

УДК: 528.7+911.2

DOI: 10.35595/2414-9179-2025-2-31-83-99

Н. В. Максимович<sup>1</sup>, А. А. Тарасов<sup>2</sup>, А. П. Кораблёв<sup>3</sup>, Н. А. Позднякова<sup>4</sup>, Т. А. Андреева<sup>5</sup>,  
О. В. Артемьева<sup>6</sup>

## КАРТОГРАФИРОВАНИЕ РАСТИТЕЛЬНОСТИ КАЛЬДЕРЫ УЗОН С ИСПОЛЬЗОВАНИЕМ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ И МАШИННОГО ОБУЧЕНИЯ

### АННОТАЦИЯ

Целью данной работы было создание картографической модели растительности части кальдеры Узон. Для этого в работе была использована генерализованная карта растительности кальдеры, составленная А. О. Пестеровым, а также данные геоботанических описаний, предоставленные сотрудниками Ботанического института РАН в качестве вспомогательных материалов для создания обучающей выборки для машинного обучения. В работе также использованы данные о высоте растительного покрова, наземной биомассы и данные нескольких каналов Sentinel-2 и Landsat-9, из которых был создан композитный растр. Помимо каналов 4–8 Sentinel-2 и 10 канала Landsat-9, в классифицируемый растр вошли данные ArcticDEM. Создание обучающей выборки производилось на основе вспомогательных данных, создание тестовой выборки — на основе экспертной оценки на основе ДДЗЗ высокого разрешения и данных геоботанических описаний. В качестве используемых методов классификации были выбраны K-means и Random Forest. Для первого применен метод локтя для оценки выраженности и отделимости классов друг от друга в классифицируемом растре. Данный способ показал несостоятельность метода неконтролируемой классификации и необходимость использования методов с обучением. Для второго применялась оценка точности на тестовых данных с заранее выявленными оптимальными параметрами обученной модели; также была составлена матрица ошибок каждого класса и посчитаны метрики качества классификации в Python 3.0. Так, общая точность (ассигасу) модели составила 90 %. Были рассчитаны статистические характеристики модели и выявлены особенности ее работы с получившимися классами растительности.

**КЛЮЧЕВЫЕ СЛОВА:** дистанционное зондирование Земли, машинное обучение, растительность, метрики точности, матрица ошибок

<sup>1</sup> Санкт-Петербургский государственный университет, Институт наук о Земле, Университетская наб., д. 7–9, Санкт-Петербург, Россия, 199034, *e-mail*: [nekit.maksimovich@mail.ru](mailto:nekit.maksimovich@mail.ru)

<sup>2</sup> Санкт-Петербургский государственный университет, Институт наук о Земле, Университетская наб., д. 7–9, Санкт-Петербург, Россия, 199034, *e-mail*: [artrar.90@mail.ru](mailto:artrar.90@mail.ru)

<sup>3</sup> Ботанический институт им. В. Л. Комарова Российской академии наук, ул. Профессора Попова, д. 2, Санкт-Петербург, Россия, 197376, *e-mail*: [akorablev@binran.ru](mailto:akorablev@binran.ru)

<sup>4</sup> Санкт-Петербургский государственный университет, Институт наук о Земле, Университетская наб., д. 7–9, Санкт-Петербург, Россия, 199034, *e-mail*: [n.pozdnyakova@spbu.ru](mailto:n.pozdnyakova@spbu.ru)

<sup>5</sup> Санкт-Петербургский государственный университет, Институт наук о Земле, Университетская наб., д. 7–9, Санкт-Петербург, Россия, 199034, *e-mail*: [t.andreeva@spbu.ru](mailto:t.andreeva@spbu.ru)

<sup>6</sup> Санкт-Петербургский государственный университет, Институт наук о Земле, Университетская наб., д. 7–9, Санкт-Петербург, Россия, 199034, *e-mail*: [o.artemeva@spbu.ru](mailto:o.artemeva@spbu.ru)

**Nikita V. Maksimovich<sup>1</sup>, Artem A. Tarasov<sup>2</sup>, Anton P. Korablev<sup>3</sup>, Natalia A. Pozdnyakova<sup>4</sup>, Tatyana A. Andreeva<sup>5</sup>, Olga V. Artemeva<sup>6</sup>**

## **MAPPING UZON CALDERA VEGETATION USING REMOTE SENSING DATA AND MACHINE LEARNING**

### **ABSTRACT**

Remote sensing data (RSD) and the machine learning methods used with them are currently used in mapping geographic processes and objects of the earth's surface associated with the functioning of ecosystems and landscapes. RSD is often used to map one of the landscape components — vegetation, often representing the main source of information about it. This also applies to such hard-to-reach areas as the Uzon caldera. Thus, the purpose of this work was to create a cartographic model of the vegetation of a part of the Uzon caldera. For this purpose, the work used a generalized map of the caldera vegetation compiled by A. O. Pesterov, data from geobotanical descriptions provided by employees of the Botanical Institute of the Russian Academy of Sciences as auxiliary materials for creating a training sample for machine learning. The work also used data on the height of the vegetation cover, aboveground biomass, as well as several Sentinel-2 and Landsat-9 channels, from which a composite raster was created. In addition to channels 4–8 of Sentinel-2 and channel 10 of Landsat-9, the classified raster included ArcticDEM data. The training sample was created based on auxiliary data; the test sample was created by expert assessment based on high-resolution remote sensing data and geobotanical description data. K-means and Random Forest were selected as the classification methods used. For the first, the elbow method was used to assess the expression and separability of classes from each other in the classified raster. This method showed the inconsistency of the unsupervised classification method and the need to use methods with training. For the second, an accuracy assessment was used on test data with pre-determined optimal parameters of the trained model; an error matrix for each class was also compiled and classification quality metrics were calculated in Python 3.0. Thus, the overall accuracy of the model was 90 %. Statistical characteristics of the model were calculated and the features of its operation with the resulting vegetation classes were identified.

