

АЭРОКОСМИЧЕСКИЕ ИССЛЕДОВАНИЯ ЗЕМЛИ, ФОТОГРАММЕТРИЯ/AEROSPACE RESEARCH OF THE EARTH, PHOTOGRAMMETRY

DOI: <https://doi.org/10.60797/IRJ.2026.163.72>

РАЗРАБОТКА ПРИЛОЖЕНИЯ ДЛЯ ПОДДЕРЖКИ КАРТОГРАФИРОВАНИЯ ЗЕМНОГО ПОКРОВА С ИСПОЛЬЗОВАНИЕМ ПЛАТФОРМЫ GOOGLE EARTH ENGINE

Научная статья

Фам Д.^{1,*}, Нгуен Т.², Нгуен Д.³

¹ ORCID : 0009-0000-7429-584X;

² ORCID : 0000-0002-0982-9080;

³ ORCID : 0009-0003-9999-2272;

^{1,2,3} Московский государственный университет геодезии и картографии, Москва, Российская Федерация

* Корреспондирующий автор (lopmaybay[at]mail.ru)

Аннотация

В данной статье представлено веб-приложение, разработанное на платформе Google Earth Engine (GEE), предназначенное для полуавтоматического картографирования земного покрова (LULC). Приложение представляет собой виртуальную исследовательскую среду (VRE), которая интегрирует доступ к обширному каталогу общедоступных данных GEE, насчитывающему около 900 наборов, с набором алгоритмов классификации. Ключевой инновацией является внедрение модуля поддержки сбора обучающих выборок, использующего алгоритмы Спектрального углового отображения (SAM), Сходства на основе скалярного произведения (Similarity based on Dot Product), Случайный лес (Random Forest) и Максимальной энтропии (Maxent). Эти методы генерируют вспомогательные карты, которые визуализируют степень сходства или вероятности каждого пикселя с эталонными классами, что позволяет пользователю интерактивно повышать качество и репрезентативность обучающих данных. В статье подробно описывается архитектура приложения, теоретические основы и практическая реализация алгоритмов поддержки выборки. Также проводится критический анализ ограничений, связанных с вычислительными ресурсами платформы GEE, и предлагаются практические рекомендации по их преодолению. Приложение доступно по ссылке <https://ee-lopmaybay.projects.earthengine.app/view/lulc>.

Ключевые слова: Google Earth Engine, классификация земного покрова, слияние данных, AlphaEarth Foundations, машинное обучение.

DEVELOPMENT OF AN APPLICATION TO SUPPORT LAND COVER MAPPING USING THE GOOGLE EARTH ENGINE PLATFORM

Research article

Pham D.^{1,*}, Nguyen T.², Nguyen D.³

¹ ORCID : 0009-0000-7429-584X;

² ORCID : 0000-0002-0982-9080;

³ ORCID : 0009-0003-9999-2272;

^{1,2,3} Moscow State University of Geodesy and Cartography, Moscow, Russian Federation

* Corresponding author (lopmaybay[at]mail.ru)

Abstract

This article presents a web application developed on the Google Earth Engine (GEE) platform for semi-automatic land use and land cover (LULC) mapping. The application is a virtual research environment (VRE) that integrates access to an extensive catalogue of publicly available GEE data, comprising approximately 900 datasets, with a set of classification algorithms. A key innovation is the introduction of a training sample collection support module that uses Spectral Angular Mapping (SAM), Similarity based on Dot Product, Random Forest, and Maximum Entropy (Maxent) algorithms. These methods generate auxiliary maps that visualise the degree of similarity or probability of each pixel with reference classes, allowing the user to interactively improve the quality and representativeness of training data. The paper describes in detail the application architecture, theoretical foundations and practical implementation of sample support algorithms. It also provides a critical analysis of the limitations associated with the computational resources of the GEE platform and offers practical recommendations for overcoming them. The app is available at <https://ee-lopmaybay.projects.earthengine.app/view/lulc>.

Keywords: Google Earth Engine, land cover classification, data fusion, AlphaEarth Foundations, machine learning.

Введение

Современные науки о Земле переживают трансформационный период, обусловленный беспрецедентным ростом объемов данных дистанционного зондирования (ДЗЗ). Увеличение числа спутниковых миссий, повышение пространственного, временного и спектрального разрешения сенсоров привели к явлению, известному как «информационный потоп» (data deluge). Традиционные подходы к обработке данных, основанные на локальных вычислениях, оказываются неэффективными для анализа петабайтных архивов спутниковых снимков. В этом контексте облачные платформы, такие как Google Earth Engine (GEE), стали парадигматическим сдвигом, предоставив научному сообществу доступ к глобальным архивам данных и колоссальным вычислительным мощностям для их анализа в планетарном масштабе [1].

Картографирование и мониторинг земельного и растительного покрова (Land Use/Land Cover, LULC) являются фундаментальными задачами для широкого спектра приложений, включая управление природными ресурсами, оценку воздействия изменения климата, городское планирование и обеспечение продовольственной безопасности [2]. Платформа GEE значительно упростила решение этих задач, однако эффективность и точность получаемых карт LULC по-прежнему сильно зависят от качества и доступности инструментов анализа. Возникает острая необходимость в разработке гибких, мощных и доступных приложений, которые не только реализуют сложные алгоритмы, но и способствуют воспроизводимости научных результатов в соответствии с принципами «Открытой науки» (Open Science) и FAIR (находимости, доступности, совместимости и переиспользования) [3].

