

DOI: <https://doi.org/10.33216/1998-7927-2022-275-5-16-20>

УДК: 004.89:504.06

МЕТОДОЛОГІЯ ІДЕНТИФІКАЦІЇ ШЛАМОНАКОПИЧУВАЧІВ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ НА ГЕОПРОСТОРОВИХ ЗОБРАЖЕННЯХ

Критська Я.О., Хмельницький Д.Б., Білобородова Т.О.

MACHINE LEARNING-BASED SLUDGE IDENTIFICATION ON GEOSPATIAL IMAGES

Krytska Y.O., Khmelnytskyi D.B., Biloborodova T.O.

В останні роки спостерігається зростання випадків деформації промислових шламонакопичувачів, результатом якої у багатьох випадках є руйнівний вплив на навколишнє середовище та екосистему. Моніторинг шламонакопичувачів має вирішальне значення для запобігання руйнівним наслідкам деформації. Традиційні методи моніторингу вимагають великих ресурсів та також малоефективні для раннього виявлення потенційної деформації. Віддалений моніторинг на основі геопросторових зображень є перспективним напрямком для моніторингу шламонакопичувачів з метою раннього виявлення потенційної деформації. В роботі представлено формалізацію етапів та визначення методології моніторингу шламонакопичувачів на основі геопросторових зображень з використанням методів машинного навчання: визначено засоби моніторингу, формалізовано етапи, розроблено технічну блок-схему процесу. Визначена методологія включає наступні етапи: (1) формування вибірки та передобробка зображень, (2) класифікація з використанням алгоритмів машинного навчання, (3) валідація результатів класифікації та визначення моделі з найвищою точністю. Методологія заснована на використанні Google Earth Engine (GEE). Засоби платформи включають інтерактивний сервер додатків, що працює з відкритим каталогом даних, обчислювальне інтегроване середовище розробки, геопросторовий API (клієнтські бібліотеки надають оболонки Python і JavaScript для веб-API в архітектурі REST. Практичну реалізацію та оцінку якості запропонованої методології проведено на основі даних зображень шламонакопичувача содового виробництва ВАТ «Лисичанська сода». Проведено попередню обробку зображень: 1) відбір зображень без хмар, для покращення результатів подальшого аналізу, 2) об'єднання шарів зображень ДДЗ, 3) анотація об'єктів, 4) розділення даних на тестовий та тренувальний набори даних пікселів. Класифікацію реалізовано з використанням алгоритмів CART, RF та SVM Ефективність моделей визначено на основі точності ідентифікації. Найвища точність на тестових даних досягнута з використанням SVM, що становила 98,05%.

Ключові слова: класифікація зображень дистанційного зондування, машинне навчання, шламонакопичувач

Вступ. Шламонакопичувачі є великими спорудами, у яких зберігаються відходи, що утворюються внаслідок промислової діяльності підприємства [1]. Останнім часом зросла кількість випадків деформації шламонакопичувачів. Результатом деформації шламонакопичувача у багатьох випадках є руйнівний вплив на навколишнє середовище та екосистему. Моніторинг шламонакопичувачів має вирішальне значення для запобігання руйнівним наслідкам деформації. Традиційні методи моніторингу часто вимагають великих витрат часу та ретельного візуального огляду. Також, незважаючи на їхню надійність і точність, ранні сигнали потенційної деформації можуть бути не виявлені, оскільки контроль обмежений локальним моніторингом. Тому в ряді випадків традиційні методи моніторингу є недостатньо ефективними [2,3].

Геопросторові зображення, отримані за допомогою засобів дистанційного зондування Землі (ДЗЗ), наразі ефективно використовуються в різноманітних областях, таких як моніторинг поверхневих вод [4], у геології, картографії, екології, лісовому та сільському господарстві, океанології, метеорології та інших галузях [5,6], що показує великий потенціал їх застосування для віддаленого моніторингу шламонакопичувачів з метою ідентифікації потенційної деформації.