**KEYWORDS:** Earth remote sensing, machine learning, vegetation, accuracy metrics, confusion matrix

### **ВВЕДЕНИЕ**

Территория Кроноцкого заповедника расположена на Восточном вулканическом поясе Камчатки. Для нее характерны горный рельеф и высотная поясность, а также различные структуры растительного покрова, свойственные вулканическим районам

---

<sup>1</sup> St. Petersburg State University, Institute of Earth Sciences, 7–9, Universitetskaya emb., St. Petersburg, 199034, Russia, *e-mail*: [nekit.maksimovich@mail.ru](mailto:nekit.maksimovich@mail.ru)

<sup>2</sup> St. Petersburg State University, Institute of Earth Sciences, 7–9, Universitetskaya emb., St. Petersburg, 199034, Russia, *e-mail*: [artrar.90@mail.ru](mailto:artrar.90@mail.ru)

<sup>3</sup> Komarov Botanical Institute of the Russian Academy of Sciences, 2, Professora Popova str., St. Petersburg, 197376, Russia, *e-mail*: [akorablev@binran.ru](mailto:akorablev@binran.ru)

<sup>4</sup> St. Petersburg State University, Institute of Earth Sciences, 7–9, Universitetskaya emb., St. Petersburg, 199034, Russia, *e-mail*: [n.pozdnyakova@spbu.ru](mailto:n.pozdnyakova@spbu.ru)

<sup>5</sup> St. Petersburg State University, Institute of Earth Sciences, 7–9, Universitetskaya emb., St. Petersburg, 199034, Russia, *e-mail*: [t.andreeva@spbu.ru](mailto:t.andreeva@spbu.ru)

<sup>6</sup> St. Petersburg State University, Institute of Earth Sciences, 7–9, Universitetskaya emb., St. Petersburg, 199034, Russia, *e-mail*: [o.artemeva@spbu.ru](mailto:o.artemeva@spbu.ru)

[*Нешатаева и др., 2021*], что говорит о сложном распределении растительности на территории исследования. Высокая мозаичность растительного покрова заповедника также объяснима наличием на его территории термальных ландшафтов. Задача классификации и картографирования растительности в кальдере Узон является непростой и интересной с точки зрения сложности создания обучающей выборки и разнообразия на территории исследования растительных сообществ — от термальной растительности в центре кальдеры до широко распространившихся на склонах бортов кальдеры ольховых и кедровых стлаников [*Нешатаева, Пестеров, 2012*]. Помимо полевых данных важными материалами при картографировании растительности выступают данные дистанционного зондирования Земли.

На данный момент ДДЗЗ являются основным источником получения информации дистанционно о растительном покрове, особенно в труднодоступных районах. Основной платформой, предоставляющей в короткие сроки данные дистанционного зондирования, является платформа Google Earth Engine (GEE)<sup>1</sup>, в которой реализован Code Editor. С его помощью существует возможность быстрого получения нужных данных дистанционного зондирования, а также их предобработки, что сильно облегчает работу с космическими снимками [*Полевицкова, Кириллов, 2023*]. Так, проанализировав спектральные характеристики космоснимков и применив различные методы классификации растительности с помощью инструментов ГИС, можно составить карты растительности и оценить их схожесть с данными, полученными полевым методом, а также определить особенности дешифрирования растительности, проанализировать и оценить точность классификации при создании тестовой выборки.

Выделение различных типов растительности при картографировании территории возможно методами неконтролируемой и контролируемой классификации. Неконтролируемая классификация используется, как правило, на первых этапах исследований и может определить растительность района с низкой точностью, выделяя лишь те классы, которые четко отделяются друг от друга спектральными и иными характеристиками. Однако с помощью методов неконтролируемой классификации можно выявить допустимое количество классов, которое поддается классификации на основе раstra. Таким образом, при наличии достаточного количества полевых данных и данных дистанционного зондирования уместнее пользоваться методами контролируемой классификации, которые позволяют создать модели, отвечающие различным требованиям, предъявляемым к выделению разных типов растительности.

Точность автоматического дешифрирования растительности лесов и других природных объектов различными методами классификации в целом остается неопределенной [*Лузаченко и др., 2008; Малышева, 2012*]. Неопределенность может быть связана с большим разнообразием растительности, физико-географических условий территорий, климатическими факторами и также зависит от поставленных задач исследования. Определить точность полученных результатов проведенного исследования позволяют данные, полученные на местности полевым методом, и составленные ранее карты. Однако качество составленных обучающих (тренировочных) и тестовых выборок также зачастую является субъективным, как и сам процесс создания выборок.

Предоставленные Ботаническим институтом РАН имени В. Л. Комарова данные наземных исследований растительности позволят оценить точность, достоверность проведенного дешифрирования. Автоматическое дешифрирование методом контролируемой классификации позволит представить новые модели пространственного распределения растительности на рассматриваемой территории и сравнить их с результатами, получен-

---

<sup>1</sup> Google Earth Engine. Электронный ресурс: <https://earthengine.google.com/platform/> (дата обращения 20.03.2025)

ными полевым методом. Использование внешних данных, характеризующих характер растительности, также может помочь при анализе растительности, ландшафтов, экосистем и интерпретации результатов. Таким образом, целью исследования является создание геоботанической карты центральной части кальдеры Узон с использованием методов машинного обучения и данных дистанционного зондирования.

## МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Полевые данные геоботанических описаний и составленная генерализованная карта А. О. Пестерова [2017] являются основными материалами, послужившими для создания обучающей выборки и оценки полученной классификации.

В качестве территории исследования была выбрана часть кальдеры Узон (выделена красным на карте А. О. Пестерова, рис. 1), имеющая высокую плотность геоботанических описаний, выполненных сотрудниками БИН РАН, которые впоследствии были привязаны и разделены на классы.

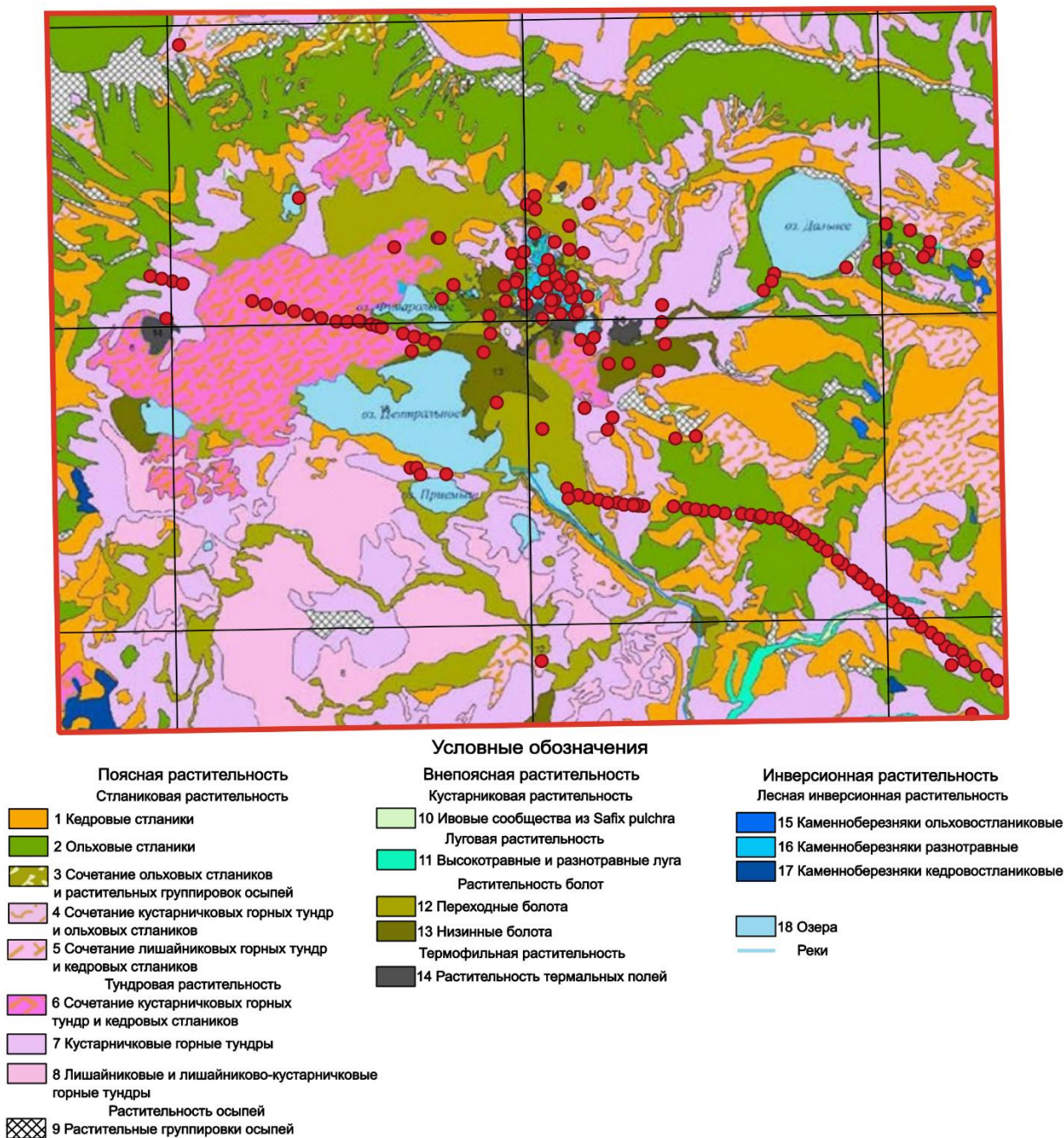
В работе были использованы: QGIS 3.28, ArcGIS PRO, платформа Google Earth Engine и Python 3.12. Необходимость применения Python для классификации и ее оценки вызвана большой гибкостью данного инструмента относительно оценки классификации и настраиваемых к созданию модели параметров.

В качестве классифицируемого изображения составлен композитный многоканальный растр, состоящий из 5 каналов Sentinel-2, 1 канала Landsat-9 (все — медианные композиты за август 2024 г.) и модели высот ArcticDEM, а также растров надземной биомассы [Boitard et al., 2025] и высоты растительности [Tolan et al., 2024], взятых из платформы GEE. Последние, разрешением в 1 м, были сгенерированы путем анализа снимков Махаг с 2017 по 2020 г. различными самообучающимися моделями ИИ. Кроме того, при формировании этих растров была проведена постобработка с использованием сверточной нейронной сети, обученной на наблюдениях GEDI. Авторы заявляют, что средняя абсолютная ошибка (MAE) обозначенных данных составляет 2,8 м, а средняя ошибка (ME) — 0,6 м.

