

ОГЛЯД МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЇ ВЕЛИКИХ ОБСЯГІВ СУПУТНИКОВИХ ДАНИХ

М.С. ЛАВРЕНЮК, О.М. НОВІКОВ

Анотація. З появою у вільному доступі великих обсягів супутникових даних дедалі більшої актуальності набуває розвиток методів машинного навчання на підставі геопросторових даних, зокрема, супутникових. Розглянуто основні методи машинного навчання і проаналізовано особливості та результати їх застосування до класифікації земного покриття за супутниковими даними високого розрізнення. Особливу увагу приділено глибинним архітектурам, зокрема згортковим нейронним мережам, що натеper є найбільш потужним і точним методом для розпізнавання візуальних образів. Визначено основні переваги методів глибинного навчання над традиційними підходами до задач класифікації, що використовувались протягом останніх десятиліть і ґрунтувались на експертних знаннях для виокремлення ознак із вхідних даних.

Ключові слова: машинне навчання, глибинне навчання, згорткова нейронна мережа, класифікація великих обсягів даних.

ОГЛЯД МЕТОДІВ МАШИННОГО НАВЧАННЯ

Машинним навчанням називається галузь комп'ютерних наук, яка вивчає методи навчання комп'ютеризованих систем на підставі даних без програмування їх поведінки [1]. Методи машинного навчання (machine-learning methods) відіграють важливу роль у багатьох аспектах сучасного суспільства: від веб-пошуку до фільтрації контенту в соціальних мережах. Системи на базі методів машинного навчання використовуються в системах машинного зору, для ідентифікації об'єктів на зображеннях, аналізу людської мови і текстів тощо [2, 3].

Традиційно розпізнавання образів (або їх класифікація) здійснювалось на основі інформаційних ознак. Отже, побудова систем розпізнавання образів (pattern-recognition) або систем, в основу яких покладено методи машинного навчання, потребувала експертних знань для розроблення методів та правил виокремлення ознак (feature extraction). Виокремлення ознак — це перетворення початкових «сирих» даних (таких як значення піксела на зображенні) у придатне подання (вектор ознак), з якого система навчання (класифікатор) може виявити і класифікувати образи, що подаються на вхід. Такі методи машинного навчання обмежені в можливостях обробляти природні дані в початковому вигляді [4].

Існуючі типи класифікаторів систематизовано за різними критеріями і їх короткі характеристики наведено в табл. 1 [5].

Таблиця 1. Різновидності підходів до класифікації залежно від критеріїв

Критерій	Тип	Короткий опис
Використання або невикористання навчальних даних	Класифікація з учителем	За підходами до класифікації з учителем вхідні дані поділяють, використовуючи набір зразків як навчальні дані
	Класифікація без учителя	Підходи до класифікації без учителя, відомі як кластеризація, не беруть до уваги мітки навчальних даних для класифікації вхідних даних
	Напівавтоматичне навчання	Навчання в напівавтоматичних підходах відбувається з використанням даних як з мітками, так і без них
Урахування або неврахування будь-якого припущення про розподіл вихідних даних	Параметричні класифікатори	Параметричні класифікатори ґрунтуються на припущенні, що функція щільності ймовірності для кожного класу відома
	Непараметричні класифікатори	Непараметричні класифікатори не обмежуються жодними припущеннями про розподіл вхідних даних
Розгляд одного класифікатора або ансамблю	Один класифікатор	Використовується єдиний класифікатор для призначення мітки для об'єкта
	Ансамбль класифікаторів	Під час визначення мітки для об'єкта враховуються результати кількох (ансамблю) класифікаторів
Використання або невикористання технології жорсткого поділу, де кожен об'єкт належить лише одному кластеру	Жорсткий класифікатор	Технології жорсткої класифікації не враховують подальші зміни різних класів
	М'який (нечіткий) класифікатор	Нечіткі класифікатори моделюють поступові граничні зміни, забезпечуючи оцінку ступеня подібності всіх класів
Видача класифікатором розподілу ймовірності належності до всіх класів	Імовірнісний класифікатор	Класифікатор здатен для заданого зразка оцінити розподіл ймовірності на множині класів
	Неймовірнісний класифікатор	Підхід визначає лише найбільш придатний клас для вхідного образу

Методи машинного навчання поділяють на дві основні категорії: навчання з учителем (supervised learning) та навчання без учителя (unsupervised learning). Методи навчання з учителем поділяють вхідні дані на набір наперед заданих класів. Для навчання такого класифікатора потрібна навчальна вибірка, яка містить марковані зразки різних класів. Методи навчання без учителя не потребують навчальних даних, проте вони не ставлять у відповідність вхідним даним певний клас, а лише вивчають закономірності у вхідних даних та поділяють вхідні дані на схожі між собою групи (кластери) [5]. Оскільки завданням роботи є аналіз методів класифікації, то надалі розглядатимемо лише методи навчання з учителем.

Класифікатори можна поділити на параметричні та непараметричні. До параметричних належать, наприклад, метод максимальної правдоподібності (maximum likelihood), оскільки він працює на припущенні, що функція

щільності ймовірності для кожного з класів подається гаусовим розподілом [6]. Непараметричні класифікатори, у свою чергу, не ґрунтуються на жодних припущеннях про розподіл вхідних даних. Ураховуючи той факт, що в більшості випадків функція розподілу невідома, непараметричні класифікатори набули значно більшого поширення.

Важливою властивістю класифікаторів є можливість не лише належності вхідних даних до певного класу (виходу класифікатора), а і визначення ймовірності належності до кожного з класів, на основі якої легко обрати найбільш достовірний клас [5]. Таку особливість має, наприклад, логістична регресія (logistic regression). Отримані ймовірності для кожного з класів можуть бути використані для постоброблення результатів класифікації, наприклад для об'єднання виходів з різних моделей в ансамблі, фільтрації шуму і т.ін.

Найбільш поширеними методами машинного навчання для задач класифікації [7] є штучні нейронні мережі (artificial neural network) [1], логістична регресія [1], метод опорних векторів Support Vector Machine (SVM) [1] та random forest [8].

Мета роботи — огляд сучасних методів машинного навчання, визначення їх переваг та недоліків для класифікації великих обсягів супутникових даних.

Штучна нейронна мережа

Штучна нейронна мережа — математична модель, а також її програмна або апаратна реалізація, побудована за принципом організації та функціонування біологічних нейронних мереж – мереж нервових клітин живого організму. Це поняття виникло у ході вивчення процесів, що перебігають у мозку, зокрема під час спроби моделювання цих процесів [9]. З математичного погляду навчання нейронних мереж — це багатопараметрична задача нелінійної оптимізації. Проблема навчання багатосарових нейронних мереж була вирішена в середині 1980-х років методом зворотного поширення помилки (backpropagation) [10]. Це ітеративний градієнтний алгоритм, який використовується з метою мінімізації помилки роботи нейронної мережі, що забезпечує отримання бажаного виходу. Для визначення якості роботи нейронної мережі використовують функцію втрат (loss function). Зазвичай за таку функцію обирають евклідову відстань, середньоквадратичну похибку або функцію кросентропії [11]. Мережа вважається навченою, якщо функція втрат набуває мінімального значення. Основна ідея цього методу полягає в поширенні сигналів помилки від виходів мережі до її входів у напрямку, зворотному прямому поширенню сигналів у звичайному режимі роботи мережі. Процедура зворотного поширення помилки можна застосувати кілька разів, щоб поширити градієнти через усі шари, починаючи з виходу (результату прямого проходження нейронної мережі) і до входів, що подаються в мережу. У процесі навчання нейронної мережі ваги зв'язків між нейронами коригуються на основі методу градієнтного спуску (gradient descent). На практиці зазвичай використовують модифікацію цього методу, коли процедура градієнтного спуску застосовується до груп навчальних прикладів. Такий підхід називається методом стохастичного градієнта (stochastic gradient descent), що значно пришвидшує час навчання нейронної мережі. Існують й

інші методи оптимізації: метод найменших квадратів (алгоритм Левенберга–Марквардта та алгоритм Ньютона–Гауса), квазіньютонівські методи (Broyden–Fletcher–Goldfarb–Shanno (BFGS)), метод спряжених градієнтів тощо [12]. У 1986 р. Румельхарт запропонував метод моментів, що запам'ятовує зміну ваг на кожній ітерації і враховує її в подальшому коригуванні ваг нейронної мережі [13]. На відміну від методу стохастичного градієнта підхід намагається зберегти той же напрямок руху протягом коригування вагових коефіцієнтів, запобігаючи коливанням. Найсучаснішими методами оптимізації є AdaGrad (adaptive gradient algorithm), RMSProp (Root Mean Square Propagation), калманівський метод стохастичного градієнтного спуску (Kalman-based Stochastic Gradient Descent — KSGD) та Adam (Adaptive Moment Estimation) [14–16]. Ці методи є модифікаціями методу стохастичного градієнту і передбачають зміни коефіцієнта швидкості навчання під час навчання нейронної мережі.