Супутник Landsat збирає та фіксує дані про нашу планету понад 40 років, що дозволяє проаналізувати зміни рельєфу тощо. Зі зростанням кількості зображень, можливості отримання інформації про зміщення поверхні постійно покращуються [7]. Також, розвиток хмарних рішень для обробки геопросторових зображень, таких як платформа Google Earth Engine (GEE) [8], надає додаткові потужності та ресурси для вирішення задачі моніторингу шламонакопичувачів. Класифікація даних геопросторо-

вих зображень з метою ідентифікації зони шламонакопичувача та визначення його площі дозволить відстежувати деформацію шламонакопичувача на великих площах, і є основою використання геопросторових зображень ДЗЗ як додаткового інструменту до традиційних засобів локального моніторингу стану шламонакопичувачів.

Для класифікації супутникових зображень й вилучення інформації використовуються два підходи: на основі об'єктів та на основі пікселів [9]. Обидва методи можуть бути інтегровані з машинним навчанням для ідентифікації шламонакопичувачів. Методи машинного навчання вважаються ефективним методом для ідентифікації об'єктів дистанційного зондування у класифікації зображень та розпізнанні об'єктів [10].

У роботі [11] проаналізовано продуктивність різних методів машинного навчання, таких як Random Forest (RF), K-найближчих сусідів, SVM на основі методу Демпстера-Шефера. У роботі [9] представлений аналіз методів машинного навчання, таких як RF [12], SVM [13] при застосуванні на оптичних чи багатодіапазонних супутникових зображеннях. RF широко використовується в різних додатках дистанційного зондування [15], є менш чутливим до складних наборів даних та згідно з [9] показує ефективні результати класифікації супутникових зображень.

Метою роботи є дослідження можливостей і перспектив використання методів машинного навчання з метою класифікації геопросторових зображень для моніторингу шламонакопичувачів.

Завдання дослідження визначено наступним чином:

- формалізація етапів та визначення методології моніторингу шламонакопичувачів на основі геопросторових зображень з використанням методів машинного навчання: визначення засобів, формалізація етапів, розробка технічної блок-схеми процесу;

- застосування запропонованої методології до ідентифікації шламонакопичувача на геопросторових зображеннях.

Методологія моніторингу шламонакопичувачів на основі геопросторових зображень ДЗЗ. Запропонована методологія заснована на використанні GEE. Засоби платформи включають інтерактивний сервер додатків, що працює з відкритим каталогом даних, обчислювальне інтегроване середовище розробки, геопросторовий API (клієнтські бібліотеки надають оболонки Python і JavaScript для веб-API в архітектурі REST).

Методологія включає наступні етапи: (1) формування вибірки та передобробка зображень, (2) класифікація з використанням алгоритмів машинного навчання, (3) валідація результатів класифікації та визначення моделі з найвищою точністю. Розроблена технічна блок-схема процесу представлена на рис.1.

Формування вибірки та передобробка зображень. Процедура формування вибірки проводиться з використанням зображень ДЗЗ, у тому числі від Landsat, MODIS, Sentinel-1 та Sentinel-2, що надаються у каталогах GEE. Місія Sentinel-2 збирає багатоспектральні зображення з високою роздільною здатністю, які корисні для широкого кола додатків, включаючи моніторинг рослинності, ґрунтового та водного покриву, зміни земного покриву, а також гуманітарні ризики та ризики стихійного лиха [16]. Згідно з каталогом даних GEE, ці зображення віднесені до рівня 2A, а зображення поверхневого відображення пройшли попередню обробку для корекції випромінювання, геометрії та атмосфери за даними GEE станом на 21 січня 2021 р. [17]. Проводиться попередня обробка зображень: 1) відбір зображень без хмар, для покращення результатів подальшого аналізу, 2) об'єднання шарів зображень ДЗЗ, 3) анотація об'єктів, 4) розділення даних на тестовий та тренувальний набори даних пікселів.