Проблема сбора обучающих данных

Несмотря на все достижения в области алгоритмов машинного обучения, краеугольным камнем контролируемой (supervised) классификации LULC остается качество обучающих данных. Точность, полнота и репрезентативность обучающей выборки напрямую определяют способность классификатора корректно разделять классы и обобщать результаты на всю исследуемую территорию [4]. Традиционные методы сбора обучающих выборок, такие как полевые работы или ручное дешифрирование снимков экспертом, обладают рядом существенных недостатков. Они чрезвычайно трудоемки, дорогостоящи и часто ограничены по пространственному охвату. Более того, ручной выбор полигонов подвержен субъективизму эксперта и может не охватывать всего спектрального разнообразия внутри каждого класса, что приводит к переобучению модели и снижению итоговой точности.

Таким образом, возникает фундаментальная проблема: как сделать процесс сбора и уточнения обучающих данных более объективным, эффективным и менее зависимым от априорных знаний эксперта? Необходимо разработать инструменты, которые бы ассистировали исследователю, направляя его внимание на наиболее информативные или неоднозначные участки изображения и помогая создавать более сбалансированные и полные обучающие выборки.

Основной целью данной работы является разработка веб-приложения на платформе GEE, представляющего собой Виртуальную Исследовательскую Среда (Virtual Research Environment, VRE) для картографирования LULC с инновационной функцией алгоритмической поддержки процесса сбора обучающих выборок.

Для достижения поставленной цели были сформулированы следующие задачи:

1. Представить теоретические основы и программную реализацию алгоритмов Спектрального углового отображения (SAM), Сходства на основе скалярного произведения (Similarity based on Dot Product), Случайный лес (Random Forest) и Максимальной энтропии (Maxent) в контексте их нового применения для создания вспомогательных карт поддержки сбора выборок.

2. Детально описать архитектуру и функциональные возможности разработанного приложения, включая его модульную структуру и механизм динамической интеграции данных.

3. Проанализировать основные вычислительные ограничения платформы GEE, с которыми могут столкнуться пользователи приложения, и предложить практические стратегии для их преодоления.

Разработанное в данной работе приложение представляют собой комплексные решения, которые предоставляют пользователям доступ к данным, программному обеспечению и вычислительным сервисам, скрывая при этом сложность базовой инфраструктуры [5]. Оно инкапсулирует сложность API GEE, предоставляя интуитивно понятный графический интерфейс для выполнения всего цикла картографирования LULC. Таким образом, оно вносит вклад в развитие научной киберинфраструктуры, демократизируя доступ к передовым методам геопространственного анализа и позволяя исследователям сосредоточиться на научных задачах, а не на технических аспектах программирования и управления данными.

Алгоритмическая основа для классификации и поддержки сбора обучающих выборок

3.1. Стандартные алгоритмы классификации

Приложение предоставляет пользователю доступ к широкому набору стандартных алгоритмов классификации, реализованных в API Google Earth Engine [6]. Эти алгоритмы можно разделить на две основные группы:

– Контролируемые (Supervised) алгоритмы: к ним относятся классификаторы, требующие наличия размеченной обучающей выборки. В приложении реализованы такие популярные методы, как Случайный Лес (Random Forest), Машина Опорных Векторов (SVM), Деревья Классификации и Регрессии (CART), Наивный Байесовский классификатор (Naive Bayes), Градиентный Бустинг (Gradient Tree Boost) и другие. Эти алгоритмы используются на заключительном этапе для создания итоговых карт LULC на основе тщательно подготовленной обучающей выборки.

– Неконтролируемые (Unsupervised) алгоритмы: эти методы, такие как K-Means, X-Means и Cobweb, не требуют обучающих данных и группируют пиксели в кластеры на основе их спектрального сходства. Хотя они могут быть полезны для предварительного анализа данных, в данном приложении их основная роль второстепенна, так как акцент сделан на повышении качества контролируемой классификации.

3.2. Инновационный модуль поддержки сбора обучающих выборок

Ключевым научным и практическим вкладом данной работы является не просто агрегация существующих классификаторов, а разработка нового рабочего процесса, основанного на инновационном применении нескольких алгоритмов для поддержки сбора обучающих данных. Вместо того чтобы использовать алгоритмы SAM, Random Forest и Maxent как конечные классификаторы, приложение перепрофилирует их ещё для создания промежуточных информационных продуктов — вспомогательных карт.

Этот подход реализует парадигму «человек-в-цикле» (human-in-the-loop), где система и эксперт работают в тандеме. Алгоритмы генерируют объективную, количественную оценку сходства или вероятности для каждого пикселя, а пользователь (эксперт) использует эту информацию для принятия осмысленных решений по корректировке обучающей выборки. Например, проанализировав карту сходства для класса «лес», пользователь может обнаружить удаленные участки леса с нетипичным спектральным откликом (например, другой породный состав или плотность),

которые были пропущены при первоначальном ручном дешифрировании. Включение этих участков в обучающую выборку значительно повышает ее репрезентативность и, как следствие, точность итоговой классификации, выполняемой более сложным алгоритмом, таким как Random Forest. Таким образом, ценность модуля заключается не в новизне самих алгоритмов, а в их интеграции в новый, итеративный рабочий процесс, направленный на фундаментальное улучшение качества входных данных для машинного обучения.

3.3. Реализация алгоритмов поддержки выборки

Метод Спектрального углового отображения (SAM - Spectral Angle Mapper)

SAM является физически обоснованным методом, который определяет спектральное сходство между двумя спектрами, рассматривая их как векторы в n -мерном пространстве, где n — количество спектральных каналов [7]. Мерой сходства служит угол между этими векторами. Чем меньше угол, тем больше сходство. Этот метод инвариантен к изменениям освещенности, поскольку угол между векторами не зависит от их длины (яркости). Угол α вычисляется по формуле [8]:

$$\alpha = \cos^{-1} \left(\frac{\sum_{i=1}^n t_i r_i}{\sqrt{\sum_{i=1}^n t_i^2} \sqrt{\sum_{i=1}^n r_i^2}} \right) \quad (1)$$

где t — вектор спектральных значений пикселя изображения, r — эталонный спектральный вектор, а n — количество каналов.