У практичних завданнях навчання з учителем найчастіше використовуються нейронні мережі прямого поширення, такі як багатошаровий перцептрон (multilayer perceptron). Проходячи від одного шару до наступного, приховані нейрони обчислюють зважену суму входів на них з попереднього шару і застосовують нелінійну функцію – функцію активації до отриманого результату (рис. 1) [4]. Важливим критерієм функції активації є її диференційовність. Найбільш відомі функції активації: гіперболічний тангенс та сигмоїдна.

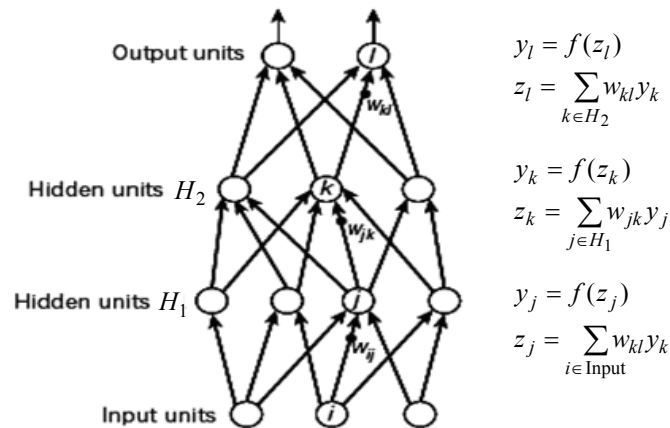


Рис. 1. Архітектура багатошарового перцептрона та принцип його роботи [4]

За теоремою про універсальну апроксимацію, доведена Хорніком у 1989 р., нейронна мережа з одним прихованим шаром може апроксимувати будь-яку неперервну функцію багатьох змінних з будь-якою точністю [17, 18]. Тому для пришвидшення навчання мережі більшість дослідників використовують архітектуру з одним прихованим шаром.

Перевагами класифікатора є його непараметричність — нейронна мережа не потребує жодних попередніх знань про розподіл вхідних даних та можливості їх поділу, які не є лінійно подільними.

Недоліки цього класифікатора:

- навчання нейронної мережі є повільним через велику кількість параметрів для оптимізації;

- збіжність до локального мінімуму;
- схильність до перенавчання (overfitting);
- є чорним ящиком.

Логістична регресія

Логістична регресія — статистичний регресійний метод моделювання залежності між векторною змінною та скаляром (вихідним значенням). Цей метод є узагальненням методу лінійної регресії з використанням softmax функції і застосовується у випадку, коли залежна змінна може набувати лише скінченну множину значень.

Параметри оцінюються на основі валідаційної вибірки, як правило, за допомогою методу максимальної правдоподібності, згідно з яким вибираються параметри, що максимізують значення функції правдоподібності на вибірці [1].

Основна відмінність та перевага такого підходу від інших моделей і алгоритмів є оцінка результату, яку можна було б розглядати як значення ймовірності для певного класу [5].

Недоліком цього підходу є нездатність побудови гіперплощини складного вигляду і, як наслідок, не дуже висока точність розпізнавання.

Метод опорних векторів

Метод опорних векторів — категорія універсальних мереж прямого поширення — запропонований в 1963 р. Вапніком [17]. Метод SVM набув поширення в останнє десятиліття для класифікації, регресії та ідентифікації новизни. Важливою властивістю SVM є те, що визначення параметрів моделі відповідає задачі випуклої оптимізації (convex optimization), і тому будь-який локальний розв'язок також є глобальним.

Такий підхід до класифікації передбачає розгляд поняття поділу (margin), яке визначається як мінімальна відстань від гіперплощини до зразків з вибірки. Подільна гіперплощина будується в такий спосіб, щоб максимізувати значення поділу. Розміщення гіперплощини визначається підмножиною точок даних, відомих як опорні вектори. Для того щоб зробити класифікатор більш потужним, у 1992 р. був запропонований спосіб створення нелінійного класифікатора, в основу якого покладено перехід від скалярних добутків до довільних ядер [4]. Це дозволило будувати нелінійні подільні гіперплощини, але не дало змоги працювати безпосередньо з початковими даними. Найбільш поширені ядра: поліноміальне, радіальна базисна функція, радіальна базисна функція Гауса та сигмоїд.

Перевагами методу опорних векторів є збіжність до глобального мінімуму та ефективність за обмеженої кількості навчальних даних [5].

Недоліки методу:

- модель SVM відносить новий приклад до тієї чи іншої категорії, що робить її не ймовірнісною, а лише бінарним класифікатором [19];
- використання ядер забезпечує роботу з лінійно неподільними множинами даних, але не дозволяє отримувати достовірні узагальнення для векторів, що віддалені від навчальних даних [4];

- навчання SVM є надто повільним і не може використовуватися для аналізу великих обсягів даних [20, 21].

Random forest

Random forest — один з поширених методів машинного навчання, що полягає у використанні ансамблю дерев рішень [21, 22]. Застосовується для задач класифікації, регресії і кластеризації. Дерево рішень будується на основі навчальної вибірки з використанням поняття інформаційної ентропії. На кожному вузлі дерева вибирається один атрибут з даних, який найбільш ефективно ділить навчальну множину на підмножини, що максимально розрізняються. Головним критерієм для вибору є нормований приріст інформації. Атрибут з найбільшим нормалізованим приростом інформації вибирається для прийняття рішення щодо поділу даних у вузлі дерева.

Більшість алгоритмів побудови дерев рішень є «жадібними». На кожному вузлі вибирається найкращий атрибут для поділу даних у поточний момент часу, що не дозволяє алгоритму з часом повернутись назад і змінити вже вибраний атрибут. Такий підхід не може гарантувати отримання оптимального розбиття.

Для навчання кожного дерева рішення, з яких складається класифікатор, навчальна вибірка ділиться на частини, що можуть містити спільні елементи. Ця технологія має назву «статистичний бутстреп» (bootstrapping) і дозволяє уникнути великої кореляції між деревами рішення, яка виникає під час навчання їх з використанням одних і тих самих даних [5].

З огляду на велику чутливість класифікатора до навчальних даних невелика зміна в навчальній вибірці призводить до зовсім інших поділів на підмножини. Для зменшення чутливості та подолання схильності до перенавчання використовують такі технології, як прорідження (pruning) уже отриманого дерева.

Перевагами цього алгоритму є можливість оцінювання значущості ознак у моделі, висока швидкість навчання та легка інтерпретація отриманої моделі.

Недоліки алгоритму: схильність до перенавчання, особливо з багатьма рівнями шумів, великий розмір отримуваних моделей.

ПОНЯТТЯ ГЛИБИННОГО НАВЧАННЯ ТА АНАЛІЗ ПІДХОДІВ ДО ЙОГО РЕАЛІЗАЦІЇ

Розглянуті вище моделі мають неглибоку (shallow) архітектуру і значною мірою залежать від якості виокремлених вручну ознак. Виокремити ознаки, які були б стійкими і в разі повертання зображення, зміщення, зашумлення чи зміни кута спостереження, є дуже складним завданням.