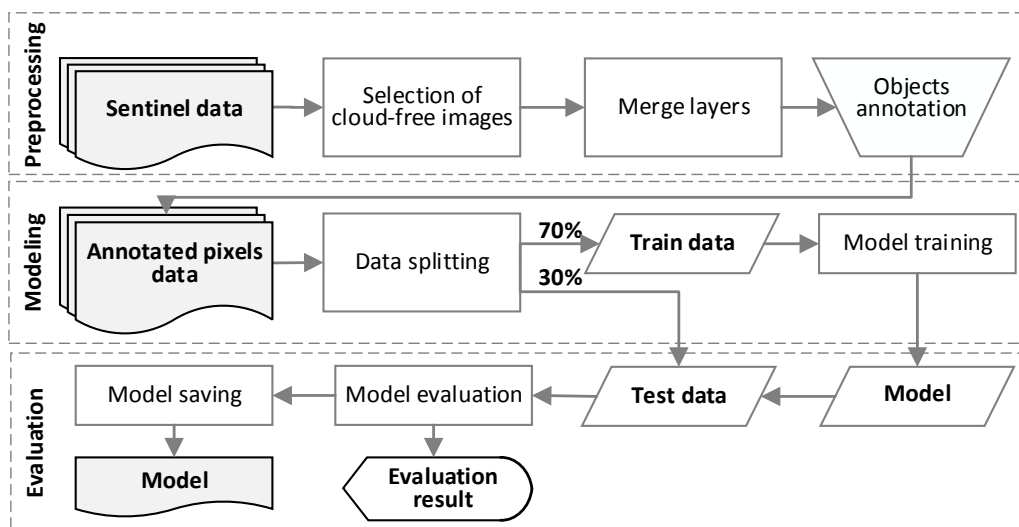


Рис. 1. Блок-схема моніторингу шламонакопичувачів з використанням методів машинного навчання на основі геопросторових зображень ДЗЗ

На основі спектральних і текстурних характеристик супутникових даних виділено п'ять категорій об'єктів зображень: дороги, рослинність, водоймища, сільськогосподарські культури та шламонакопичувачі. Об'єкти зображення є наборами зв'язаних пікселів, що мають однакове ціле значення. Процес анотації об'єктів є процедурою ідентифікації об'єкта зображення та присвоєння унікального ідентифікатора всім пікселям об'єкта. Результатом анотації є копія вхідного зображення з додатковим шаром міток, що з'єднує пікселі зі значенням ідентифікатора об'єкта на основі зв'язності пікселів у першому шарі зображення.

Усі точки вибірки об'єднані в одному шарі для формування набору даних, де 70% пікселів обрано випадковим чином для навчання моделі, а решта 30% для перевірки та оцінки точності алгоритму.

Класифікація з використанням алгоритмів машинного навчання. Для класифікації запропоновано використання наступних алгоритмів машинного навчання: CART, Random Forest (RF) та SVM.

Валідація результатів класифікації. Матриця плутанини є основним методом оцінки точності, який описує точність класифікації на тестових даних. На основі оцінки точності визначається оптимальна модель для ідентифікації шламонакопичувачів. На основі ідентифікованого об'єкту визначається площа шламонакопичувача. Зміни площі шламонакопичувача з плином часу можуть бути використані для моніторингу його деформації.

Ідентифікація шламонакопичувача заводу «Ліссода» на геопросторових зображеннях та визначення його площі. Оцінка якості запропонованої методології проведена на основі даних зображень шламонакопичувача заводу «Ліссода», який виділено червоним прямокутником на рис. 2.



Рис. 2. Географічне розташування шламонакопичувача заводу «Ліссода» на мапі

Вибірка сформована з використанням зображень Sentinel-2 від GEE. Для дослідження викорис-

тано дані контрольних зображень за 2015-2022 рік, а хвильовий код Quality Assessment 60 (QA60) використано в якості фільтру для зменшення каламутних частинок та усунення впливу хмар на зображеннях.

Реалізовано анотування наступних об'єктів зображень: дороги, рослинність, водоймища, сільськогосподарські культури та шламонакопичувач. Анотація шламонакопичувача проведена за двома характеристиками шламонакопичувача: відходи 1 - для відходів, що мають низьке залягання, та відходи 2 - для відходів, що мають високе залягання. Процес анотації об'єктів представлено на рис. 3, де а) анотація рослинності, б) анотація водойм, в) дороги, г) відходи 1, д) відходи 2, е) інше.