Шаги составления композитного (многоканального) растра:

1. На платформе GEE составлен медианный композит снимков Sentinel-2 SR с помощью команды `.median()` и далее производилось получение и скачивание на облачный диск композита с помощью команды `Export.Image.toDrive()`. Условиями подбора космоснимков были: низкая облачность (менее 10 %) и полное или частичное покрытие территории исследования. Использован растр с пятью каналами: B4, B5, B6, B7, B8.
2. Получение 10-го канала Landsat-9 (TIRS 1), который косвенно характеризует температуру поверхности Земли. Данный канал использовался для повышения точности классификации термальной растительности, распространенной в кальдере Узон.
3. Для повышения точности последующей классификации применены данные цифровой модели рельефа ArcticDEM (имеющие пространственное разрешение в 2 м/пикс), взятые с сайта университета Миннесоты<sup>1</sup>, а также данные о высоте растительности и ее надземной биомассе, полученные с помощью платформы GEE.
4. Приведение всех полученных растров к одному пространственному разрешению в 10 м и к одной проекции зоны (UTM zone 57N) с помощью инструмента «Экспорт» в QGIS. При этом был использован для всех растров был использован метод ближайшего соседа.

<sup>1</sup> ArcticDEM — Polar Geospatial Center. Электронный ресурс: <https://www.pgc.umn.edu/data/arcticdem/> (дата обращения 08.07.2025)



*Рис. 1. Карта А. О. Пестерова и точки геоботанических описаний (красным), привязанные к территории исследования [Пестеров, 2017]*

*Fig. 1. Map by A. O. Pesterov and points of geobotanical descriptions (red) linked to the study area [Pesterov, 2017]*

Использование многоканальных растров, состоящих из разных данных, в процессе машинного обучения приводит к повышению точности классификации растительности и других объектов (точность полученной модели при проведении экспериментов повысилась на 10–15 %), что подтверждается во многих исследованиях, в которых методы машинного обучения применялись для классификации [Котлов и др., 2024; Шашков и др., 2024]. Повышение точности классификации (в первую очередь с обучением) происходит за счет выявления новых характеристик или закономерностей, разграничивающих классы в

гиперпространстве признаков. Так, для растительности важной характеристикой является один из ее ландшафтных признаков — характер рельефа (высота), получаемый с цифровой модели рельефа (ЦМР). Еще одним важным признаком термальных экосистем, находящихся на территории исследования, является температура поверхности, интерпретируемая с помощью 10 канала спутника Landsat-9. Главным отличительным признаком березняков является их высота (более 5–7 м), что было выявлено с помощью растра, содержащего высоту растительности.

Следующий шаг исследования заключался в создании обучающей (тренировочной) и тестовой выборок. Процесс создания обучающей выборки в QGIS включает использование привязанной карты А. О. Пестерова, полевых геоботанических описаний, а также космических снимков ESRI в качестве подложки. Всего было создано 608 точек, включающих в себя 8 классов растительности (ольховый и кедровый стланики, кустарничковые и лишайниковые тундры, растительность на осыпях, болота, термальная растительность, березняки) и 2 класса поверхности — водные объекты и снег. Основными принципами создания обучающей выборки были равномерное распределение точек по территории исследования и стремление к сбалансированности классов относительно количества создаваемых точек. Однако последнее не всегда удавалось соблюдать из-за изначальной нераспространенности или редкости некоторых видов (формаций) растительности на территории.

Тестовая выборка создавалась также на основе вышеперечисленных вспомогательных данных, но уже в количестве 425 точек. Основным принципом создания тестовой выборки создание буфера в 15 м вокруг каждой точки. Таким образом, как предполагается авторами, можно достичь максимальной объективности оценки результатов классификации, т. к. это «стандартизирует» выставленные точки тестовой выборки, основанные на снимках высокого разрешения и геоботанических описаниях (размер которых составляет в среднем 10×10 или 20×20 м на местности) относительно классифицируемого растра, пространственное разрешение которого — 10 м/пикс. Вследствие того, что пиксель по сути является собирательным значением спектральных характеристик той или иной территории, из-за данного подхода объективность оценки точности возрастет за счет учета 1–2 соседних пикселей, которые могут быть отнесены к верному классу. Впоследствии производился учет всех классифицированных пикселей, входящих в каждый созданный буфер, с помощью метода *within()* в Python.

Основной частью исследования было создание модели машинного обучения, основанной на методе Random Forest. Однако перед этим использовался метод локтя для определения оптимального делимого количества классов и их делимости в целом. Так, для понятия делимости классов и количества кластеров, в рамках которых возможна наиболее точная классификация, был произведен расчет внутрикластерной суммы квадратов с помощью библиотеки Scikit-learn в Python.

Формула (1) расчета внутрикластерной суммы квадратов (WCSS):

$$WCSS = \sum_{i=1}^n \sum_{j=1}^k ||x_i - \mu_j|| \quad (1),$$

где  $x_i$  — точка данных,  
 $\mu_j$  — центр  $j$ -го кластера,  
 $n$  — количество точек данных,  
 $k$  — количество кластеров.

Для интерпретации полученных вычислений был построен график внутрикластерных сумм квадратов относительно каждого количества кластеров (от 1 до 20), с помощью которого была проанализирована способность кластеров «разделяться» друг от друга.