В контексте данного приложения, для каждого класса LULC (например, «лес») сначала вычисляется единый эталонный вектор r_{class} . Это делается путем усреднения спектральных векторов всех пикселей, попавших в обучающие полигоны данного класса. По сути, r_{class} представляет собой «спектральный центроид» или усредненный спектральный отклик класса. Затем для каждого пикселя t на всем изображении вычисляется его сходство с этим эталонным вектором. В коде для оптимизации вычислений используется не сам угол α , а непосредственно скалярное произведение числителя формулы, которое пропорционально косинусу угла и напрямую отражает сходство. В результате для каждого класса генерируется одноканальное изображение (карта сходства), где более высокие значения соответствуют пикселям, спектрально более похожим на «усредненного» представителя этого класса.

Метод сходства на основе скалярного произведения (Similarity based on Dot Product)

Данный метод оценивает спектральное сходство между пикселем изображения и эталонными спектрами классов, используя скалярное произведение векторов в n -мерном пространстве спектральных каналов. Мерой сходства служит не нормализованное скалярное произведение, которое зависит как от угла между векторами, так и от их длин (яркости). Таким образом, метод не является инвариантным к изменениям освещенности. Сходство между спектром пикселя изображения t и эталонным спектром r вычисляется по формуле:

$$\text{Similarity} = \sum_{i=1}^n t_i \cdot r_i \quad (2)$$

где:

- t_i — значение i -го спектрального канала пикселя;
- r_i — значение i -го канала эталонного спектра;
- n — количество спектральных каналов.

В отличие от SAM, где используется усредненный центроид, здесь сохраняется всё многообразие спектральных характеристик класса. Для каждого пикселя изображения вычисляется скалярное произведение с каждым эталонным спектром внутри класса. Из полученных значений выбирается максимальное значение, которое считается мерой сходства пикселя с классом. Это позволяет учесть внутриклассовую изменчивость. Пиксель относится к тому классу, для которого максимальное скалярное произведение (сходство) является наибольшим. Результат — тематическое изображение, где каждый пиксель помечен ID класса с наивысшим сходством. Для ускорения расчетов используется прямое скалярное произведение без нормализации. Это эквивалентно использованию косинуса угла между векторами, умноженного на их длины, что делает метод чувствительным к яркости.

Метод максимальной энтропии (Maxent)

Maxent — это метод машинного обучения, который изначально был разработан для моделирования экологических ниш и предсказания географического распространения видов [9]. Его ключевая особенность — способность работать с данными только о присутствии вида («presence-only data»), не требуя данных об его отсутствии. Принцип максимальной энтропии гласит, что из всех возможных вероятностных распределений, согласующихся с имеющимися данными (ограничениями), следует выбирать то, которое имеет наибольшую энтропию, то есть является наиболее равномерным или «наименее предвзятым» [10]. Для каждого класса c функция вероятности имеет вид:

$$P(c | \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp \left(\sum_{i=1}^n \lambda_{i,c} f_{i,c}(\mathbf{x}) \right), \quad (3)$$

где:

- \mathbf{x} — вектор спектральных признаков пикселя;
- $\lambda_{i,c}$ — весовые коэффициенты для i -го признака класса c ;
- $f_{i,c}(\mathbf{x})$ — функции признаков (значения спектральных каналов);
- $Z(\mathbf{x})$ — нормирующая константа, обеспечивающая $\sum P(c | \mathbf{x}) = 1$.

Процесс обучения осуществляется с использованием реализации ее Classifier.amnhMaxent(), которая автоматически подбирает оптимальные параметры модели для каждого класса. После обучения для каждого пикселя вычисляются вероятности принадлежности ко всем классам, и окончательная классификация производится по правилу максимальной вероятности.

$$C(\mathbf{x}) = \arg \max_c P(c | \mathbf{x}) \quad (4)$$

Метод находит особенно успешное применение в задачах моделирования распределения видов растительности и классификации сложных ландшафтов с плавными переходами между классами. Его способность предоставлять вероятностные оценки делает его ценным инструментом для экологов и специалистов по дистанционному зондированию, позволяя не только классифицировать объекты, но и оценивать достоверность полученных результатов.

Метод классификации на основе ансамбля случайных деревьев (Random Forest)

Алгоритм Random Forest (Случайный лес) представляет собой ансамблевый метод машинного обучения, основанный на построении множества решающих деревьев. Каждое дерево обучается на случайной подвыборке объектов и признаков, что обеспечивает декорреляцию между отдельными деревьями и повышает общую обобщающую способность модели. Финальное предсказание формируется путем агрегирования прогнозов всех деревьев ансамбля [11].

Для классификационного вектора признаков x алгоритм вычисляет функцию классификации:

$$C(x) = \text{mode} \{T_1(x), T_2(x), \dots, T_K(x)\} \quad (5)$$

где:

- $T_k(x)$ — прогноз k -го дерева ансамбля;
- K (равен 100 по умолчанию) — количество деревьев в ансамбле;
- mode — модальная функция, возвращающая наиболее частый класс.

Вероятностная оценка для класса c вычисляется как:

$$P(c | x) = \frac{1}{K} \sum_{k=1}^K I(T_k(x) = c) \quad (6)$$

где I — индикаторная функция.