Останнім часом дедалі більшого поширення набуває клас методів глибокого навчання [4], яке включає етап навчання ознак або подань.

Навчання ознак (feature learning) або навчання подань (representation learning) — набір методів для перетворення даних початкового вигляду в подання, що може бути використано для навчання класифікатора [4].

Методи глибинного навчання є методами навчання подань з багатьма їх рівнями, отриманими через об'єднання простих нелінійних модулів, кожен з яких переводить подання з одного рівня (починаючи з початкових даних) на вищий, більш абстрактний рівень. Об'єднання достатньої кількості таких трансформувальних забезпечує навчання функції будь-якої складності. Ключовим аспектом глибинного навчання є те, що шари особливостей не розробляються експертами в тій чи іншій галузі, а автоматично виокремлюються із вхідних даних.

Метод групового врахування аргументів (Group Method of Data Handling (GMDH)), запропонований українським науковцем Олексієм Івахненком, є одним з перших прикладів глибинного навчання для багатосарових нейронних мереж (у його працях розглядається мережа з 8 шарами), який використовує поліноми Колмогорова–Габора [23, 24]. Інші архітектури глибинного навчання, зокрема побудовані зі штучних нейронних мереж, беруть свій початок з неокогнітрону, запропонованого в 1980 р. Куніхіко Фукусімою [25]. Неокогнітрон — це перша двовимірна згортовка нейронна мережа (Convolutional Neural Network (CNN)). Проте в цій мережі ваги налаштовувались не за допомогою методу зворотного поширення помилок, а локально за допомогою підходу «переможець отримує все» (winner-take-all). Тому цей тип мереж належить до навчання без учителя [12]. У 1989 р. метод зворотного поширення помилки успішно застосував Ян Лекун до глибинної двовимірної згорткової нейронної мережі в задачі розпізнавання рукописних цифр [26]. Після цього багато уваги приділялось розвитку глибинних мереж переконань (Deep Belief Networks (DBN)), які являють собою стек обмежених машин Больцмана (Restricted Boltzmann Machines (RBMs)), які, у свою чергу, є машиною Больцмана з тим обмеженням, що нейрони мусять формувати двочастковий граф: з'єднань між вузлами в межах групи немає [27]. Почали розвиватись також рекурентні глибинні мережі для оброблення послідовностей даних, що залежать від часу [28]. Проте в 1991 р. було помічено, що дуже глибокі нейронні мережі, особливо рекурентні нейронні мережі, складно навчити методом зворотного поширення помилки [29]. Ця проблема називається проблемою зникнення або вибуху градієнта (Vanishing or Exploding Gradients): з використанням стандартних функцій активації сукупна помилка сигналу зворотного поширення або дуже швидко прямує до нуля, або дуже швидко експоненційно зростає. Для подолання цієї проблеми в 1997 р. було запропоновано рекурентні мережі з моделлю довгої короткочасної пам'яті (Long Short-Term Memory (LSTM)) [30].

Після разового успішного використання Яном Лекуном згорткових нейронних мереж вони не здобули популярності [31]. І лише в 2012 р. А. Крижевський та інші науковці відновили інтерес до згорткових нейронних мереж після того, як показали вражаюче високу точність класифікації зображень на змаганні ImageNet Large Scale Visual Recognition Challenge (ILSVRC). На цьому змаганні нейронні мережі застосовувалися до набору даних, що налічував понад мільйон зображень з інтернету, які містили більше ніж 1000 різних класів. Цей успіх розпочав революцію в напрямку комп'ютерного зору із застосуванням глибинних мереж у різноманітних напрямках.

Сьогодні методи глибинного навчання показують значні успіхи у вирішенні проблем, які неодноразово намагалися розв'язати протягом останніх десятиліть за допомогою штучного інтелекту: установлення рекордів у розпізнаванні зображень [32, 33], виявленні транспортних засобів [34], розпізнаванні людської мови [35], реконструкції структури мозку [36], перекладі з однієї мови на іншу [37].

Важливою вимогою для застосування методів глибинного навчання є наявність навчальних вибірок великого обсягу, оскільки недостатній обсяг навчальних даних спричиняє проблему «перенавчання» (overfitting), коли модель не узагальнює отриману під час навчання інформацію, а просто її запам'ятовує. У такому випадку на навчальних даних модель демонструє хороші результати, але не показує такої точності на невідомих даних.

АКТУАЛЬНІСТЬ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ГЕОПРОСТОРОВОГО АНАЛІЗУ ТА СУПУТНИКОВОГО МОНІТОРИНГУ

Для геопросторового аналізу джерелами інформації є супутникові спостереження, дані безпілотних літальних апаратів та інші типи геопросторової інформації, наприклад, зібраної шляхом краудсорсингу (crowdsourcing). Нині людство вступило в нову інформаційну еру, яка характеризується наявністю у вільному доступі великих обсягів різноманітних безкоштовних супутникових даних. Протягом 2013–2016 рр. запущено декілька оптичних та радарних супутників високого розрізнення (10–30 м), що надають безоплатні дані. Sentinel-1A/B та Sentinel-2A були запущені Європейським космічним агентством (ESA) [38, 3] та Landsat-8 у межах проекту Landsat, що є спільною ініціативою між Геологічною службою США (USGS) та Національним управлінням з авіації та дослідження космічного простору (NASA) [40]. Ці набори даних поновлюються регулярно та надаються користувачам у вільному доступі на оперативній основі. Це відкриває безпрецедентні можливості для геопросторового аналізу за переваги наявності часових рядів даних та можливості злиття інформації з різних джерел [41, 42]. Серед усіх продуктів, які можна отримати із супутникових даних, найважливішими є карти класифікації [43], що надзвичайно актуальні, особливо для країн, що розвиваються, оскільки офіційна статистика часто є неповною або недостовірною [44]. Карти класифікації використовуються для багатьох прикладних завдань, таких як моделювання та оцінювання змін довкілля, клімату, прогнозування врожайності і т.ін. Окрім використання карт класифікацій в різних прикладних завданнях, виникає потреба в регулярному оновленні карт класифікацій для ідентифікації вирубок лісів, аналізу пошкоджень від затоплень, посух або будь-яких інших надзвичайних ситуацій. Для більшості таких завдань не вистачає використання лише одного джерела даних, що змушує розробляти методи злиття даних з джерел з різними характеристиками, різним просторовим та часовим розрізненням [43, 45].

Різновиди класифікаторів, призначених для аналізу багатоспектральних даних, наведено в табл. 2.

Таблиця 2. Різновиди підходів до класифікації багатоспектральних даних

Критерій	Класифікатори	Короткий опис
Урахування просторової інформації	Спектральні	Розглядаються гіперспектральні зображення без урахування просторових залежностей
	Просторові	Вхідні дані класифікуються з використанням інформації про «сусідні» (просторово прилеглі) піксели
	Просторово-спектральні	Для класифікації гіперспектральних даних враховується сукупність спектральної та просторової інформації
Тип використуваної інформації	Попіксельний класифікатор	На вхід до класифікатора подаються вектори значень для кожного пікселя
	Об'єктний класифікатор	Кожному пікселю на зображенні ставиться у відповідність мітка таким чином, щоб піксели з однаковими мітками мали певні спільні візуальні характеристики. Класифікація проводиться на основі виділених об'єктів, а не лише одного пікселя

Специфіка класифікації геопросторових даних полягає в тому, що переважна більшість методів орієнтовані на попіксельну класифікацію — спостереження для кожного конкретного пікселя формують вхідний вектор, довжина якого дорівнює кількості спектральних каналів на знімку і на виході моделі кожному пікселю призначається значення певного класу. Для розв'язання задачі класифікації багатоспектральних супутникових знімків застосовується розвинений математичний апарат інтелектуальних методів.

