



So sánh sản phẩm phân loại lớp phủ toàn cầu Dynamic World và sản phẩm phân loại dựa trên thuật toán SVM tại khu vực Hà Nội

Võ Ngọc Hải¹, Đỗ Như Hiệp^{1*}, Hà Thị Minh Tuyền²

¹Đại học Tài nguyên và Môi trường Hà Nội, 41A Phú Diễn, Bắc Từ Liêm, Hà Nội

²Ban quản lý khu các làng dân tộc, Đông Mô, Sơn Tây, Hà Nội

Email tác giả liên hệ: dnhiep@hunre.edu.vn

<https://doi.org/10.5281/zenodo.14302467>

Tóm tắt

Hiện nay, việc sử dụng các sản phẩm về phân loại lớp phủ bề mặt trên toàn cầu đã có thể đem lại những hiệu quả nhất định cho các công tác nghiên cứu, ứng dụng cho nhiều ngành. Trong đó, các sản phẩm có độ phân giải không gian cao đạt đến 10m có khả năng đem lại hiệu quả trong quản lý và nghiên cứu, phân tích dữ liệu trong quản lý đất đai. Các sản phẩm này gồm nhiều nguồn, nổi bật là sản phẩm Dynamic World (DW), Esri's Land Cover (Esri) và ESA's World Cover (EWC)... là các sản phẩm sử dụng tư liệu ảnh vệ tinh và tổng hợp kết quả phân loại lớp phủ trên toàn cầu với độ phân giải cao và thời gian gần thực. Nghiên cứu này sử dụng mô hình trí tuệ nhân tạo, thuật toán học máy thực hiện phân loại lớp phủ bề mặt khu vực Hà Nội bằng ảnh Sentinel-2 và so sánh kết quả với các sản phẩm đã có trên thế giới để đánh giá hiệu quả sử dụng. Kết quả cho thấy rằng các sản phẩm trên tương đối hiệu quả và có thể sử dụng cho một số lĩnh vực chuyên ngành tại Việt Nam, nhất là đối với lĩnh vực quản lý đất đai.

Từ khóa: Lớp phủ bề mặt, Sentinel, Google Earth Engine, Dynamic World, World Cover

Ngày nhận bài: 25/09/2024

Ngày sửa lại: 19/10/2024

Ngày chấp nhận đăng: 22/10/2024

Ngày xuất bản: 30/12/2024

Comparison of the Dynamic World global classification product with the SVM algorithm-based classification product in the Hanoi area

Vo Ngoc Hai¹, Do Nhu Hiep^{1*}, Ha Thi Minh Tuyen²

¹Hanoi University Natural Resources and Environment, 41A Phu Dien, Bac Tu Liem, Hanoi, Vietnam

²Village Forethnic, Dong Mo, Son Tay, Hanoi, Vietnam

Corresponding Author Email: dnhiep@hunre.edu.vn

Abstract

Currently, the use of global land cover products can yield certain benefits for research and applications across various industries. Among these, products with high spatial resolution, reaching up to 10 meters, have the potential to provide significant advantages in land management and data analysis. These products encompass various sources, with notable examples being Dynamic World (DW), Esri's Land Cover (Esri) and ESA's World Cover (EWC)... which utilize satellite imagery and synthesize global land cover results with high resolution and near-real-time updates. This study employs artificial intelligence models and machine learning algorithms to classify land cover in the Hanoi area using Sentinel-2 imagery and compares the results with existing global products to assess their effectiveness. The results indicate that these products are relatively effective and can be applied to specific specialized fields in Vietnam, especially on the land management.