Для каждого пикселя вычисляется максимальная вероятность классификации

$$P_{\max}(x) = \max \{P(c_1 | x), P(c_2 | x), \dots, P(c_M | x)\} \quad (7)$$

где M — количество классов.

Метод успешно применяется для задач классификации земного покрова по данным дистанционного зондирования, особенно когда требуется не только точная классификация, но и оценка достоверности полученных результатов. Вероятностный вывод позволяет идентифицировать проблемные участки со смешанной спектральной сигнатурой.

Архитектура и функциональность приложения

Приложение построено на базе GEE JavaScript API и имеет модульную архитектуру, обеспечивающую логическое разделение функциональности и удобство использования. Пользовательский интерфейс состоит из трех ключевых компонентов (Рис. 1):

1. Панель управления (Control Panel) является основным центром для настройки всех параметров анализа. Она содержит виджеты для выбора временных интервалов, источников данных, алгоритмов классификации, определения количества и характеристик классов, а также для задания границ исследуемой территории.

2. Панель отчетов и консоль (Report Panel & Console) служит для вывода результатов анализа и служебной информации. Она включает в себя консоль для отображения сообщений о ходе выполнения операций и возможных ошибках, а также секцию для генерации и хранения подробных отчетов по каждому запуску классификации. Отчеты содержат информацию об использованных параметрах, статистику по площадям классов, метрики точности и время выполнения, что критически важно для обеспечения воспроизводимости исследований.

3. Основное окно карты (Main Map Window) служит для интерактивной визуализации исходных спутниковых данных, обучающих выборок, результатов классификации и вспомогательных карт. Интегрированные инструменты рисования позволяют пользователю в интерактивном режиме создавать и редактировать полигоны для обучающих выборок и границ исследуемой территории.

Такая структура позволяет пользователю последовательно проходить все этапы рабочего процесса в едином окне, от настройки параметров до анализа и интерпретации результатов.



Рисунок 1 - Архитектура созданного приложения
DOI: <https://doi.org/10.60797/IRJ.2026.163.72.1>

Данное приложение использует специальный ассет (`projects/ee-lopmaybay/assets/gee_catalog`), который представляет собой FeatureCollection, содержащий метаданные для почти 900 общедоступных наборов данных платформы. В интерфейсе приложения реализована функция поиска, которая позволяет пользователю вводить ключевые слова (например, «Landsat», «DEM», «precipitation») и легко интегрировать в свой анализ не только мультиспектральные данные, но и радарные снимки, цифровые модели рельефа, данные о климате, ночных огнях и многие другие, что открывает широкие возможности для мультисенсорного анализа и слияния данных.

Типичный рабочий процесс в приложении представляет собой последовательность логических шагов, направленных на итеративное улучшение результата классификации:

1. Настройка источников данных: Пользователь выбирает один или несколько наборов данных из стандартного списка или через поиск по каталогу GEE. Для каждой выбранной коллекции можно указать конкретные спектральные каналы и метод агрегации данных за год (например, `mosaic`, `median`, `mean`).

2. Определение временных параметров: Задаются два ключевых временных параметра: «Год для сбора выборки» (`Sampling Year`), на изображении за который будет производиться отрисовка и уточнение обучающих полигонов, и «Годы для классификации» (`Classification Years`), для которых будут созданы итоговые карты LULC.

3. Задание исследуемой территории (ROI): Границы анализируемой области могут быть определены тремя способами:

- (а) поиск по названию административной единицы с использованием базы данных FAO/GAUL;
- (б) интерактивная отрисовка полигона или прямоугольника непосредственно на карте;
- (в) загрузка границ из собственного `shape`-файла, предварительно загруженного в GEE Assets.

4. Определение классов LULC: Пользователь указывает количество классов, задает для каждого из них уникальное имя (например, «лес», «вода», «город») и выбирает цвет для визуализации на карте. Набор классов также можно загрузить из собственного `shape`-файла, предварительно загруженного в GEE Assets.

5. Первичный сбор обучающих выборок: Для каждого класса создается начальная обучающая выборка. Это можно сделать путем интерактивной отрисовки полигонов на изображении за «Год для сбора выборки» или путем загрузки готовых выборок из GEE Asset. Приложение также предлагает возможность загрузить предустановленные наборы выборок для региона Хайфон, что удобно для тестирования и обучения.

6. Генерация вспомогательных карт (ключевой этап): После создания первичной выборки пользователь запускает один или несколько алгоритмов поддержки выборки (SAM, Similarity, Random Forest, Maxent). Инструмент вычисляет для каждого пикселя на изображении оценку, отражающую его соответствие существующим обучающим классам, а затем создает итоговую карту «Максимального сходства» (или «Максимальной вероятности»). Низкие значения на этой карте указывают на пиксели, которые спектрально не очень похожи ни на один из существующих классов. Эти «темные области» являются основными кандидатами для дополнительного сбора образцов, поскольку они могут представлять собой либо недостаточно охваченные вариации уже существующих классов, либо совершенно новые, неучтенные классы. Лучше всего на данном этапе используется набор данных AlphaEarth Foundations (AEF). AEF представляет собой фундаментальную модель, обученную на петабайтах мультимодальных данных дистанционного зондирования. Модель ассимилирует информацию из разнообразных источников, включая оптические (Landsat, Sentinel-2), радарные (Sentinel-1), высотные (GLO-30 DEM) и климатические (ERA5-Land) данные, для создания единого, семантически богатого представления поверхности Земли. Основное преимущество использования AEF заключается в том, что модель абстрагирует сложный и вычислительно затратный процесс ручного инжиниринга признаков (например, расчет многочисленных спектральных индексов, таких как NDVI, NDBI, текстурных признаков и т.д.). Вместо этого AEF предоставляет готовый, более информативный и дискриминационный набор признаков, который превосходит традиционные подходы, особенно в условиях ограниченного количества обучающих данных.