АНАЛІЗ МЕТОДІВ КЛАСИФІКАЦІЇ ЗЕМНОГО ПОКРИВУ ТА ОРНИХ ЗЕМЕЛЬ ЗА СУПУТНИКОВИМИ ДАНИМИ

Донедавна найефективнішим методом для задач класифікації земного покриття за одним знімком вважався SVM. Це підтверджується результатами комплексного дослідження сучасних (state-of-the-art) попіксельних методів навчання з учителем для побудови карт земного покриття, отриманими Р. Катамі та співавторами [46]. В експерименті класифікація виконувалась за допомогою лише одного знімка. Майже такий самий результат показав нейромережвий підхід. Проте метод SVM потребує значно більше обчислювальних ресурсів. Його складність лежить у діапазоні між $O(n^2)$ та $O(n^3)$ залежно від типу ядра та реалізації [47, 48], у той же час виявляється складність багатослового перцептрона $O(n)$ [49, 50]. Як наслідок використовувати метод SVM для аналізу великих обсягів даних неможливо, а традиційні моделі багатослового перцептрона, які застосовувалися для розпізнавання зображень, через повну зв'язність між вузлами потерпають від «прокляття розмірності», тому не масштабуються на зображення вищого розрізнення, а величезна кількість параметрів швидко веде до перенавчання. У таких мережових моделях не береться до уваги просторова структура даних; однаково розглядаються вхідні піксели, розміщені далеко та близько один від одного.

Методи класифікації земного покриття на основі «великих» обсягів супутникових даних

Поява великих обсягів даних високого розрізнення не тільки відкриває нові можливості, але й породжує нові проблеми, зумовлені їх обробленням. Річ у тім, що методи геопросторового аналізу, класифікації та формування навчальної вибірки, що добре зарекомендували себе у разі оброблення одного знімка або невеликої території, виявляються непридатними для великих територій. Основні проблеми пов'язані з великими обсягами оброблюваних даних, малою площею покриття одним знімком, зашумленістю даних (хмарністю), необхідністю одночасної класифікації різночасових знімків і т.ін. Саме тому в останні роки найпоширенішими й ефективними методами класифікації типів земного покриття на основі різнорідних часових рядів супутникових даних стали ансамблеві методи [51–54] та методи глибинного навчання [55–60]. На відміну від загальноприйнятої схеми такі підходи працюють одразу з початковими даними та використовують усі наявні вхідні дані, а не виокремлюють лише важливі ознаки, як це було раніше [61, 62]. Методи глибинного навчання підтвердили свою ефективність для оброблення як оптичних (гіперспектральних та мультиспектральних), так і радарних зображень, побудови різних типів земної поверхні: ідентифікації доріг, будинків [55, 63–65]. Найбільш поширені моделі в глибинному навчанні для аналізу геопросторових даних — це згорткові нейронні мережі, глибинні автокодувальники (Deep Auto-encoders (DAE)), глибинні мережі переконань та рекурентні мережі з моделлю тривалої короткочасної пам'яті [55, 65–68].

Натепер багато праць демонструють перевагу методів глибинного навчання над загальноприйнятими методами неглибокої архітектури в задачах геопросторового аналізу. Так, у праці [44] для оцінювання індексу врожайності кукурудзи на території США використано два різні регресійні методи: Convolutional Architecture for Fast Feature Embedding (CAFFE) та Support Vector Regression (SVR). У результаті метод глибинного навчання перевершив SVR. Коефіцієнт кореляції для нього становив 0,81, середньоквадратичне відхилення — 6,298, тоді як коефіцієнт кореляції та середньоквадратичне відхилення для SVR дорівнювали 0,644 та 8,204 відповідно.

Автори [34] порівняли метод глибинного навчання на основі згорткової нейронної мережі та комбінації ансамблю SVM з різними методами для виокремлення ознак для задачі ідентифікації транспортних засобів за супутниковими знінками. Запропонований метод глибинного навчання істотно перевершив за точністю всі комбінації методу SVM.

У праці [19] порівняно метод глибинного навчання на основі розрідженого автокодувальника (sparse auto-encoders) та SVM для класифікації супутникових знімків. Отримані результати свідчать про те, що SVM варто використовувати у випадку нестачі навчальних даних. У цьому разі він виявився продуктивнішим ніж розріджений автокодувальник. Проте для великої кількості даних для навчання доцільно використовувати методи глибинного навчання. Також показано більшу чутливість SVM до шуму в навчальних даних порівняно з автокодувальником.

Для розпізнавання цілей, у тому числі і військових, з використанням даних з радара із синтезованою апертурою (РСА) існує три загальні проблеми: наявність шуму, зміщення цілі та поворот цілі. Для цього використову-

валась згорткова нейронна мережа та SVM [31]. Точності, отримані за допомогою SVM на звичайних даних, — 75,68%, на зашумлених даних — 70,58%. Точності згорткової нейронної мережі на звичайних даних — 93,16%, на зашумлених даних — 91,89%, що демонструє суттєву перевагу методів глибинного навчання над SVM.

Розпізнати різні типи поверхні Землі за допомогою оптичних супутникових знімків є надзвичайно важливим завданням. У праці [69] розв'язувалась задача ідентифікації полів для гольфу за допомогою оптичних супутникових знімків Landsat-8 з просторовим розрізненням 30 м. У ході такого експерименту перевірялась залежність якості класифікації за допомогою згорткової нейронної мережі та SVM від кількості даних у навчальній вибірці. Як метрику для оцінювання методів класифікації використовувалась не загальна точність класифікації, а F-value у зв'язку з наявністю величезної кількості від'ємних даних як у навчальній, так і в тестовій вибірці. Для всіх розмірів навчальної вибірки згорткова нейронна мережа точніше класифікувала поля для гольфу ніж SVM (рис. 2). Притому час навчання SVM значно перевищує час навчання згорткової нейронної мережі, і зі збільшенням розміру навчальної вибірки їх різниця істотно зростає (рис. 3).

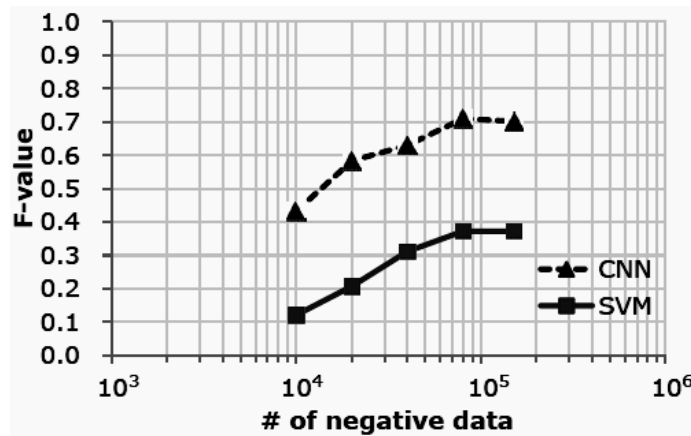


Рис. 2. Результат порівняння точності CNN і SVM [69]

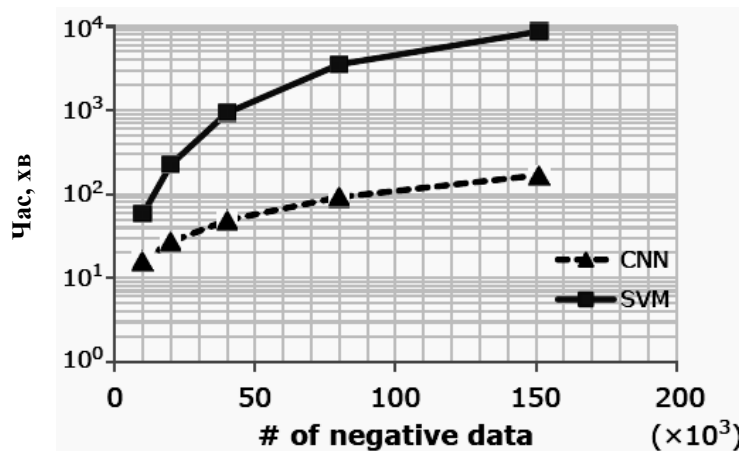


Рис. 3. Результат порівняння часу навчання CNN і SVM

Традиційний процес класифікації супутникових даних

Традиційно класифікація геопросторових даних виконувалась за загально-прийнятою схемою (рис. 4) [61], за якою навчання класифікатора охоплювало широкий спектр складних етапів: вибір галузі інтересу, поділ її на частини, виокремлення ознак, підготовка навчальної вибірки, класифікація та сегментація.