Keywords: Land cover, Sentinel, Google Earth Engine, Dynamic World, World Cover

Submission received: 25/09/2024

Revised: 19/10/2024

Accepted: 22/10/2024

Published: 30/12/2024

1. Mở đầu

Có nhiều phương pháp đã được áp dụng tỏ ra rất hiệu quả để phục vụ việc phân loại các lớp phủ bề mặt, các lớp phủ sử dụng đất (Land Use Land Cover – LULC) từ ảnh viễn thám. Một trong những phương pháp, công nghệ tiên tiến hiện nay là ứng dụng công nghệ trí tuệ nhân tạo (AI - Artificial Intelligence) và các thuật toán học máy nhằm nâng cao khả năng tự động hóa, tự chủ trong xử lý dữ liệu của người dùng, giảm bớt phụ thuộc vào các hệ thống phần mềm chuyên dụng có giá thành cao. Trên thế giới cũng như tại Việt Nam đã có những nghiên cứu ứng dụng AI, học máy để thực hiện phân loại các lớp phủ bề mặt [1, 2]. Hiện nay, các sản phẩm này cũng được tạo ra và lưu trữ trên các nền tảng điện toán đám mây có thể truy cập miễn phí để lập bản đồ lớp phủ sử dụng đất [3, 4]. Đã có nhiều thuật toán được ứng dụng một cách hiệu quả trong việc phân loại



lớp phủ bề mặt đất từ ảnh vệ tinh như Maximum Likelihood Classifier (MLC), Minimum Distance Classifier (MDC), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Classification and regression tree (Cart) và Random Forest (RF). Trong đó, các thuật toán Cart, RF và SVM là các thuật toán đã áp dụng cho việc học máy được sử dụng nhiều trong khai phá cơ sở dữ liệu ảnh vệ tinh trực tuyến từ Google Earth Engine (GEE). Việc khai thác các thuật toán học máy đã được nhiều nghiên cứu, ứng dụng cho thấy tính hiệu quả trong việc phân loại lớp phủ bề mặt từ ảnh vệ tinh [5, 6]. Các sản phẩm LULC ở phạm vi toàn cầu đã được tạo ra nhanh chóng nhờ dữ liệu vệ tinh có độ phân giải trung bình và độ phân giải tương đối cao có sẵn như Landsat và Sentinel. Trên thế giới có những sản phẩm toàn cầu như Google's Dynamic World (DW), ESA's World Cover (EWC) và Esri's Land Cover (Esri)... các sản phẩm là bản đồ LULC của DW, EWC, Esri được một số nghiên cứu đánh giá và kết quả cho thấy sự tương ứng về không gian mạnh mẽ (tức là ước tính diện tích gần bằng nhau) cho các lớp phủ mặt nước, khu vực công trình xây dựng, thực phủ và một số loại cây trồng. Tuy nhiên, kết quả đánh giá cũng cho thấy rằng sản phẩm của EWC thiên về việc ước tính các lớp phủ là lớp cỏ, trong khi đó Esri thiên về lớp phủ cây bụi còn sản phẩm DW thiên về các lớp phủ băng tuyết [7, 8].

Để có thể có các ứng dụng phù hợp từ những sản phẩm toàn cầu sẵn có trên, nghiên cứu này tiến hành phân loại lớp phủ bề mặt khu vực Hà Nội từ ảnh vệ tinh Sentinel-2 và tiến hành so sánh sản phẩm này với sản phẩm đã có của Google là DW. Việc phân loại lớp phủ bề mặt khu vực Hà Nội được thực hiện trên nền tảng điện toán đám mây GEE với thuật toán học máy RF và dữ liệu là ảnh vệ tinh Sentinel-2.

2. Dữ liệu nghiên cứu

Ảnh vệ tinh Sentinel-2 TOA ngày 18 tháng 5 năm 2023 có độ phân giải không gian 10m ở kênh ảnh toàn sắc được sử dụng để phân loại lớp phủ bề mặt khu vực Hà Nội. Ngoài ra còn sử dụng sản phẩm đã có là DW (của Google), đây cũng là sản phẩm sử dụng dữ liệu là ảnh vệ tinh Sentinel-2 để phân loại lớp phủ bề mặt.