7. Итеративное уточнение выборки: На основе анализа карты «Максимального сходства» пользователь добавляет новые точки или корректирует существующие, итеративно улучшая репрезентативность обучающей выборки. На рис. 2. и рис. 3 представляется изменение значения максимального сходства пикселей при изменении количества обучающих точек, при этом алгоритм «случайный лес» (Random Forest) является более чутким к изменениям.

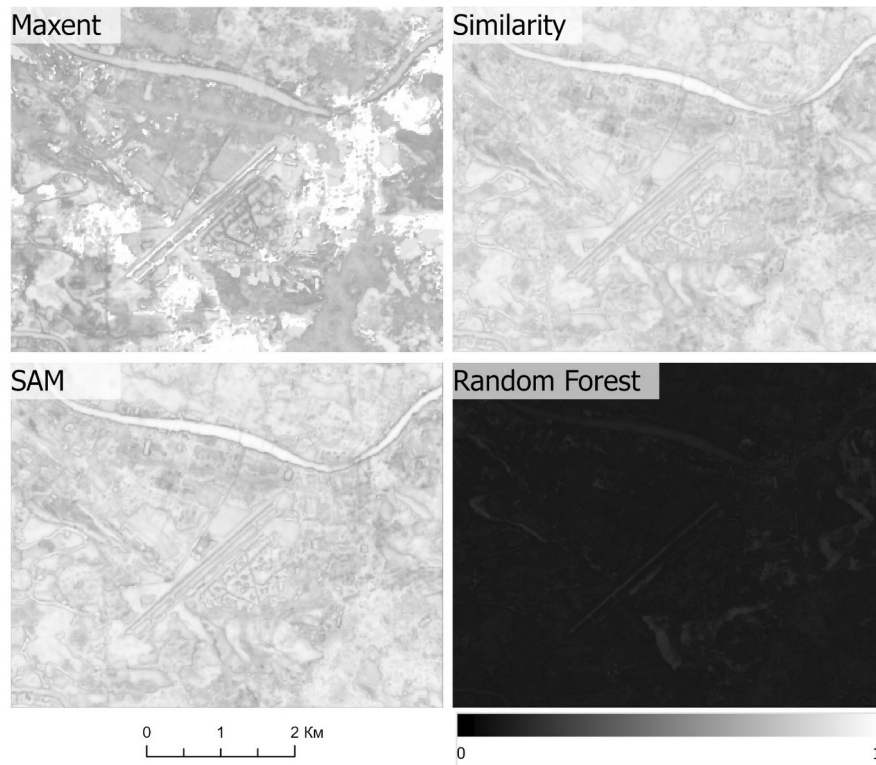


Рисунок 2 - Карта индекса «Максимальной вероятности» 4 алгоритмов для фрагмента снимки AlphaEarth Foundations в случае, когда каждый класс имеет от 1 до 5 обучающих точек
DOI: <https://doi.org/10.60797/IRJ.2026.163.72.2>

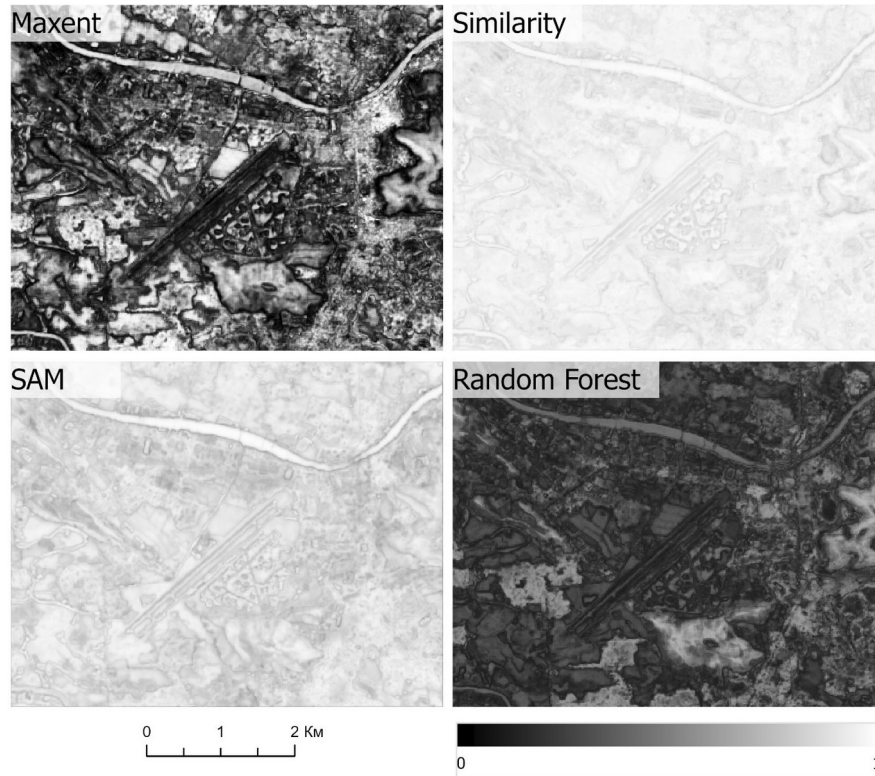


Рисунок 3 - Карта индекса «Максимальной вероятности» 4 алгоритмов для фрагмента снимки AlphaEarth Foundations в случае, когда каждый класс имеет от 30 до 60 обучающих точек
DOI: <https://doi.org/10.60797/IRJ.2026.163.72.3>

8. Финальная классификация: После того как обучающая выборка признана удовлетворительной, пользователь выбирает основной алгоритм классификации (например, Random Forest) и запускает процесс.