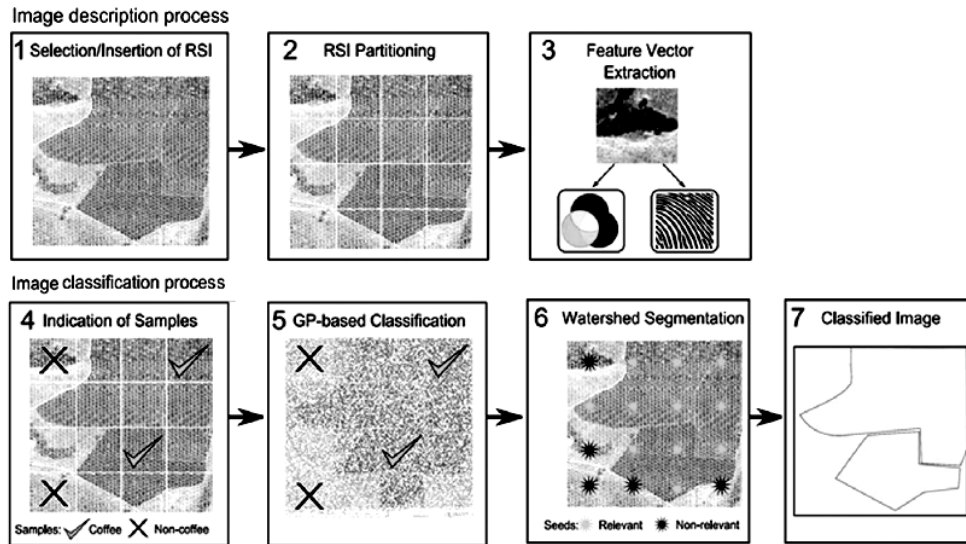


Рис. 4. Схема загальноприйнятого процесу класифікації [61]

Найбільш складною частиною був етап виокремлення ознак, який, як і в інших предметних галузях, потребував ґрунтовних експертних знань специфіки геопросторових даних та властивостей досліджуваних об'єктів. Протягом багатьох років для ручного виокремлення ознак використовувались такі методи, як Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF) та Histogram of Oriented Gradients (HOG) [70]. Проте методи такого типу виокремлюють ознаки лише на низькому рівні (локальні характеристики зображення). Ідентифікація ознак на вищих рівнях, таких як виявлення контурів (edge detection), їх перетинів та частин об'єктів є надзвичайно складним завданням [34, 70].

Класифікація на основі радарних даних

За наявності хмар і тіней від них в оптичних знімках останнім часом дедалі більше використовуються радарні дані для класифікації типів земної поверхні. Радарні дані не залежать від погоди і можуть надавати достовірну інформацію не лише вдень, а й уночі. Проте найбільшою проблемою у використанні радарних даних є великий рівень шуму в них [65]. Зазвичай класифікація радарних даних виконується попіксельно з використанням таких ознак, як інтенсивність, амплітуда та фаза, хоча цього недостатньо для отримання класифікації без наявності великої кількості шумів. Необхідно розглядати текстуру радарних знімків. Багато різних фільтрів для аналізу текстури використовують середнє значення, дисперсію, контраст і т.ін. Після застосування цих фільтрів попіксельні методи машинного навчання,

такі як SVM, широко застосовуються для класифікації радарних даних. Але використання такого підходу є складним і займає багато часу, оскільки потребує ґрунтовних знань для підбору параметрів та придатних фільтрів, що можуть суттєво відрізнитись залежно від території, властивостей об'єктів, розміщених на ній, особливо з урахуванням необхідності оброблення багатьох супутникових знімків. Тому доцільно використовувати методи глибинного навчання для автоматичного виокремлення ознак та врахування текстур. Можливості методів глибинного навчання для радарних даних, отриманих із супутника TerraSAR-X, та порівняння їх із загальноприйнятим підходом проаналізовано у праці [65]. Для цього багатосаровий автокодувальник (SAE), класифікатор з розрідженим поданням (SRC) та згортковий автокодувальник (DCAE) порівнювалися з методом SVM на основі RBF ядра. Розглянуті методи глибинного навчання перевершили SVM. Проте найкращим виявився DCAE, що за рахунок свого підвибіркового шару дав змогу відчутно знизити рівень шуму на вихідній карті класифікації (рис. 5).

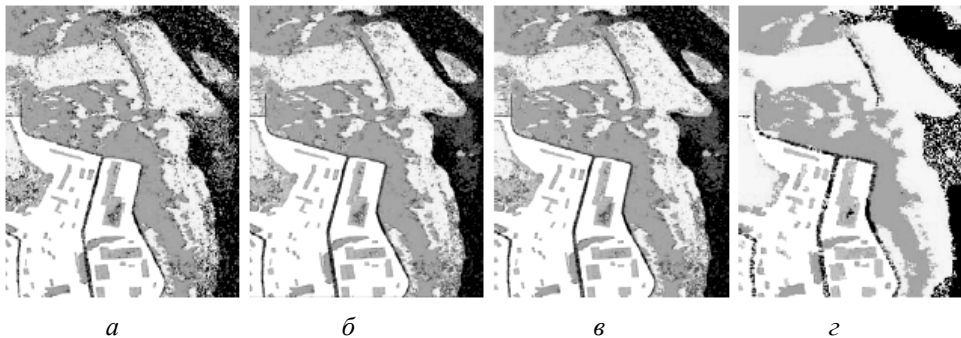


Рис. 5. Порівняння карт класифікації радарних даних, отриманих за різними методами: а — SVM; б — SRC; в — SAE; г — DCAE [65]

Із розглянутих вище праць випливає, що методи глибинного навчання — це найефективніший інструмент оброблення великих обсягів багатоспектральних супутникових даних серед методів машинного навчання. Результати експериментів показують, що методи глибинного навчання в усіх випадках випереджають інші загальноприйняті досліджувані методи машинного навчання неглибокої архітектури, що протягом десятиліть використовувались як найкращі методи для класифікації супутникових даних.

ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ В ЗАДАЧАХ КЛАСИФІКАЦІЇ СУПУТНИКОВИХ ДАНИХ

Найбільш прийнятною глибинною архітектурою для оброблення геопросторових даних і, як наслідок, для розв'язання поставленої задачі класифікації посівів є згорткові нейронні мережі, оскільки вони призначені для оброблення даних, які подаються у вигляді багатовимірних масивів. Наприклад, стандартне кольорове зображення складається з трьох двовимірних масивів, що містять інтенсивності пікселів у кожному із трьох каналів (R — червоний, G — зелений та B — голубий). Багатоспектральні дані подаються саме у вигляді багатовимірних масивів. Одновимірні масиви використовуються

для подання сигналів та послідовностей, зокрема мови. Двовимірні масиви застосовують для зображень, аудіоспектрограм, геопросторових даних. Тривимірні подання використовують для відео- або об'ємних зображень, у тому числі багатоспектральних супутникових знімків або їх рядів. Варто зазначити, що методи попиксельної класифікації геопросторових даних часто потерпають від наявності шуму на вихідній карті класифікації, що називається шумом типу «сіль та перець» (salt-and-pepper). Це притаманно не лише радарним даним, а й оптичним знімкам високого розрізнення. Саме тому для класифікації важливо розглядати не лише спектральні характеристики окремих пікселів, а і просторові, що і робить згорткова нейронна мережа.

Архітектура згорткової нейронної мережі

Можна виокремити чотири ключові властивості, які забезпечують перевагу використання згорткових нейронних мереж для оброблення геопросторових даних: локальні з'єднання, спільні ваги (shared weights), підвибірка (pooling) та використання багатьох шарів.