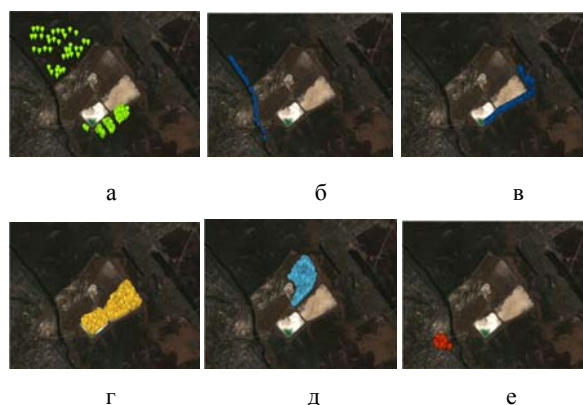


Рис. 3. Анотація об'єктів:
а – анотація рослинності, б – анотація водойм, в – дороги,
г – відходи 1, д – відходи 2, е – інше

Проведено класифікацію з використанням алгоритмів CART, RF та SVM. Приклад результатів класифікації представлено на рис. 4, де а) CART, б) RF та в) SVM.

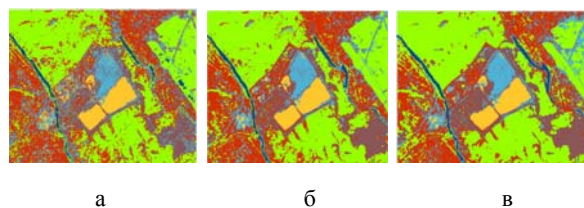


Рис. 4. Відокремлення області шламонакопичувача:
а – CART, б – RF, в – SVM

Проведено відокремлення класифікованої області шламонакопичувача відходів від загальної мапи. Приклад результатів відокремлення представлено на рис. 5, де а) CART, б) RF та в) SVM.

Порівняльний аналіз алгоритмів ідентифікації шламонакопичувача проведено шляхом оцінки точності класифікації (дивись таблицю).

Найвищу точність ідентифікації шламонакопичувача на зображеннях, що дорівнює 98,05%, показав алгоритм SVM.

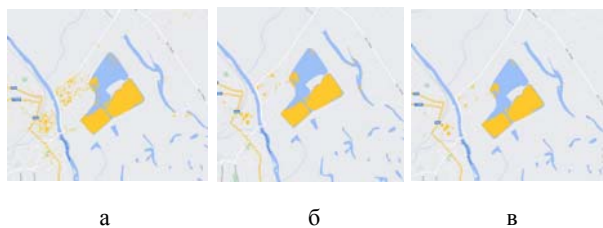


Рис. 5. Відокремлення області шламонакопичувача: а – CART, б – RF, в – SVM

Таблиця

Результати оцінки точності класифікації	
Алгоритм	Точність, %
CART	73,12
RF	85,62
SVM	98,05

Висновки. В роботі представлено формалізацію етапів та визначення методології моніторингу шламонакопичувачів на основі геопросторових зображень з використанням методів машинного навчання: визначено засоби моніторингу, формалізовано етапи, розроблено технічну блок-схему процесу. Представлено практичну реалізацію формалізованих етапів методології на прикладі ідентифікації шламонакопичувача заводу «Ліссода», що включила наступні етапи: (1) формування вибірки та передобробка зображень, (2) класифікація з використанням алгоритмів машинного навчання, (3) валідація результатів класифікації та визначення моделі з найвищою точністю. Практична реалізація проведена з використанням платформи Google Earth Engine. Попередня обробка зображень включала: 1) відбір зображень без хмар, для покращення результатів подальшого аналізу, 2) об'єднання шарів зображень ДДЗ, 3) анотація об'єктів, 4) розділення даних на тестовий та тренувальний набори даних пікселів. Класифікацію на основі даних пікселів реалізовано з використанням алгоритмів CART, RF та SVM. Ефективність моделей визначено на основі точності ідентифікації шламонакопичувача. Порівняльний аналіз алгоритмів ідентифікації шламонакопичувача проведено шляхом оцінки точності класифікації. Найвищу точність ідентифікації шламонакопичувача на зображеннях, що дорівнює 98,05%, показав алгоритм SVM

Л і т е р а т у р а

- Celenza, G., 2019. Industrial waste treatment process engineering: Biological Processes. CRC Press.
- Grebby S. et al. Advanced analysis of satellite data reveals ground deformation precursors to the Brumadinho Tailings Dam collapse. *Communications Earth & Environment*. – 2021. – Т. 2. – №. 1. – С. 1-9.
- Robertson, P. K., L.d. Melo, D.J. Williams, and G.W. Wilson. Report of the Expert Panel on the Technical Causes of the Failure of Feijão Dam I. 2019. URL:

<https://www.resolutionmineeis.us/documents/robertson-et-al-2019> (дата звернення 10.12.2022).