DW là sản phẩm phân loại lớp phủ bề mặt từ ảnh vệ tinh Sentinel-2 TOA dựa trên phương pháp học máy, và có thể cập nhật lớp phủ toàn cầu từ 2 đến 5 ngày một lần tùy thuộc vào vị trí. DW được thành lập dựa trên nền tảng dữ liệu điện toán đám mây của Google Earth Engine và mô hình AI. Sản phẩm DW được xây dựng phù hợp với Nguyên tắc AI của Google. Việc phân loại lớp phủ bề mặt của sản phẩm DW được phát triển bằng cách sử dụng dữ liệu đào tạo của Hiệp hội địa lý quốc gia (National Geographic Society - NGS). Đây là nguồn dữ liệu được thành lập bởi dự án hợp tác giữa NGS và Google cùng với Viện Tài nguyên Thế giới (World Resources Institute VRI). Sản phẩm bao gồm 9 loại lớp phủ được mô tả tại bảng 1 dưới đây:

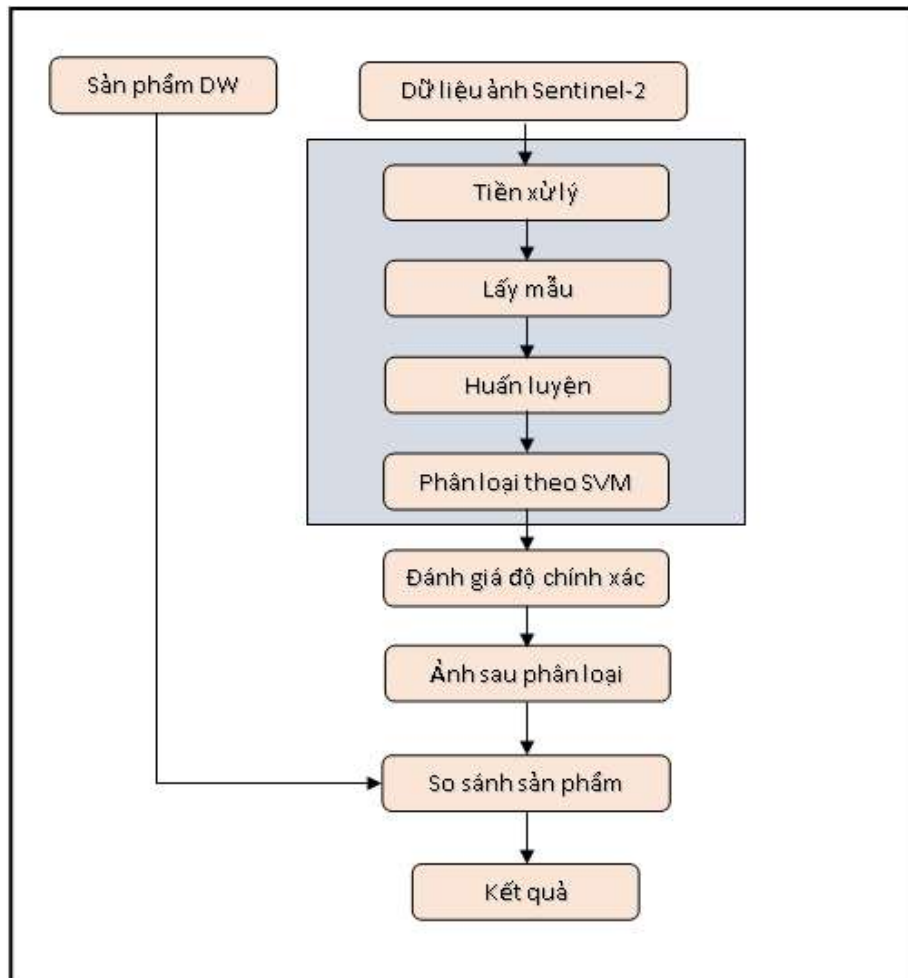
Bảng 1. Mô tả về các loại lớp phủ của DW

Stt	Tên lớp phủ	Mô tả
1	Mặt nước	Các vùng nước vĩnh viễn và theo mùa
2	Cây cối	Bao gồm rừng nguyên sinh và rừng thứ sinh, cũng như rừng trồng
3	Cỏ	Đồng cỏ tự nhiên, đồng cỏ chăn nuôi và công viên
4	Thảm thực vật ngập nước	Rừng ngập mặn và các hệ sinh thái ngập nước khác
5	Cây trồng	Bao gồm cây trồng và lúa
6	Cây bụi	Cây bụi và các thảm thực vật thưa thớt đến rậm rạp
7	Khu vực xây dựng	Các tòa nhà chung cư, đường xá, làng mạc
8	Sa mạc	Sa mạc, đất trống và đá lộ thiên
9	Băng, Tuyết	Băng, tuyết phủ vĩnh viễn và theo mùa

