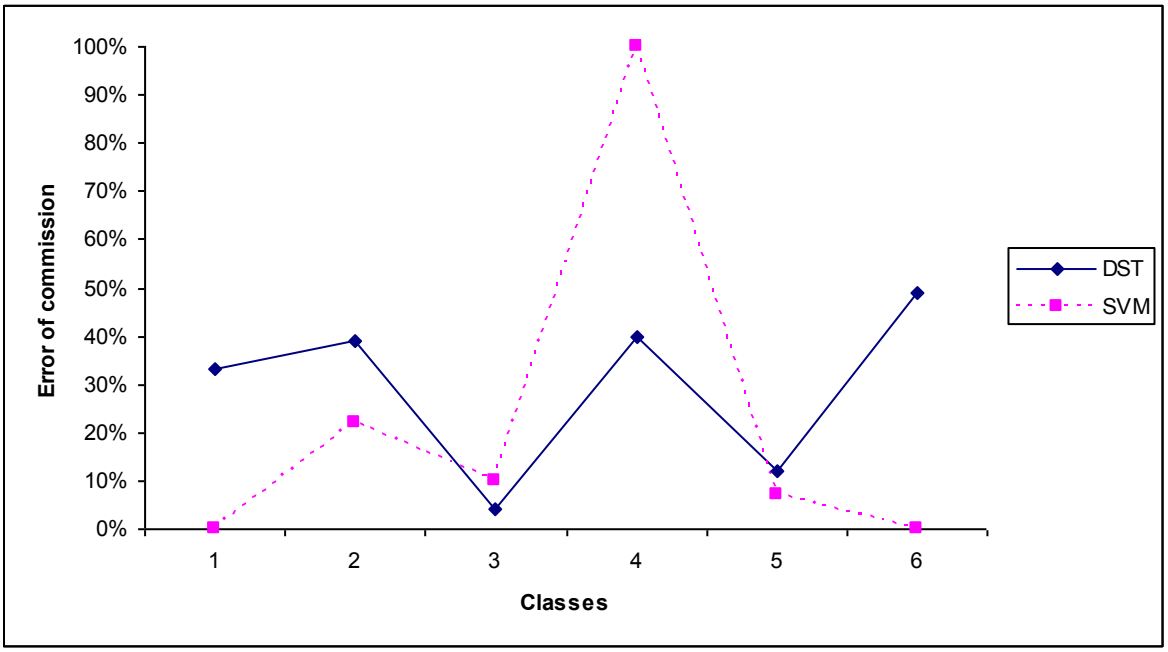


Рис. Д.6. Розподіл помилок ототожнення із класом при забрудненні 45%