9. Анализ результатов: Приложение отображает на карте итоговые LULC карты для всех указанных «Годов для классификации» и алгоритмов. В панели отчетов генерируется подробная информация, включающая статистику по площадям каждого класса и, для года сбора выборки, метрики точности (общая точность, матрица ошибок), рассчитанные на основе самих обучающих данных (рис. 4). После этого, пользователь может сохранить все исходные данные, результаты и отчеты для дальнейшей работы в других ГИС. На рис. 5 демонстрирует разные результаты классификации разными алгоритмами, что помогает сравнить эффективность и точность классификаций.

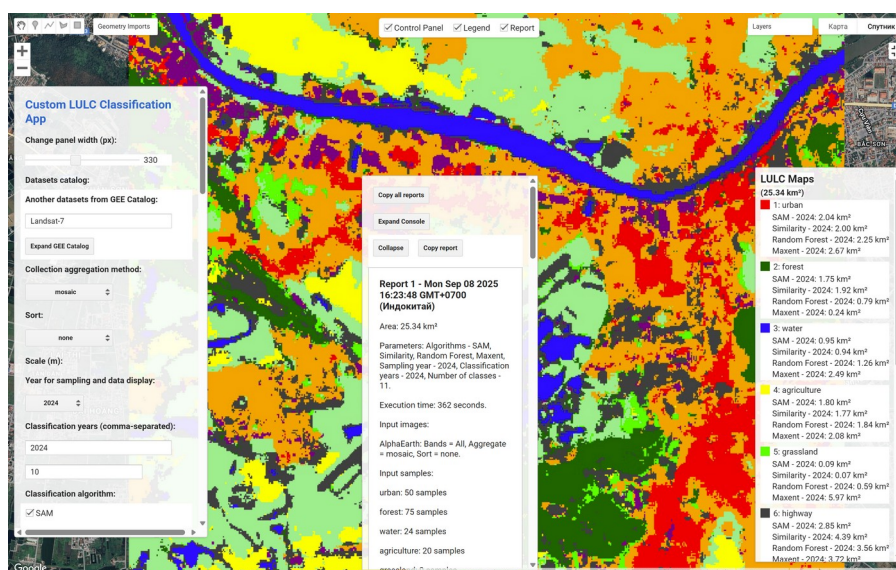


Рисунок 4 - Интерфейс приложения после процесса классификации и анализа результатов
DOI: <https://doi.org/10.60797/IRJ.2026.163.72.4>

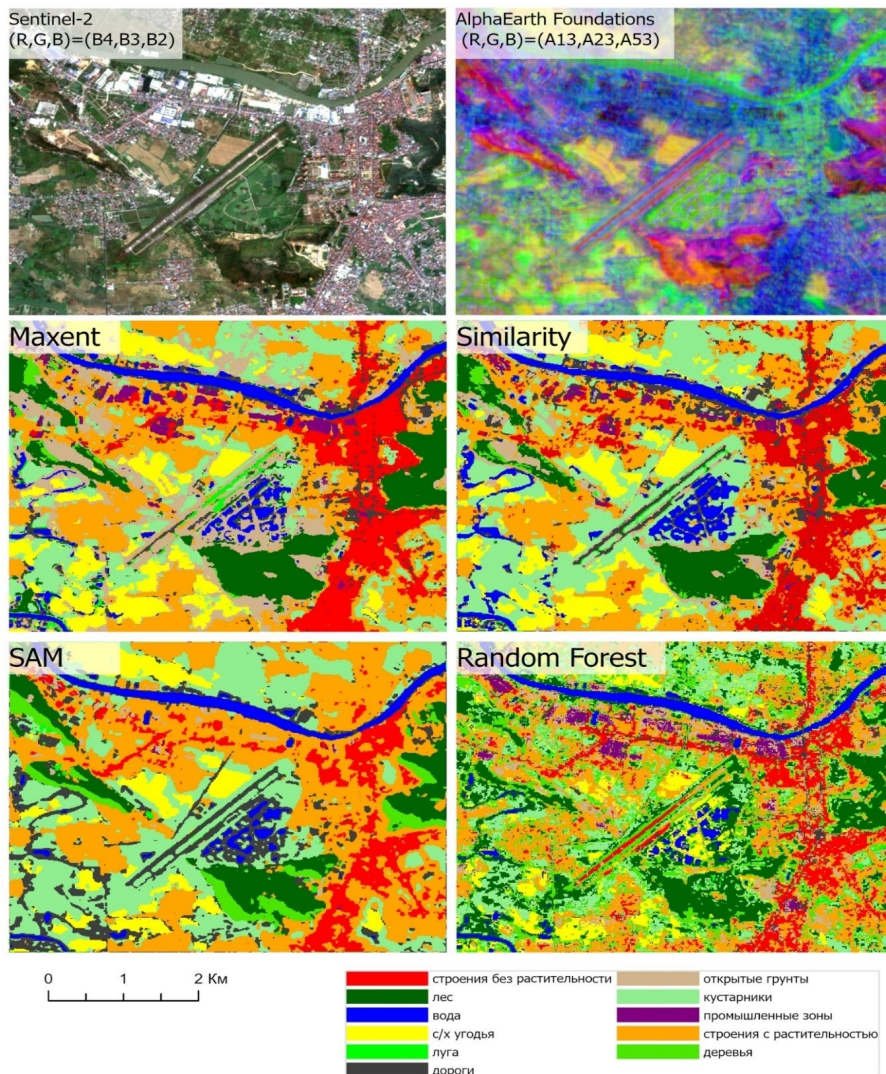


Рисунок 5 - Фрагменты снимков и результата классификации 4 алгоритмами для снимки AlphaEarth Foundations
DOI: <https://doi.org/10.60797/IRJ.2026.163.72.5>

Обсуждение

Несмотря на огромные вычислительные возможности, платформа Google Earth Engine имеет ряд встроенных ограничений, направленных на справедливое распределение ресурсов между пользователями и обеспечение стабильности системы. При работе с большими территориями, данными высокого разрешения или сложными алгоритмами пользователи приложения неизбежно столкнутся с ошибками, такими как «User memory limit exceeded» (Превышен лимит памяти пользователя), «Computation timed out» (Время вычисления истекло) или «Too many concurrent aggregations» (Слишком много одновременных агрегаций) [12].