3. Phương pháp và kết quả nghiên cứu

Sản phẩm thứ nhất của nghiên cứu được thực hiện bằng cách sử dụng nền tảng GEE để trích xuất dữ liệu DW tháng 6 năm 2023, đây là sản phẩm có thời điểm trùng với sản phẩm 2 (do nhóm nghiên cứu tự phân loại) để có cơ sở so sánh đánh giá hai sản phẩm này với nhau.

Đối với sản phẩm hai, nhóm tác giả sử dụng nền tảng GEE và ngôn ngữ lập trình JavaScript để phân loại LULC của khu vực Hà Nội, Việt Nam. Việc phân loại được thực hiện bằng mô hình AI dựa trên thuật toán SVM. Dữ liệu sử dụng để phân loại là ảnh vệ tinh Sentinel-2 được thu nhận bằng cách tổng hợp trong tháng 6 năm 2023. Số lượng lớp phủ được phân loại là 6 lớp bao gồm: Đất trồng, Mặt nước, Khu vực xây dựng, Cây cối dày đặc, Cây bụi, Giao thông. Quy trình nghiên cứu được thể hiện tại hình 1. Kết quả nghiên cứu này sử dụng sản phẩm DW cùng với phân loại ảnh Sentinel 2 năm 2023 để so sánh, đánh giá khả năng sử dụng của các sản phẩm. Các sản phẩm toàn cầu DW đã được một số nghiên cứu đánh giá có khả năng sử dụng cho một số lĩnh vực cụ thể với mức độ tổng quan [9]. Thuật toán học máy SVM được sử dụng để phân loại các lớp phủ. Đối với khu vực Hà Nội nhóm nghiên cứu phân loại gồm 6 lớp phủ cơ bản như đã nêu ở trên.