Типова згорткова мережа містить декілька рівнів оброблення інформації (рис. 6) [4]. Перші кілька рівнів складаються з двох типів шарів: згорткових, що чергуються з підвибірковими шарами. Згорткові та підвибіркові шари безпосередньо пов'язані з класичними уявленнями про прості та складні клітини в нейробіології, а архітектура в цілому нагадує організацією зорової кори тварин [4].

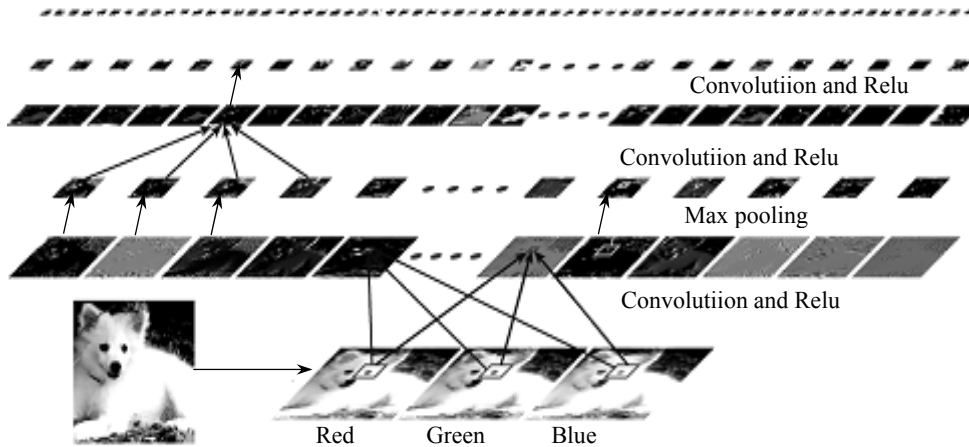


Рис. 6. Архітектура типової згорткової нейронної мережі [4]

Згортковий шар є основним блоком мережі. Параметри шару складаються з набору фільтрів або ядер, які мають невелике рецептивне поле по ширині та довжині, але простягаються на всю глибину вхідного багатовимірного масиву. Однією з головних переваг згорткових мереж є використання спільної ваги у згорткових шарах: для кожного пікселя певного шару використовується один і той же фільтр (weights bank). Основна ідея полягає в тому, що ідентифікація об'єкта в одній частині вхідних даних дасть змогу знайти такий самий об'єкт в іншій частині даних. Це допомагає зменшити ємність необхідної пам'яті для збереження архітектури мережі та

підвищити продуктивність мережі за рахунок інваріантності до зсуву вхідних даних. Під час прямого проходження кожен фільтр здійснює згортку по ширині та висоті вхідного масиву даних, обчислюючи скалярний добуток вхідних даних та фільтра, формуючи двовимірну карту активації цього фільтра. Об'єднання всіх активаційних карт, отриманих усіма фільтрами вздовж виміру глибини, формує вихідний масив згорткового шару. За вхідних даних великої розмірності, таких як супутникове зображення, недоцільно з'єднувати нейрони з усіма нейронами попереднього шару, оскільки у такій архітектурі мережі не враховуються просторові властивості даних. У згорткових мережах використовується локальна просторова кореляція шляхом локального з'єднання між нейронами сусідніх шарів: кожен нейрон з'єднується лише з невеликою ділянкою вхідного масиву [4]. Розмір цієї ділянки називається рецептивним полем нейрона і задається параметром. Після кожного згорткового шару застосовується нелінійна функція активації, як і для звичайного багатозарового перцептрона.

Підвибірковий (pooling) шар є різновидом нелінійного зниження потужності множини дискретних значень шляхом об'єднання семантично подібних ознак в одну. Існує декілька нелінійних функцій для реалізації підвибірки, серед яких найпоширенішою є підвибірка максимізації (maxpooling). Вона поділяє вхідне зображення на множину квадратів, які не перетинаються, і для кожного з них залишає максимальне значення. Ідея полягає в тому, що точне положення знайденої ознаки не так важливе, як її грубе положення відносно інших ознак. Підвибірковий шар забезпечує інваріантність відносно зсуву та спотворення об'єкта на вхідному зображенні.

Після декількох послідовних чергувань згорткового та підвибіркового шарів використовують один або декілька повнозв'язних шарів (fully-connected) нейронів. Нейрони у повнозв'язному шарі з'єднані з усіма нейронами попереднього шару, як у звичайних нейронних мережах прямого поширення. Звичайний метод зворотного поширення похибки дозволяє навчати згорткову нейронну мережу, підбираючи ваги у фільтрах аналогічно до того, як навчаються звичайні багатозарові мережі.

Основним недоліком традиційних нейронних мереж вважається їх збіжність до локального мінімуму. Цей недолік притаманний і глибинним нейронним мережам. Проте на практиці збіжність до локального мінімуму рідко є проблемою для глибинних мереж. Незалежно від початкових умов система майже завжди досягає розв'язку однакового рівня якості [4]. Нещодавні теоретичні та емпіричні результати переконливо свідчать про те, що локальні мінімуми загалом не є проблемою [72, 73]. Річ у тім, що поверхня функції втрат містить багато сідлових точок, де градієнт дорівнює нулю і в переважній їх більшості значення функції втрат дуже схожі. Таким чином, не має великого значення, у якій з цих точок алгоритм навчання мережі зупиниться.

Особливості та перспективи методів глибинного навчання для класифікації супутникових даних

У розглянутих вище працях методи глибинного навчання продемонстрували відмінні результати оброблення геопросторової інформації. Відзначимо, що

для забезпечення надійного поділу класів варто використовувати всі інформативні канали в багатоспектральних зображеннях. Проте згорткові нейронні мережі розглядались лише для класифікації звичайних двовимірних зображень, що зберігаються у RGB форматі або у відтінках сірого (grayscale), а технологія застосування їх до гіперспектральних зображень дуже рідко подається в літературі.

Для класифікації гіперспектральних зображень застосувати згорткову нейронну мережу напрямку не вдається [74]. Проте можна кожен піксел супутникового знімка розглядати як двовимірне зображення, висота якого дорівнює 1. Тоді на вході в мережу матимемо вектор, розмірність якого дорівнює кількості каналів у супутниковому зображенні. Таку архітектуру можна вважати одновимірною згортковою нейронною мережею. У праці [74] її ефективність порівнюється з методом RBF-SVM для класифікації гіперспектрального зображення. Показано, що цей підхід не лише перевершив RBF-SVM у термінах загальної точності класифікації, а й забезпечив вищу точність класифікації майже всіх окремих класів на трьох різних наборах даних.

Варто відзначити, що в більшості праць за тематикою аналізу геопросторових даних метод глибинного навчання використовується лише для класифікації одного знімка за конкретну дату. Таку постановку задачі не можна вважати проблемою оброблення великих обсягів даних (Big Data), а отже, багат шарова нейронна мережа не є оптимальною моделлю для їх класифікації. Для забезпечення достовірної класифікації сільськогосподарських культур чи типів земної поверхні необхідно використовувати часовий ряд знімків або навіть об'єднувати (зливати) дані з різних супутників, які доповнюють один одного важливою інформацією про спостережувану поверхню. Типовим прикладом такого злиття інформації є спільне використання оптичних (чутливих до сонячного освітлення, наявності хмар) та радарних даних (нечутливих до сонячного освітлення, майже зовсім нечутливих до хмарності) [43]. За спроби застосування загальноприйнятих методів глибинного навчання до задачі класифікації часового ряду супутникових знімків виникають такі недоліки: великий розмір знімків (щонайменше 10000 на 10000 пікселів), не однаковий розмір знімків та не повне накладання їх між собою, особливо з різних супутників, зашумленість даних (хмари та тіні від них), відносно невеликий розмір навчальних вибірок, що можна зібрати. Через експоненційне збільшення розмірів супутникових знімків різної природи актуальним завданням є розвиток моделей глибинного навчання, спроможних працювати з часовими рядами різнорідних супутникових даних.