Самым важным и продолжительным по времени шагом в исследовании было обучение модели на основе алгоритма Random Forest. Данный метод является, пожалуй, самым популярным среди методов машинного обучения, используемых с данными дистанционного зондирования. С его помощью возможно классифицировать как простые данные, предполагающие бинарную классификацию, так и сложные многомерные данные, предполагающие создание большого количества классов. В своих исследованиях [Thanh Noi, Kappas, 2017; Mohammadpour, 2022] авторы описывают метод Random Forest как один из лучших ансамблевых методов машинного обучения, который может быть применен для классификации сложных данных, основанных на дистанционном зондировании Земли. В исследовании [Boateng et al., 2020] описано, что Random Forest нечувствителен к помехам и к несбалансированной обучающей выборке в отличие от некоторых других методов машинного обучения, что также подходит для достижения цели нашего исследования.

Суть данного метода — разделение выборки на подвыборки по такому критерию, который показывает наибольший прирост информации, позволяющей разделять классы. Когда вся выборка разделена на подвыборки с максимальным разделением классов друг от друга, им присваиваются значения классов. Такой процесс повторяется в нескольких деревьях в несколько итераций. Чем глубже каждое дерево, тем больше таких итераций произведет алгоритм. Кроме того, стоит обратить внимание на то, что таких деревьев в данном методе может быть от одного до нескольких сотен, что регулируется параметром количества деревьев. В конце алгоритма среди всех деревьев и полученных итогов классификации по каждому дереву берется среднее их значение, поэтому метод называется ансамблевым.

Не менее важным является подбор оптимальных параметров для выбранного метода обучения модели. Для Random Forest существует большое количество изменяемых параметров, таких как глубина каждого дерева, упомянутое количество деревьев и др. Все они влияют на качество обучения [Pal, 2005]. Характер влияния тех или иных параметров может повышать или понижать разные метрики качества классификации, поэтому важно представлять конечную задачу классификации для правильного подбора параметров. Авторами были выбраны следующие параметры, с которыми проводились эксперименты с фиксированием точности (ассурасу) классифицированной тестовой выборки:

1. Глубина дерева — количество итераций алгоритма разделения выборки на подвыборки, произведенных деревом. Ее значение может быть от 3 для простых моделей до 30 для сложных [Шелухин, Полковников, 2019].
2. Количество деревьев — параметр, который оказывает менее сильное влияние на модель, нежели глубина деревьев, однако также влияющий на точность модели. Означает количество деревьев, используемых во всем «случайном лесу»; из итогов впоследствии находится среднее значение.

Оценка классификации проводилась с помощью построения матрицы ошибок и расчета на ее основе метрик качества модели, которые были следующими:

1. Ассурасу — доля правильных ответов алгоритма от всех возможных ответов. Самый неиспользуемый и неинформативный показатель, который, однако, служит хорошей метрикой для оценки параметров модели. Рассчитывается по формуле (2):

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2),$$

где TP — истинно положительный результат (true positive),  
 TN — истинно отрицательный результат (true negative),  
 FP — ложноположительный результат (false positive),  
 FN — ложноотрицательный результат (false negative).

2. Precision — доля истинно положительных результатов ко всем положительным результатам, предсказанным классификатором. Можно интерпретировать как долю объектов, названных классификатором положительными и при этом действительно являющимися положительными. Рассчитывается по формуле (3):

$$precision = \frac{TP}{TP + FP} \quad (3).$$

3. Recall (полнота) — доля истинно положительных результатов ко всем положительно размеченным моделью данным. Можно интерпретировать как способность модели обнаруживать истинно тот или иной класс среди всех положительно отнесенных к этому классу объектов. Рассчитывается по формуле (4):

$$recall = \frac{TP}{TP + FN} \quad (4).$$

4. F1-score — среднее гармоническое между recall и precision. Показывает, насколько похоже размечали данные авторы и модель. Рассчитывается по формуле (5):

$$f1 - score = \frac{2 * precision * recall}{precision + recall} \quad (5).$$

В итоге была создана обучающая выборка, которая впоследствии стала тренировочной, а также была создана тестовая выборка на основе вспомогательных материалов, и на ее основе были рассчитаны метрики качества модели. Кросс-валидация не проводилась, т. к. использовалась вся тестовая выборка для проверки, которая составлена экспертным образом и которая покрывала всю территорию исследования равномерно, вследствие чего ее разделение не предполагалось. Помимо этого, не предполагается применение модели на другую территорию, т. к. модель создана именно для рассматриваемой нами области с уникальными ландшафтами и экосистемами.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

На основе полученного растра, содержащего в себе каналы Sentinel-2, Landsat-9 и данные ArcticDEM, биомассы и высоты растительности, была рассчитана внутрикластерная сумма квадратов (WCSS) для определенного количества кластеров (от 1 до 20), и на их основе был составлен график (рис. 2). На нем можно наблюдать постепенное снижение значения WCSS с увеличением количества кластеров. Однако можно заметить, что только при количестве 3 или 4 кластеров значение показателя начинает снижаться не так резко. Это говорит об оптимальном количестве кластеров в 3 или 4, что не подходит для целей картографирования и классификации растительности. Из-за этого метод K-means в дальнейшем не рассматривался. Кроме того, наблюдается низкая разделяемость классов на протяжении всего графика.