- Критська Я. О., Білобородова Т.О. Дослідження методів обробки та аналізу геопросторових зображень для віддаленого моніторингу поверхневих вод. *Вісник СХУ імені Володимира Даля*. – 2022. – №. 1 (271). – С. 11-17.
- Imran, Shabir Ahmad, Do Hyeun Kim. Quantum GIS based descriptive and predictive data analysis for effective planning of waste management. *IEEE*. – 2020. – Т. 8. – С. 46193-46205.
- Kodwo Miezah, Kwasi Obiro-Danso. Municipal solid waste characterization and quantification as a measure towards effective waste management in Ghana. *Waste management*. – 2015. – Т. 46. – С. 15-27.
- Cigna, F., & Tapete, D. (2021). Present-day land subsidence rates, surface faulting hazard and risk in Mexico City with 2014–2020 Sentinel-1 IW InSAR. *Remote Sensing of Environment*. – 2021. – Т. 253. – С. 112161.
- Google Earth Engine. Developers.google.com URL: <https://developers.google.com/earth-engine> (дата звернення 10.12.2022).
- Ghorbanzadeh O., Blaschke T., Gholamnia K., Meena S.R., Tiede D., Aryal J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*. – 2019. – Т. 11. – №. 2. – С. 196.
- Roodposhti M. S., Aryal J., Bryan B. A. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. *Environmental modelling & software*. – 2019. – Т. 112. – С. 70-81.
- Mezaal M. R., Pradhan B., Rizeei H. M. Improving landslide detection from airborne laser scanning data using optimized Dempster-Shafer. *Remote Sensing*. – 2018. – Т. 10. – №. 7. – С. 1029.
- Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M. Machine learning feature selection methods for landslide susceptibility mapping. *Mathematical geosciences*. – 2014. – Т. 46. – №. 1. – С. 33-57.
- Huang Y., Zhao L. Review on landslide susceptibility mapping using support vector machines. *Catena*. – 2018. – Т. 165. – С. 520-529.
- Kavzoglu T., Colkesen I., Sahin E. K. Machine learning techniques in landslide susceptibility mapping: a survey and a case study. *Landslides: Theory, practice and modelling*. – 2019. – С. 283-301.
- Melville B., Lucieer A., Aryal J. Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. *International journal of applied earth observation and geoinformation*. – 2018. – Т. 66. – С. 46-55.
- Shaikh M. A. Using GIS in Solid Waste Management Planning: A case study for Aurangabad, India. – Institutionen för datavetenskap, 2006.
- Sentinel-2 MSI: MultiSpectral Instrument, Level-2A. Developers.google.com URL: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR. (дата звернення 10.12.2022).