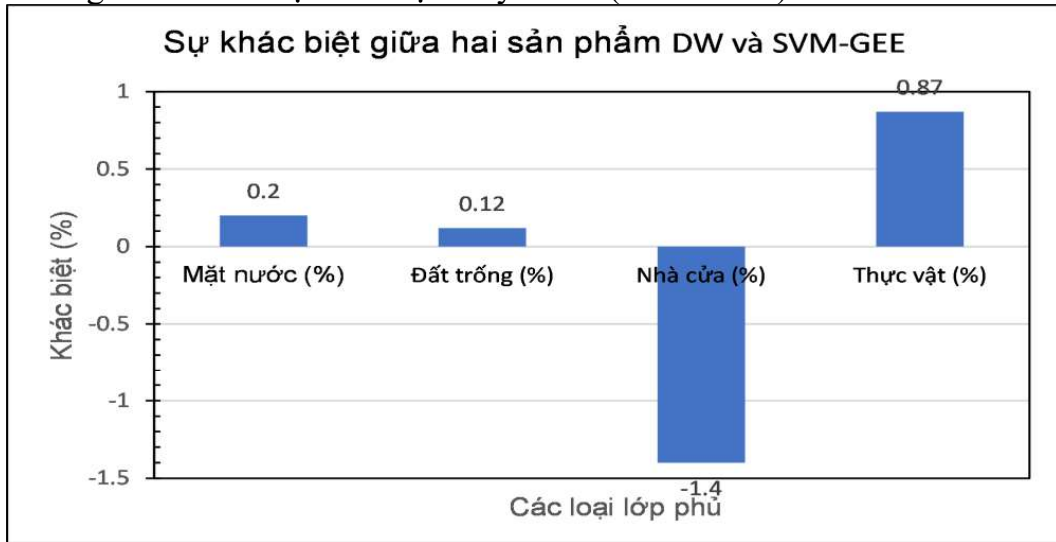


Hình 1. Sơ đồ tổng thể quy trình nghiên cứu

Kết quả đánh giá độ chính xác phân loại dựa trên các thông số như độ chính xác tổng thể (Overall Accuracy - OA) và hệ số Kappa của ma trận nhầm lẫn (Confusion-matrix). Trong đó hệ số Kappa có giá trị từ 0,4 đến 0,6 được đánh giá là đạt kết quả trung bình, giá trị từ lớn hơn 0,6 đến 0,8 là tốt và hơn 0,8 đến 1,0 là rất tốt [10]. Thực hiện đánh giá trên nền GEE với hàm errorMatrix, trong đó sử dụng 70% số lượng mẫu

dùng để phân loại ảnh và 30% số lượng mẫu dùng để kiểm tra đánh giá. Cụ thể, kết quả đánh giá độ chính xác đạt được giá trị OA và Kappa tương ứng là 0,95 và 0,91. Đây là các giá trị hoàn toàn có khả năng đáp ứng được yêu cầu trong phân loại các lớp phủ sử dụng đất từ dữ liệu ảnh vệ tinh.

Hình 2 thể hiện kết quả so sánh tỷ lệ các lớp phủ cơ bản là: Mặt nước, Đất trống, Khu vực xây dựng được phân loại theo các sản phẩm DW và sản phẩm được phân loại trên nền tảng GEE với thuật toán học máy SVM (GEE-SVM).



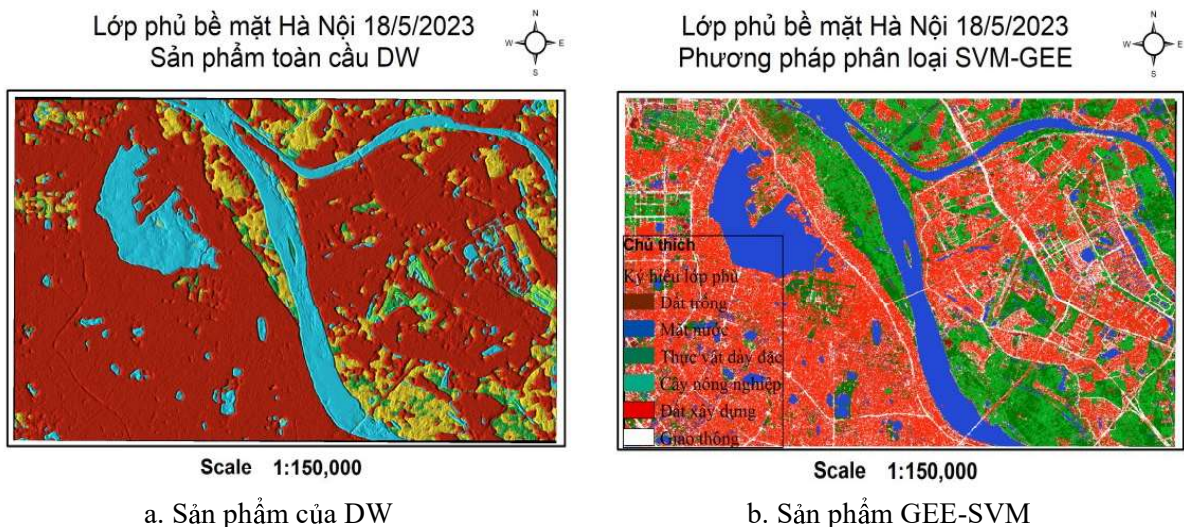
Hình 2. So sánh tỷ lệ khác biệt giữa các lớp phủ sử dụng đất

Các kết quả trên có sự khác biệt rõ ràng giữa các sản phẩm DW và nghiên cứu phân loại với thuật toán SVM trên nền tảng dữ liệu điện toán đám mây GEE.