Важно понимать, что эти ошибки не являются сбоями в работе приложения, а представляют собой фундаментальные ограничения интерактивного режима GEE, который имеет строгие квоты на использование памяти и времени процессора для каждого запроса [12]. Для успешной работы с приложением в реальных исследовательских задачах необходимо применять стратегии, которые позволяют обходить эти ограничения. На основе анализа документации и лучших практик можно сформулировать следующие рекомендации:

1. Масштабирование области интереса (ROI): перед запуском анализа на всей исследуемой территории рекомендуется проводить интерактивную настройку параметров, сбор и уточнение выборки на небольшом, но репрезентативном участке. Это значительно снижает нагрузку на серверы GEE и позволяет быстро получать обратную связь.

2. Управление масштабом (scale): параметр scale определяет пространственное разрешение, с которым выполняются вычисления. Для предварительного анализа или при работе с очень большими территориями можно временно увеличить значение scale (например, до 30 или 100 метров), чтобы уменьшить количество обрабатываемых пикселей и избежать превышения лимитов.

3. Оптимизация геометрии: следует избегать использования очень сложных полигонов с тысячами вершин в качестве ROI или обучающих выборок при работе в интерактивном режиме, так как это может привести к превышению лимита на размер запроса.

Представленное приложение является открытой, гибкой и расширяемой платформой. Дальнейшее развитие может идти по нескольким направлениям:

– Интеграция методов активного обучения (Active Learning): текущий процесс уточнения выборки является интерактивным, но ведомым человеком. Следующим шагом может стать интеграция алгоритмов активного обучения, которые бы автоматически предлагали пользователю наиболее информативные пиксели для включения в обучающую выборку, тем самым еще больше автоматизируя и оптимизируя этот процесс [4].

– Использование семантических признаков (Foundational Models): приложение уже включает возможность использования набора данных AlphaEarth, который является примером спутниковых эмбедингов (satellite embeddings). Дальнейшее развитие может включать более глубокую интеграцию с так называемыми «фундаментальными моделями» (foundational models), которые предварительно обучены на огромных массивах данных и способны извлекать семантически богатые признаки из изображений.

– Расширение для прикладных задач: архитектура приложения может быть адаптирована для решения специфических задач. Например, для оперативной оценки ущерба после стихийных бедствий, где рабочий процесс, включающий быструю классификацию на основе ограниченных данных, является критически важным.

Заключение

В данной работе было представлено и апробировано веб-приложение, реализованное на платформе Google Earth Engine и функционирующее как виртуальная исследовательская среда для картографирования земельного и растительного покрова. Приложение доступно по ссылке <https://ee-lorpmaybay.projects.earthengine.app/view/lulc>. Приложение успешно сочетает в себе доступ к обширным архивам спутниковых данных, гибкость в выборе методов анализа и интуитивно понятный пользовательский интерфейс, что значительно упрощает и ускоряет процесс LULC-анализа.

Ключевым научным вкладом работы является разработка и демонстрация эффективности нового рабочего процесса, основанного на алгоритмической поддержке сбора обучающих выборок. Инновационное применение алгоритмов SAM, Similarity, Random Forest и Maxent не только в качестве конечных классификаторов, но и инструментов для генерации вспомогательных карт сходства и вероятности, позволяет реализовать эффективный полуавтоматический механизм. Как показал выше, этот итеративный подход к уточнению обучающих данных позволяет значительно повысить качество и репрезентативность выборки, что напрямую ведет к существенному росту точности итоговых карт LULC.

Практическая ценность приложения заключается в его гибкости, обеспечиваемой механизмом динамического подключения данных из каталога GEE, и в предоставлении пользователям четких и понятных стратегий для работы с вычислительными ограничениями платформы. Сочетание вычислительной мощи облачных технологий с интеллектуальными, ориентированными на пользователя рабочими процессами делает данное приложение ценным инструментом для исследователей, преподавателей и специалистов-практиков. Разработка подобных VRE вносит значительный вклад в развитие принципов «Открытой науки», повышая доступность, прозрачность и воспроизводимость исследований в области наук о Земле.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Ковальчук А.В. Earth Engine от Google — уникальная платформа для анализа больших данных Д33 / А.В. Ковальчук [и др.] // *Habr*. — 2020. — URL: <https://habr.com/ru/articles/500020/> (дата обращения: 01.10.2024).
2. Liu Y. Land use and land cover classification and terrestrial ecosystem carbon storage changes in Vietnam based on Sentinel images / Y. Liu, K. Yang, Z. Peng [et al.] // *Scientific Reports*. — 2025. — Vol. 15, № 1. — P. 22114. — DOI: 10.1038/s41598-025-04765-z
3. Collombin M. The Challenges of Reproducibility for Research Based on Geodata Web Services / M. Collombin, O. Ertz, G. Giuliani [et al.] // *Preprints*. — 2023. — DOI: 10.20944/preprints202312.2316.v1
4. Berger K. A Survey of Active Learning for Quantifying Vegetation Traits from Terrestrial Earth Observation Data / K. Berger, J.P. Rivera Caicedo, L. Martino [et al.] // *Remote Sensing*. — 2021. — Vol. 13, № 2. — P. 287. — DOI: 10.3390/rs13020287
5. Gesing S. Science Gateways and AI/ML: How Can Gateway Concepts and Solutions Meet the Needs in Data Science? / S. Gesing, M. Pierce, S. Marru [et al.] // *IntechOpen*. — 2023. — DOI: 10.5772/intechopen.110144
6. Uzhinskiy A. Google Earth Engine and machine learning for Earth monitoring / A. Uzhinskiy // *CEUR Workshop Proceedings*. — 2022. — Vol. 3429. — P. 21–34. — URL: <http://ceur-ws.org/Vol-3429/paper2.pdf> (accessed: 01.10.2024).
7. Spectral Angle Mapper Algorithm for Remote Sensing Image Classification // *International Journal of Innovative Science, Engineering & Technology (IJSET)*. — 2014. — Vol. 1, № 4. — P. 192–196. — URL: https://www.ijset.com/v1s4/IJSET_V1_I4_27.pdf (accessed: 01.10.2024).