*Рис. 2. WCSS для кластеров классификации K-means*  
*Fig. 2. WCSS for K-means classification clusters*

Дальнейшая работа состояла из неоднократного применения модели на тестовых данных для оценки получающихся классов, их разделимости и оценки параметров модели, оптимальных для обучения. Одной из задач в т. ч. являлось выделение двух видов болот (низинных и переходных) и выделение высокотравных лугов в поймах рек. Однако в дальнейшем, при подсчете accuracy (точности) для данных классов, выяснилось, что эти три класса относительно неразделимы друг от друга, а их точность составляет 20 % и менее. При этом, судя по составленной в начале исследования матрице путаницы, луга путаются почти со всеми остальными классами, а болота между собой. Из-за этого было принято решение убрать класс лугов, а два класса болот объединить. Важным моментом было решение об удалении классов сочетаний растительности. Как показали эксперименты, добавление в модель классов сочетаний растительности снижает ее точность и не позволяет создать обучающую и тестовую выборки более объективными. Таким образом, остальные классы имели точность выше 70 %, а большинство из них обладали точностью в 80 % и более, поэтому было принято решение оставить то количество классов, которое было указано ранее (8 классов растительности и 2 класса поверхности). На этом же этапе было принято использовать данные ЦМР и 10 канал Landsat-9, что позволило увеличить общую точность модели на 8–10 %, а для некоторых классов (термальная растительность, растительность на осыпях) — на 20 % и более.

При окончательном формировании обучающей выборки и испытании метода локтя для кластеризации K-means дальнейшим шагом работы явилось проведение экспериментов для выявления лучших параметров для модели, основанной на Random Forest. Так, были протестированы параметры глубины деревьев и их количества. Для реализации данных экспериментов были использованы библиотеки Matplotlib и Seaborn для Python. Полученный график изменения точности от глубины деревьев говорит об оптимальной

глубине дерева, равной 10, т. к. именно на ней точность имеет наивысшее значение. После него модель, как предполагается авторами, переобучается, что видно на понижающейся точности тестовой выборки и неизменной точности тренировочной (обучающей) выборки (рис. 3).

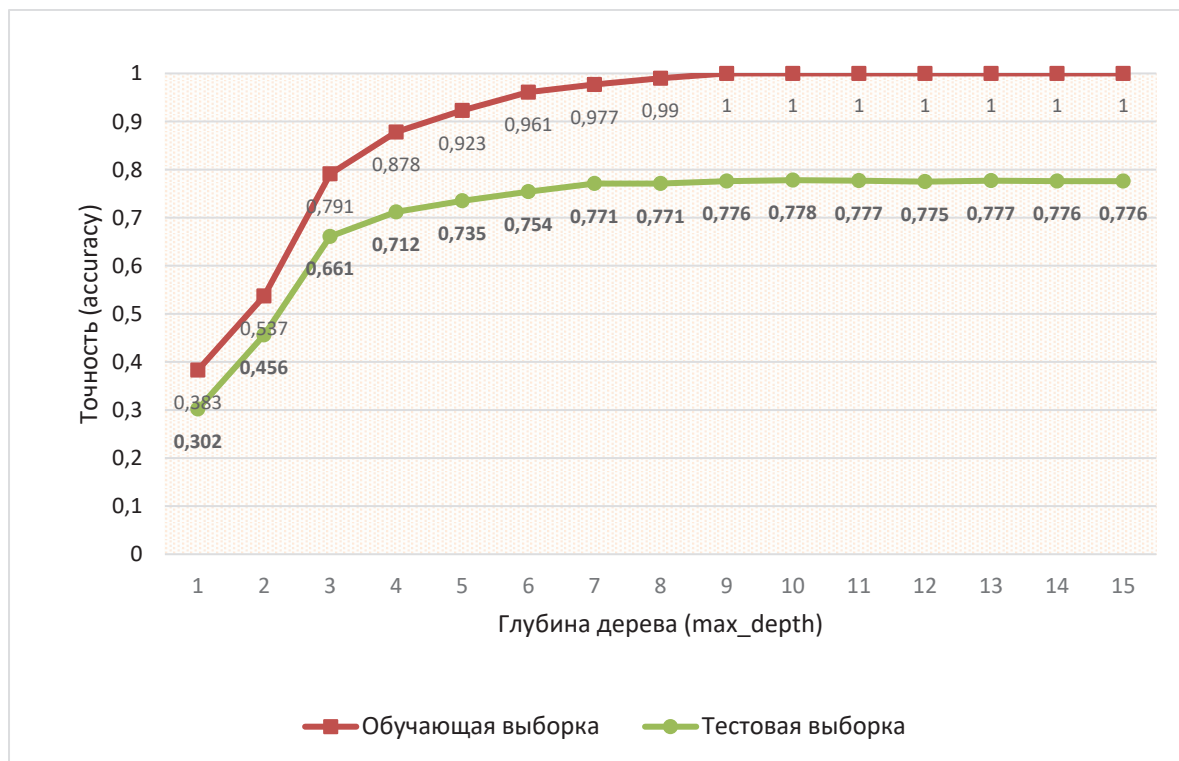


Рис. 3. Изменение точности для тестовой и обучающей выборок с изменением глубины дерева

Fig. 3. Changing the accuracy for test and training sets with changing tree depth

При анализе количества деревьев их диапазон для экспериментов был выбран от 0 до 175. Каждый раз при увеличении количества деревьев фиксировалось значение точности модели (рис. 4). При увеличении до количества деревьев от 0 до 75 происходил быстрый рост точности, однако после 75 существенного роста не наблюдается. Авторами было выбрано значение 150 в качестве оптимального и дающего наибольшую точность.

Таким образом, при выставлении всех оптимальных параметров, была получена модель, общая точность (ассигасу) которой составляет 90 %. При этом значение точности разнится от класса к классу (рис. 5). Были выделены следующие классы растительного покрова и типов поверхности:

- 1 — ольховый стланик,
- 2 — кедровый стланик,
- 5 — горные кустарничковые тундры,
- 6 — лишайниковые тундры,
- 7 — растительность на осыпях,
- 9 — болота,
- 10 — термальная растительность,
- 11 — березняки,
- 12 — водные объекты,
- 13 — снег.