References

1. Celenza, G., 2019. Industrial waste treatment process engineering: Biological Processes. CRC Press.
2. Grebby S. et al. Advanced analysis of satellite data reveals ground deformation precursors to the Brumadinho Tailings Dam collapse. Communications Earth & Environment. – 2021. – Т. 2. – №. 1. – С. 1-9.
3. Robertson, P. K., L.d. Melo, D.J. Williams, and G.W. Wilson. Report of the Expert Panel on the Technical Causes of the Failure of Feijão Dam I. 2019. URL: <https://www.resolutionmineeis.us/documents/robertson-et-al-2019> (дата звернення 10.12.2022).
4. Krytska Y.O., Biloborodova T.O. Geospatial images processing and analysis for remote surface water monitoring. Herald of East Ukrainian National University. Severodonetsk : EUNU. – 2022. – №. 1 (271). – С. 11-17.
5. Imran, Shabir Ahmad, Do Hyeun Kim. Quantum GIS based descriptive and predictive data analysis for effective planning of waste management. IEEE. – 2020. – Т. 8. – С. 46193-46205.
6. Kodwo Miezah, Kwasi Obiro-Danso. Municipal solid waste characterization and quantification as a measure towards effective waste management in Ghana. Waste management. – 2015. – Т. 46. – С. 15-27.
7. Cigna, F., & Tapete, D. (2021). Present-day land subsidence rates, surface faulting hazard and risk in Mexico City with 2014–2020 Sentinel-1 IW InSAR. Remote Sensing of Environment. – 2021. – Т. 253. – С. 112161.
8. Google Earth Engine. Developers.google.com URL: <https://developers.google.com/earth-engine> (дата звернення 10.12.2022).
9. Ghorbanzadeh O., Blaschke T., Gholamnia K., Meena S.R., Tiede D., Aryal J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing. – 2019. – Т. 11. – №. 2. – С. 196.
10. Roodposhti M. S., Aryal J., Bryan B. A. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. Environmental modelling & software. – 2019. – Т. 112. – С. 70-81.
11. Mezaal M. R., Pradhan B., Rizeei H. M. Improving landslide detection from airborne laser scanning data using optimized Dempster-Shafer. Remote Sensing. – 2018. – Т. 10. – №. 7. – С. 1029.
12. Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M. Machine learning feature selection methods for landslide susceptibility mapping. Mathematical geosciences. – 2014. – Т. 46. – №. 1. – С. 33-57.
13. Huang Y., Zhao L. Review on landslide susceptibility mapping using support vector machines. Catena. – 2018. – Т. 165. – С. 520-529.
14. Kavzoglu T., Colkesen I., Sahin E. K. Machine learning techniques in landslide susceptibility mapping: a survey and a case study. Landslides: Theory, practice and modelling. – 2019. – С. 283-301.
15. Melville B., Lucieer A., Aryal J. Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. International journal of applied earth observation and geoinformation. – 2018. – Т. 66. – С. 46-55.
16. Shaikh M. A. Using GIS in Solid Waste Management Planning: A case study for Aurangabad, India. – Institutionen för datavetenskap, 2006.
17. Sentinel-2 MSI: MultiSpectral Instrument, Level-2A. Developers.google.com URL: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR. (дата звернення 10.12.2022).

Krytska Y.O., Khmelnytskyi D.B., Biloborodova T.O. Machine learning-based sludge identification on geospatial images.

In recent years, there has been an increase in cases of deformation of industrial sludge, which in many cases has a devastating impact on the environment and the ecosystem. Monitoring of Waste accumulation sites is crucial to prevent the destructive effects of deformation. Traditional monitoring methods require large resources and are also ineffective for early detection of potential deformation. Remote monitoring based on geospatial images is a promising area for monitoring of sludge caps with the purpose of early detection of potential deformation. The work presents the formalization of stages and determination of the methodology of Waste accumulation sites monitoring based on geospatial images using machine training methods: identified monitoring tools, formalized stages, developed a technical block diagram of the process. The defined methodology includes the following stages: (1) sampling and imagery, (2) classification using machine learning algorithms, (3) Validation of classification and model determination with the highest accuracy. The methodology is based on the use of Google Earth Engine (GEE). Platform tools include an interactive server of applications with an open data directory, computing integrated development environment, geospatial application programming interface (client libraries provide Python and JavaScript shells for web-API in representational state transfer (REST) architecture). The practical implementation and quality assessment of the proposed methodology was carried out based on image data of Waste accumulation sites of the open joint-stock company (OJSC) «Lysychanska soda». Pre-processing of images: 1) selection of images without clouds, to improve the results. 2) unification of layers of remote sensing images. 3) annotation of objects, 4) separation of data into test and training sets of pixel data. The classification is implemented using Classification and Regression Trees (CART), Random Forest (RF) and Support Vector Machine (SVM) algorithms. The effectiveness of the models is determined based on the accuracy of identification. The highest accuracy on test data was achieved using SVM, which was 98.05%.

Keywords: remote sensing images classification, machine learning, waste accumulation sites

Критська Я.О. – д.ф., доцент кафедри комп'ютерних наук та інженерії Східноукраїнського національного університету імені Володимира Даля, e-mail: kritskayana@gmail.com

Хмельницький Д.Б. – магістр кафедри комп'ютерних наук та інженерії Східноукраїнського національного університету імені Володимира Даля, e-mail: dimon22052000@gmail.com

Білобородова Т.О. – к.т.н., доцент, доцент кафедри комп'ютерних наук та інженерії Східноукраїнського національного університету імені Володимира Даля, e-mail: beloborodova.t@gmail.com

Стаття подана 25.09.2022.