Bảng 2. Kết quả khác biệt giữa sản phẩm lớp phủ toàn cầu DW và kết quả phân loại lớp phủ theo thuật toán SVM trên nền tảng GEE

Loại sản phẩm	Mặt nước (%)	Đất trống (%)	Nhà cửa (%)	Thực vật (%)
DW và SVM-GEE	0,2	0,12	-1,4	0,87

Sự khác biệt trên cũng có thể nhìn nhận một cách tương đối rõ ràng trên các hình ảnh của các sản phẩm phân loại lớp phủ sử dụng đất tại Hình 3.



a. Sản phẩm của DW

b. Sản phẩm GEE-SVM

Hình 3. Các sản phẩm LULC toàn cầu và kết quả thực hiện phân loại với thuật toán SVM theo mô hình AI trên nền tảng dữ liệu điện toán đám mây GEE

Các khác biệt về kết quả phân loại giữa các sản phẩm có thể do nhiều nguyên nhân bao gồm cả khách quan và chủ quan. Từ bảng kết quả cho thấy kết quả tương đối gần



nhanh hơn giữa sản phẩm DW và GEE-SVM. Để so sánh các kết quả DW và đánh giá độ tin cậy của sản phẩm nào tốt hơn với các điều kiện tại khu vực nghiên cứu tại thời điểm này là thực sự còn nhiều khó khăn. Tuy nhiên kết quả này cũng thể hiện sự tương đồng với một số nghiên cứu khác về so sánh, đánh giá sản phẩm LULC toàn cầu [11, 12]. Có nhiều nguyên nhân tồn tại trong việc đánh giá độ tin cậy của các sản phẩm toàn cầu DW để phục vụ các công tác nghiên cứu cũng như quản lý sử dụng. Một trong những nguyên nhân ban đầu là mỗi sản phẩm hướng đến một trong những mục tiêu sử dụng khác nhau, có các lớp phủ sử dụng đất khác nhau nên gây khó khăn trong việc đánh giá độ chính xác của sản phẩm. Với 9 lớp phủ của sản phẩm DW, và 6 lớp phủ được phân loại bằng ảnh Sentinel-2 tại khu vực Hà Nội, Việt Nam là khó khăn trong đánh giá độ tin cậy phân loại các lớp phủ của sản phẩm DW. Tuy nhiên nghiên cứu này cũng đã sử dụng 4 lớp phủ cơ bản trùng lặp với 4 lớp phủ của sản phẩm DW để so sánh, đánh giá các kết quả như đã phân tích ở trên. Kết quả đánh giá cũng đã cho thấy được một số sự tương đồng của 2 sản phẩm. Sản phẩm DW và GEE-SVM đều là sản phẩm được phân loại từ dữ liệu ảnh vệ tinh Sentinel-2 thu nhận ngày 18 tháng 5 năm 2023, đây cũng là các sản phẩm được coi như là sản phẩm gần thời gian thực (Near Real Time – NRT) nên khả năng tương đồng về kết quả phân loại có khả năng cao hơn như kết quả thể hiện tại Bảng 2. Trong đó lớp đất trống với diện tích bao phủ thấp nên sự khác nhau không lớn, chỉ là 0,12%; lớp Nhà cửa và lớp Thực vật nói chung có khác biệt lớn tới -1,4% và 0,87%. Riêng về lớp Mặt nước thể hiện sự tương đồng cao giữa hai sản phẩm với giá trị khác biệt là khoảng 0,2%.

4. Kết luận

Việc chuyển đổi phương pháp khai thác, phân tích dữ liệu ảnh vệ tinh trực tuyến và các sản phẩm toàn cầu phục vụ công tác quản lý đất đai đang ngày càng tỏ ra hiệu quả. Điều này cũng góp phần tích cực vào công cuộc chuyển đổi số nói chung hiện nay và của ngành Quản lý đất đai nói riêng.

Sản phẩm DW là một trong những sản phẩm về các lớp phủ toàn cầu có nhiều ưu điểm như là gần thời gian thực, đây là một trong những yếu tố rất quan trọng trong các phân tích, đánh giá phục vụ công tác quản lý nhà nước về đất đai, môi trường. Nghiên cứu này cũng cho thấy sự tương đồng về kết quả tốt hơn giữa sản phẩm DW và SVM-GEE.