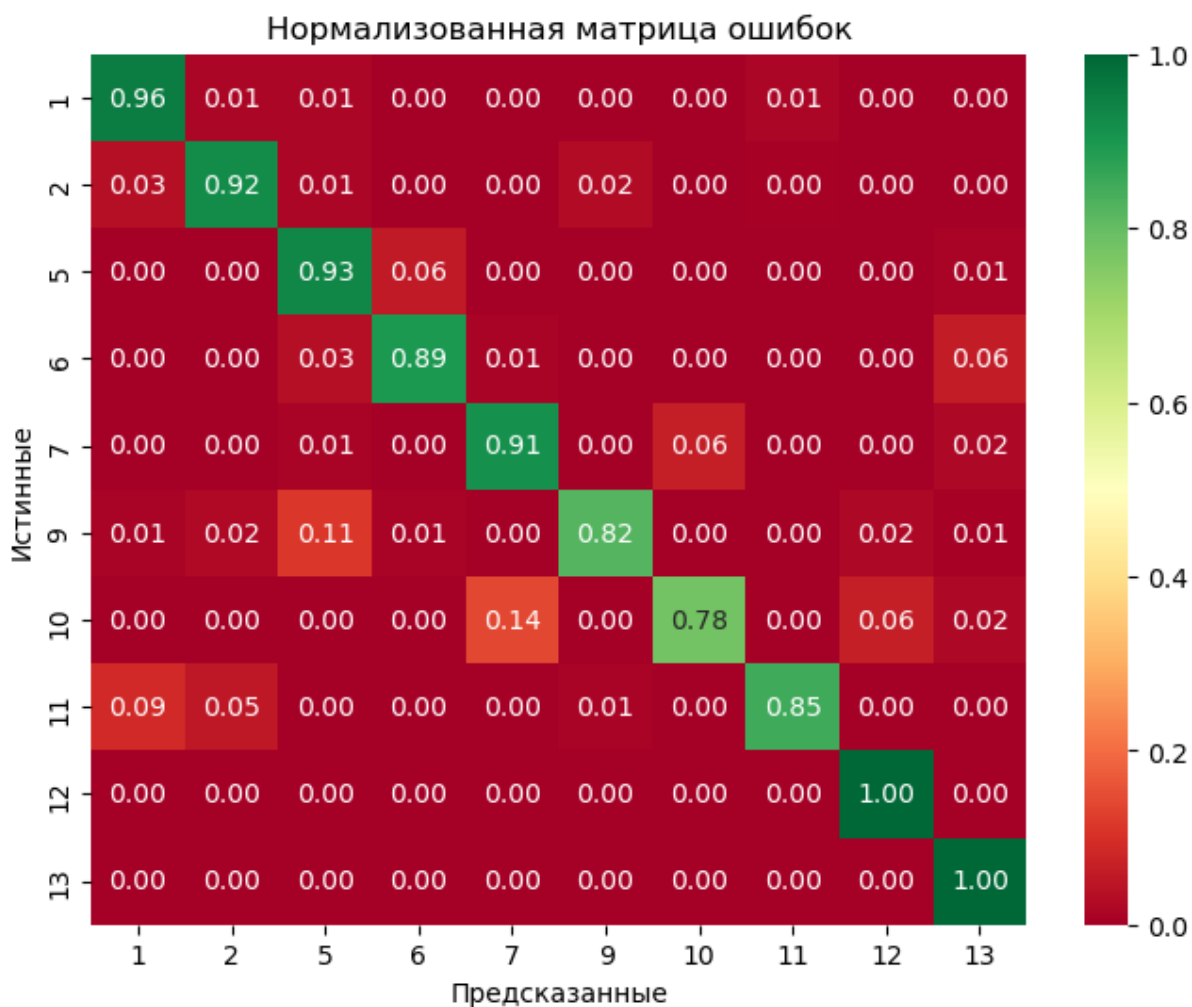
Рис. 4. Точность модели в зависимости от количества деревьев

Fig. 4. Model accuracy depending on the number of trees

Класс 8 отсутствует, т. к. изначально это был класс лугов, который авторы решили исключить из обучающей и тестовой выборок. Кроме того, классы 3 и 4 не представлены на графике, т. к. ими являлись сочетания тундр и ольховых и кедровых стлаников, которые были удалены в ходе проведения оценки точности моделей при проведении экспериментов. Так, для классов 9–11 точность оказалась относительно низкой, а для классов 12 и 13 — максимально высокой. Первое связано со сложностью дешифрирования классов и выделения их признаков. Для болот (похожих на тундры), термальной растительности (похожей на растительность на осыпях) и березняков (похожих на стланик) важными оказались дополнительные данные о высоте растительности и биомассе, с помощью которых удалось достичь относительно высокой точности классификации. Стоит также учесть, что максимальной объективности в создании обучающей выборки и оценке точности данных классов достичь крайне затруднительно из-за того, что принадлежность пикселя к классу не всегда очевидна даже опытному ботанику и специалисту по дешифрированию растительности. Классы снежников (снега) и водных объектов оказались легко разделимы из-за их выраженных спектральных характеристик. Однако полученный показатель точности (ассигасу) говорит лишь об обобщенной точности модели относительно всех ее статистических показателей (TP, TN, FP, FN), поэтому необходим анализ остальных метрик.