8. Kruse F.A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data / F.A. Kruse, A.B. Lefkoff, J.W. Boardman [et al.] // *Remote Sensing of Environment*. — 1993. — Vol. 44, № 2-3. — P. 145–163. — DOI: 10.1016/0034-4257(93)90013-N
9. How Presence-only Prediction (MaxEnt) works // *ArcGIS Pro Documentation*. — URL: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-presence-only-prediction-works.htm> (accessed: 01.10.2024).
10. Morales N.S. Land-Cover Classification Using MaxEnt: Can We Trust in Model Quality Metrics for Estimating Classification Accuracy? / N.S. Morales, I.C. Fernández // *Entropy*. — 2020. — Vol. 22, № 3. — P. 342. — DOI: 10.3390/e22030342
11. Breiman L. Random Forests / L. Breiman // *Machine Learning*. — 2001. — Vol. 45, № 1. — P. 5–32. — DOI: 10.1023/A:1010933404324
12. Earth Engine quotas // *Google for Developers*. — URL: <https://developers.google.com/earth-engine/guides/usage> (accessed: 01.10.2024).

Список литературы на английском языке / References in English

1. Kovalchuk A.V. Earth Engine or Google — unikal'naya platforma dlya analiza bol'shikh dannykh DZZ [Google Earth Engine — a unique platform for analysis of big Earth observation data] / A.V. Kovalchuk [et al.] // *Habr*. — 2020. — URL: <https://habr.com/ru/articles/500020/> (accessed: 01.10.2024). [in Russian]
2. Liu Y. Land use and land cover classification and terrestrial ecosystem carbon storage changes in Vietnam based on Sentinel images / Y. Liu, K. Yang, Z. Peng [et al.] // *Scientific Reports*. — 2025. — Vol. 15, № 1. — P. 22114. — DOI: 10.1038/s41598-025-04765-z
3. Collombin M. The Challenges of Reproducibility for Research Based on Geodata Web Services / M. Collombin, O. Ertz, G. Giuliani [et al.] // *Preprints*. — 2023. — DOI: 10.20944/preprints202312.2316.v1
4. Berger K. A Survey of Active Learning for Quantifying Vegetation Traits from Terrestrial Earth Observation Data / K. Berger, J.P. Rivera Caicedo, L. Martino [et al.] // *Remote Sensing*. — 2021. — Vol. 13, № 2. — P. 287. — DOI: 10.3390/rs13020287
5. Gesing S. Science Gateways and AI/ML: How Can Gateway Concepts and Solutions Meet the Needs in Data Science? / S. Gesing, M. Pierce, S. Marru [et al.] // *IntechOpen*. — 2023. — DOI: 10.5772/intechopen.110144
6. Uzhinskiy A. Google Earth Engine and machine learning for Earth monitoring / A. Uzhinskiy // *CEUR Workshop Proceedings*. — 2022. — Vol. 3429. — P. 21–34. — URL: <http://ceur-ws.org/Vol-3429/paper2.pdf> (accessed: 01.10.2024).
7. Spectral Angle Mapper Algorithm for Remote Sensing Image Classification // *International Journal of Innovative Science, Engineering & Technology (IJSET)*. — 2014. — Vol. 1, № 4. — P. 192–196. — URL: https://www.ijset.com/v1s4/IJSET_V1_I4_27.pdf (accessed: 01.10.2024).
8. Kruse F.A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data / F.A. Kruse, A.B. Lefkoff, J.W. Boardman [et al.] // *Remote Sensing of Environment*. — 1993. — Vol. 44, № 2-3. — P. 145–163. — DOI: 10.1016/0034-4257(93)90013-N
9. How Presence-only Prediction (MaxEnt) works // *ArcGIS Pro Documentation*. — URL: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-presence-only-prediction-works.htm> (accessed: 01.10.2024).
10. Morales N.S. Land-Cover Classification Using MaxEnt: Can We Trust in Model Quality Metrics for Estimating Classification Accuracy? / N.S. Morales, I.C. Fernández // *Entropy*. — 2020. — Vol. 22, № 3. — P. 342. — DOI: 10.3390/e22030342
11. Breiman L. Random Forests / L. Breiman // *Machine Learning*. — 2001. — Vol. 45, № 1. — P. 5–32. — DOI: 10.1023/A:1010933404324
12. Earth Engine quotas // *Google for Developers*. — URL: <https://developers.google.com/earth-engine/guides/usage> (accessed: 01.10.2024).