ЛІТЕРАТУРА

1. *Bishop C.M.* Pattern Recognition and Machine Learning / C.M. Bishop. — NY: Springer. — 2006.
2. *Zgurovsky M.Z.* System Analysis: Theory and Applications / M.Z. Zgurovsky, N.D. Pankratova // Springer. — 2007. — 448 p.
3. *Novikov A.* The synthesis of information protection systems with optimal properties / A. Novikov, A. Rodionov // Complexity and Security. — Vol. 37. — 2008. — 307 p.
4. *LeCun Y.* Deep learning / Y. LeCun, B. Yoshua, H. Geoffrey // Nature. — Vol. 521, N 7553. — 2015. — P. 436–444.

5. *Ghamisi P.* Advanced Spectral Classifiers for Hyperspectral Images: A review / P. Ghamisi, J. Plaza, Y. Chen et al // IEEE Geoscience and Remote Sensing Magazine. — Vol. 5, N 1. — 2017. — P. 8–32.
6. *Fukunaga K.* Introduction to statistical pattern recognition / K. Fukunaga // Academic press, 2013. — 591 p.
7. *Maulik U.* Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques / U. Maulik, D. Chakraborty // IEEE Geoscience and Remote Sensing Magazine. — Vol. 5, N 1. — 2017. — P. 33–52.
8. *Gislason P.O.* Random forests for land cover classification / P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson // Pattern Recognition Letters. — Vol. 27, N 4. — 2006. — P. 294–300.
9. *McCulloch W.S.* A logical calculus of the ideas immanent in nervous activity / W.S. McCulloch, W. Pitts // The bulletin of mathematical biophysics. — Vol. 5, N 4. — 1943. — P. 115–133.
10. *Rumelhart D.E.* Learning representations by back-propagating errors / D.E. Rumelhart, G.E. Hinton, R.J. Williams // Cognitive modeling. — Vol. 5, N 3. — 1988. — P. 213–220.
11. *Amari S.I.* Statistical theory of learning curves under entropic loss criterion / S.I. Amari, N. Murata // Neural Computation. — Vol. 5, N 1. — 1993. — P. 140–153.
12. *Schmidhuber J.* Deep learning in neural networks: An overview / J. Schmidhuber // Neural networks. — Vol. 61. — 2015. — P. 85–117.
13. *Rumelhart D.E.* Learning internal representations by error propagation / D.E. Rumelhart, G.E. Hinton, R.J. Williams // Parallel distributed processing. MIT Press. — 1986. — P. 318–362.
14. *Duchi J.* Adaptive subgradient methods for online learning and stochastic optimization / J. Duchi, E. Hazan, Y. Singer // Journal of Machine Learning Research. — 2011. — P. 2121–2159.
15. *Patel V.* Kalman-Based Stochastic Gradient Method with Stop Condition and Insensitivity to Conditioning / V. Patel // SIAM Journal on Optimization. — Vol. 26, N 4. — 2016. — P. 2620–2648.
16. *Kingma D.* Adam: A method for stochastic optimization / D. Kingma, J. Ba // arXiv preprint arXiv. — Vol. 1412, N 6980. — 2014. — P. 1–15.
17. *Haykin S.* Neural networks and learning machines / S. Haykin // Upper Saddle River. — NJ, USA: Pearson, 2009. — Vol. 3. — 938 p.
18. *Hornik K.* Multilayer feedforward networks are universal approximators / K. Hornik, M. Stinchcombe, H. White // Neural networks. — Vol. 2, N 5. — 1989. — P. 359–366.
19. *Liu P.* SVM or deep learning? A comparative study on remote sensing image classification / P. Liu, K.K.R. Choo, L. Wang, F. Huang // Soft Computing. — Vol. 21, N 23. — 2017. — P. 7053–7065.
20. *Huang F.J.* Large-scale learning with SVM and convolutional nets for generic object categorization / F.J. Huang, Y. LeCun // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — 2006. — P. 284–291. DOI: 10.1109/CVPR.2006.164.
21. *Pirotti F.* Benchmark of machine learning methods for classification of a Sentinel-2 image / F. Pirotti, F. Sunar, M. Piragnolo // International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. — Vol. 41. — 2016. — P. 335–340.
22. *Breiman L.* Random forests / L. Breiman // Machine learning. — Vol. 45, N 1. — 2001. — P. 5–32.
23. *Ivakhnenko A.G.* The group method of data handling—a rival of the method of stochastic approximation / A.G. Ivakhnenko // Soviet Automatic Control. — Vol. 13, N 3. — 1968. — P. 43–55.
24. *Ivakhnenko A.G.* Polynomial theory of complex systems / A.G. Ivakhnenko // IEEE transactions on Systems, Man, and Cybernetics. — Vol. 1, N 4. — 1971. — P. 364–378.

25. *Fukushima K.* Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position / K. Fukushima // *Biol. Cybern.* — Vol. 36. — 1980. — P. 193–202. Doi:10.1007/bf00344251.
26. *LeCun Y.* Backpropagation applied to handwritten zip code recognition / Y. LeCun, B. Boser, J.S. Denker et al. // *Neural computation.* — Vol. 1, N 4. — 1989. — P. 541–551.
27. *Hinton G.E.* A fast learning algorithm for deep belief nets / G.E. Hinton, S. Osindero, Y.W. Teh // *Neural computation.* — Vol. 18, N 7. — 2006. — 1527–1554 p.
28. *Connor J.T.* Recurrent neural networks and robust time series prediction / J.T. Connor, R.D. Martin, L.E. Atlas // *IEEE transactions on neural networks.* — Vol. 5, N 2. — 1994. — P. 240–254.
29. *Hochreiter S.* Untersuchungen zu dynamischen neuronalen Netzen / S. Hochreiter // *Diss. diploma thesis, institut für informatik, lehrstuhl prof. brauer, technische universität münchen.* — 1991.
30. *Hochreiter S.* Long short-term memory / S. Hochreiter, J. Schmidhuber // *Neural computation.* — Vol. 9, N 8. — 1997. — P. 1735–1780.
31. *Ding J.* Convolutional neural network with data augmentation for SAR target recognition / J. Ding, B. Chen, H. Liu, M. Huang // *IEEE Geoscience and Remote Sensing Letters.* — Vol. 13, N 3. — 2016. — P. 364–368.
32. *Krizhevsky A.* Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // *In Advances in neural information processing systems.* — 2012. — P. 1097–1105.
33. *Ngiam J.* Multimodal deep learning / J. Ngiam, A. Khosla, M. Kim et al. // *In Proceedings of the 28th international conference on machine learning (ICML-11).* — 2011. — P. 689–696.
34. *Chen X.* Vehicle detection in satellite images by hybrid deep convolutional neural networks / X. Chen, S. Xiang, C.L. Liu, C.H. Pan // *IEEE Geoscience and remote sensing letters.* — Vol. 11, N 10. — 2014. — P. 1797–1801.
35. *Hinton G.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups / G. Hinton, L. Deng, D. Yu et al. // *IEEE Signal Processing Magazine.* — Vol. 29, N 6. — 2012. — P. 82–97.
36. *Helmstaedter M.* Connectomic reconstruction of the inner plexiform layer in the mouse retina / M. Helmstaedter, K.L. Briggman, S.C. Turaga et al. // *Nature.* — Vol. 500, N 7461. — 2013. — P. 168–174.
37. *Sutskever I.* Sequence to sequence learning with neural networks / I. Sutskever, O. Vinyals, Q.V. Le // *In Advances in neural information processing systems.* — 2014. — P. 3104–3112.
38. *Drusch M.* Sentinel-2: ESA's optical high-resolution mission for GMES operational services / M. Drusch, U. Del Bello, S. Carlier et al. // *Remote Sensing of Environment.* — Vol. 120. — 2012. — P. 25–36.
39. *Torres R.* GMES Sentinel-1 mission / R. Torres, P. Snoeij, D. Geudtner et al. // *Remote Sensing of Environment.* — Vol. 120. — 2012. — P. 9–24.
40. *Roy D.P.* Landsat-8: Science and product vision for terrestrial global change research / D.P. Roy, M.A. Wulder, T.R. Loveland et al. // *Remote Sensing of Environment.* — Vol. 145. — 2014. — P. 154–172.
41. *Zhang J.* Multi-source remote sensing data fusion: status and trends / J. Zhang // *International Journal of Image and Data Fusion.* — Vol. 1, N 1. — 2010. — P. 5–24.
42. *Dalla M.M.* Challenges and opportunities of multimodality and data fusion in remote sensing / M.M. Dalla, S. Prasad, F. Pacifici et al. // *Proceedings of the IEEE.* — Vol. 103, N 9. — 2015. — P. 1585–1601.
43. *Gómez-Chova L.* Multimodal classification of remote sensing images: A review and future directions / L. Gómez-Chova, D. Tuia, G. Moser, G. Camps-Valls // *Proceedings of the IEEE.* — Vol. 103, N 9. — 2015. — 1560–1584.
44. *Kuwata K.* Estimating crop yields with deep learning and remotely sensed data / K. Kuwata, R. Shibasaki // *Geoscience and Remote Sensing Symposium (IGARSS).* — 2015. — P. 858–861.