Các kết quả phân loại ảnh đã được đánh giá và thể hiện được độ tin cậy, tuy nhiên để có những căn cứ chắc chắn hơn nữa, các nghiên cứu tiếp theo sau này nên có sự so sánh với các loại bản đồ như là bản đồ hiện trạng sử dụng đất qua các thời kỳ.

Ngoài ra, để nâng cao hơn nữa hiệu quả sử dụng các sản phẩm toàn cầu phục vụ công tác nghiên cứu, quản lý đất đai một cách nhanh chóng, chính xác, cần có những nghiên cứu đánh giá cụ thể của từng loại sản phẩm DW và các sản phẩm toàn cầu khác cho từng loại lớp phủ và đối với từng khu vực khác nhau.

Cam kết của các tác giả

Tất cả các tác giả có tên trong bài báo cam kết sự đồng thuận và không có xung đột lợi ích trong công bố khoa học tại bài báo này.

Tài liệu tham khảo

- [1] Alshari EA, Abdulkareem MB, Gawali BW, "Classification of land use/land cover using artificial intelligence (ANN-RF)", *Frontiers in Artificial Intelligence*, 5(964279), 2023.
- [2] Civco DL, "Artificial neural networks for land-cover classification and mapping" *International journal of geographical information science*, 7(2): 173-186, 1993.



- [3] Tùng ĐT, "KHAI THÁC TRỰC TUYẾN CƠ SỞ DỮ LIỆU ẢNH VỆ TINH, SO SÁNH THUẬT TOÁN HỌC MÁY VỀ PHÂN LOẠI LỚP PHỦ TRÊN NỀN GOOGLE EARTH ENGINE" *Hội thảo Quốc gia GIS & Viễn thám*, 2022.
- [4] Capolupo A, Monterisi C, Caporusso G, Tarantino E, "Extracting land cover data using GEE: A review of the classification indices", *Computational Science and Its Applications–ICCSA 2020: 20th International Conference, Cagliari, Italy, July 1–4, 2020, Proceedings, Part IV 20, Springer*, pp: 782-796, 2020.
- [5] MAO L, LI H, XUE X, LI J, LI M, "Developing remote sensing based methods for land cover change detection in National Parks from GEE platform: a case study from the Qianjiangyuan National Park pilot area", *OURNAL OF NANJING FORESTRY UNIVERSITY*, 46(2): 213, 2022.
- [6] Svoboda J, Štych P, Laštovička J, Paluba D, Kobliuk N, "Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia", *Remote Sensing*, 14(5): 1189, 2022.
- [7] Wang Z, Mountrakis G, "Accuracy Assessment of Eleven Medium Resolution Global and Regional Land Cover Land Use Products: A Case Study over the Conterminous United States", *Remote Sensing*, 15(12): 3186, 2023.
- [8] Venter ZS, Barton DN, Chakraborty T, Simensen T, Singh G, "Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover", *Remote Sensing*, 14(16): 4101, 2022.
- [9] Iye Zou, Paul D. Loprinzi, Jane Jie Yu, Lin Yang, Chunxiao Li, Albert S. Yeung, ZK, Shin-Yi Chiou and Tao Xiao, "Superior Effects of Modified Chen-Style Tai Chi versus 24-Style Tai Chi on Cognitive Function, Fitness, and Balance Performance in Adults over 55", *Brain Sci*, 10(9): 102, 2019.
- [10] Visa S, Ramsay B, Ralescu AL, Van Der Knaap EJM, "Confusion matrix-based feature selection", 710(120-127), 2011.
- [11] Orusa T, Cammareri D, Borgogno Mondino E, "A Scalable Earth Observation Service to Map Land Cover in Geomorphological Complex Areas beyond the Dynamic World: An Application in Aosta Valley (NW Italy)", *Applied Sciences*, 13(1): 390, 2022.
- [12] Villegas Rugele GM, Ochoa D, Menendez JM, Van Coillie F, "Evaluating the Applicability of Global LULC Products and an Author-Generated Phenology-Based Map for Regional Analysis: A Case Study in Ecuador's Ecoregions", *Land*, 12(5): 1112, 2023.