45. *McNairn H.* Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2 / H. McNairn, A. Kross, D. Lapen, R. Caves, J. Shang // *International Journal of Applied Earth Observation and Geoinformation*. — Vol. 28. — 2014. — P. 252–259.
46. *Khatami R.* A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research / R. Khatami, G. Mountrakis, S.V. Stehman // *Remote Sensing of Environment*. — Vol. 177. — 2016. — P. 89–100.
47. *Abdiansah A.* Time Complexity Analysis of Support Vector Machines (SVM) in LibSVM / A. Abdiansah, R. Wardoyo // *International Journal of Computer Applications*. — Vol. 128, N 3. — 2015. — P. 28–34.
48. *Chapelle O.* Training a Support Vector Machine in the Primal / O. Chapelle // *Neural Computation*. — Vol. 19, N 5. — 2007. — P. 1155–1178.
49. *Sun X.* meProp: Sparsified back propagation for accelerated deep learning with reduced overfitting / X. Sun, X. Ren, S. Ma, H. Wang // *arXiv preprint arXiv*. — Vol. 1706, N 06197. — 2017. — P. 1–10.
50. *Gallagher M.* Multi-layer perceptron error surfaces: visualization, structure and mod-elling / M. Gallagher // *Doctoral dissertation*. — 2000.
51. *Han M.* (2012). Remote sensing image classification based on neural network ensemble algorithm / M. Han, X. Zhu, W. Yao // *Neurocomputing*. — Vol. 78, N 1. — 2012. — P. 133–138.
52. *Huang X.* An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery / X. Huang, L. Zhang // *IEEE Transactions on Geoscience and Remote Sensing*. — Vol. 51, N 1. — 2013. — P. 257–272.
53. *Lavreniuk M.S.* Large-scale classification of land cover using retrospective satellite data / M.S. Lavreniuk, S.V. Skakun, A.J. Shelestov et al. // *Cybernetics and Systems Analysis*. — Vol. 52, N 1. — 2016. — P. 127–138.
54. *Kussul N.* Parcel-based crop classification in ukraine using landsat-8 data and sentinel-1A data / N. Kussul, G. Lemoine, F.J. Gallego et al. // *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. — Vol. 9, N 6. — 2016. — P. 2500–2508.
55. *Chen Y.* Deep learning-based classification of hyperspectral data / Y. Chen, Z. Lin, X. Zhao et al. // *IEEE Journal of Selected topics in applied earth observations and remote sensing*. — Vol. 7, N 6. — 2014. — P. 2094–2107.
56. *Zhao W.* Learning multiscale and deep representations for classifying remotely sensed imagery / W. Zhao, S. Du // *ISPRS Journal of Photogrammetry and Remote Sensing*. — Vol. 113. — 2016. — P. 155–165.
57. *Kussul N.N.* Land cover changes analysis based on deep machine learning technique / N.N. Kussul, N.S. Lavreniuk, A.Y. Shelestov et al. // *Journal of Automation and Information Sciences*. — Vol. 48, N 5. — 2016. — P. 42–54.
58. *Kussul N.* Geospatial Intelligence and Data Fusion Techniques for Sustainable Development Problems / N. Kussul, A. Shelestov, R. Basarab et al. // *ICTERI*. — 2015. — P. 196–203.
59. *Zhang F.* Saliency-guided unsupervised feature learning for scene classification / F. Zhang, B. Du, L. Zhang // *IEEE Transactions on Geoscience and Remote Sensing*. — Vol. 53, N 4. — 2015. — P. 2175–2184.
60. *Zhang F.* Scene classification via a gradient boosting random convolutional network framework / F. Zhang, B. Du, L. Zhang // *IEEE Transactions on Geoscience and Remote Sensing*. — Vol. 54, N 3. — 2016. — P. 1793–1802.
61. *Lary D.J.* Machine learning in geosciences and remote sensing / D.J. Lary, A.H. Alavi, A.H. Gandomi, A.L. Walker // *Geoscience Frontiers*. — Vol. 7, N 1. — 2016. — P. 3–10.
62. *Peña-Barragán J.M.* Object-based crop identification using multiple vegetation indices, textural features and crop phenology / J.M. Peña-Barragán, M.K. Ngugi, R.E. Plant, J. Six // *Remote Sensing of Environment*. — Vol. 115, N 6. — 2011. — P. 1301–1316.

63. *Mnih V.* Learning to detect roads in high-resolution aerial images / V. Mnih, G. Hinton // Computer Vision–ECCV. — 2010. — P. 210–223.
64. *Mnih V.* Learning to label aerial images from noisy data / V. Mnih, G. Hinton // Proceedings of the 29th International Conference on Machine Learning (ICML-12). — 2012. — P. 1–8.
65. *Geng J.* High-resolution SAR image classification via deep convolutional autoencoders / J. Geng, J. Fan, H. Wang et al. // IEEE Geoscience and Remote Sensing Letters. — Vol. 12, N 11. — 2015. — P. 2351–2355.
66. *Chen Y.* Spectral–spatial classification of hyperspectral data based on deep belief network / Y. Chen, X. Zhao, X. Jia // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. — Vol. 8, N 6. — 2015. — P. 2381–2392.
67. *Liang H.* Hyperspectral imagery classification using sparse representations of convolutional neural network features / H. Liang, Q. Li // Remote Sensing. — Vol. 8, N 2. — 2016. — P. 1–16.
68. *Lyu H.* Learning a transferable change rule from a recurrent neural network for land cover change detection / H. Lyu, H. Lu, L. Mou // Remote Sensing. — Vol. 8, N 6. — 2016. — P. 1–22.
69. *Ishii T.* Surface object recognition with CNN and SVM in Landsat 8 images / T. Ishii, R. Nakamura, H. Nakada et al. // In Machine Vision Applications (MVA), 14th IAPR International Conference. — 2015. — P. 341–344.
70. *Deng L.* Deep learning: methods and applications / L. Deng, D. Yu // Foundations and Trends® in Signal Processing. — Vol. 7, N 3–4. — 2014. — P. 197–387.
71. *Längkvist M.* Classification and segmentation of satellite orthoimagery using convolutional neural networks / M. Längkvist, A. Kiselev, M. Alirezaie, A. Loutfi // Remote Sensing. — Vol. 8, N 4. — 2016. — P. 1–21.
72. *Dauphin Y.N.* Identifying and attacking the saddle point problem in high-dimensional non-convex optimization / Y.N. Dauphin, R. Pascanu, C. Gulcehre et al. // In Advances in neural information processing systems. — 2014. — P. 2933–2941.
73. *Choromanska A.* The loss surfaces of multilayer networks / A. Choromanska, M. Henaff, M. Mathieu et al. // In Artificial Intelligence and Statistics. — 2015. — P. 192–204.
74. *Hu W.* Deep convolutional neural networks for hyperspectral image classification / W. Hu, Y. Huang, L. Wei et al. // Journal of Sensors. — 2015. — P. 1–13.

Надійшла 29.11.2017