Для дальнейшей оценки классификации и работы модели необходимо проанализировать упомянутые ранее метрики, рассчитанные для каждого класса (табл. 1). Проанализировав таблицу можно прийти к выводу о том, что при выделении растительности для некоторых классов модель «недорабатывает», а для остальных — «перерабатывает». Об этом говорит разница между значениями метрик precision и recall. Так, модель

«недорабатывает» для большинства классов. На примере с березняками: она будет точно выделять березняки там, где они действительно есть, пропуская при этом их часть, и не выделяя то, что к ним не относится. Наоборот, в случае, когда модель перерабатывает, она дает показатель полноты (recall) больше показателя точности (precision) и выделяет «лишние» пиксели как данный класс, что характерно для снежников и кустарничковой тундры. Показатель f1-score выявил относительную согласованность в разметке классов между авторами и моделью, исключая класс «снег», который показал высокую точность (accuracy) при большой разнице между показателями precision и recall, что говорит о завышенном количестве пикселей, классифицируемых как этот класс.



*Рис. 5. Нормализованная матрица ошибок для тестовой выборки*  
*Fig. 5. Normalized confusion matrix for the test sample*

Помимо этого, была проанализирована информативность каналов классифицируемого растра (рис. 6). Так, наибольшей информативностью обладают ЦМР и 2, 5, 6 каналы Sentinel-2. Однако предполагается, что важное значение в разделении сложных классов имеют 10-й канал Landsat-9, данные о биомассе и высоте растительности.

Табл. 1. Значения метрик оценки классификации для каждого класса  
 Table 1. Classification evaluation metric values for each class

Класс	precision	recall	f1-score
Ольховый стланик	0,90	0,96	0,93
Кедровый стланик	0,93	0,92	0,92
Горные кустарничковые тундры	0,82	0,93	0,93
Лишайниковые тундры	0,94	0,89	0,91
Растительность на осыпях	0,91	0,91	0,91
Болота	0,96	0,82	0,89
Термальная растительность	0,85	0,78	0,81
Березняки	0,97	0,85	0,91
Водные объекты	0,92	1,00	0,96
Снег	0,77	1,00	0,77

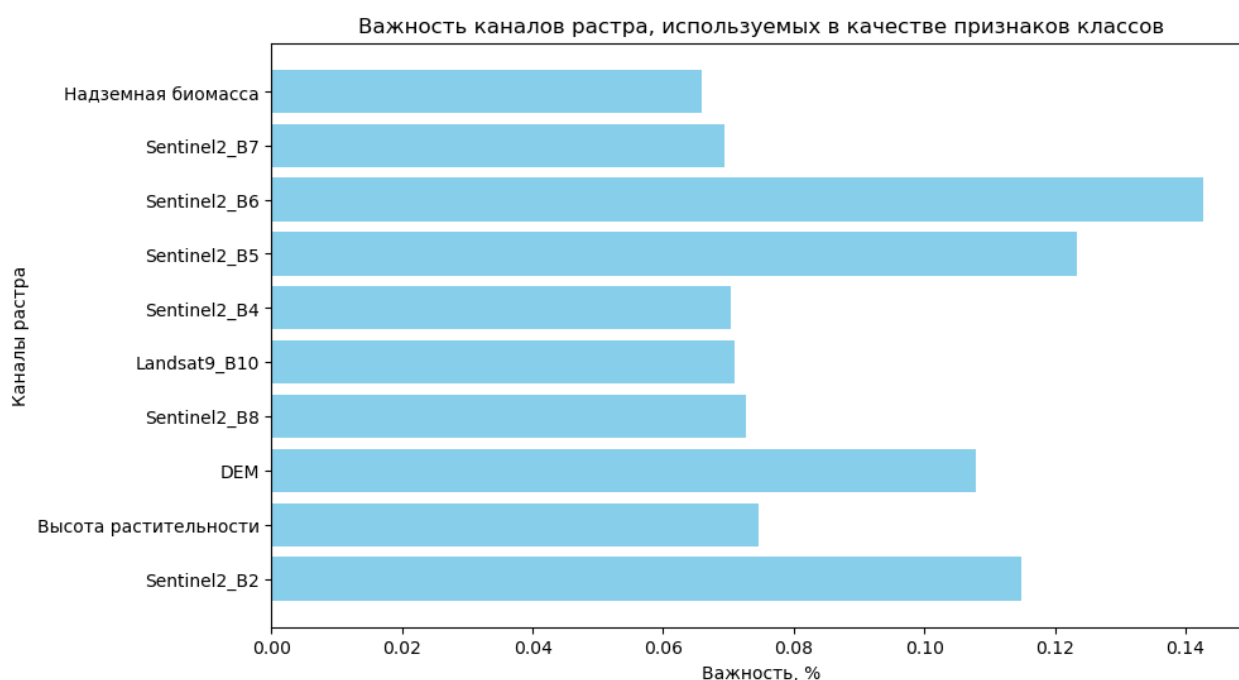


Рис. 6. Информативность признаков для обучения модели  
 Fig. 6. Informativeness of features for model training

Далее выполнен перевод получившегося в результате классификации растра в вектор с помощью соответствующего инструмента в QGIS, на основе которого была составлена картографическая модель растительности части кальдеры Узон (рис. 7).

С помощью расчета площадей было выявлено, что большую часть территории занимают ольховые стланики (почти 20 %), кедровые стланики (16 %), меньшую — термальная растительность, березняки и снег (0,5, 0,4 и 0,3 % соответственно).

## Карта растительности части кальдеры Узон

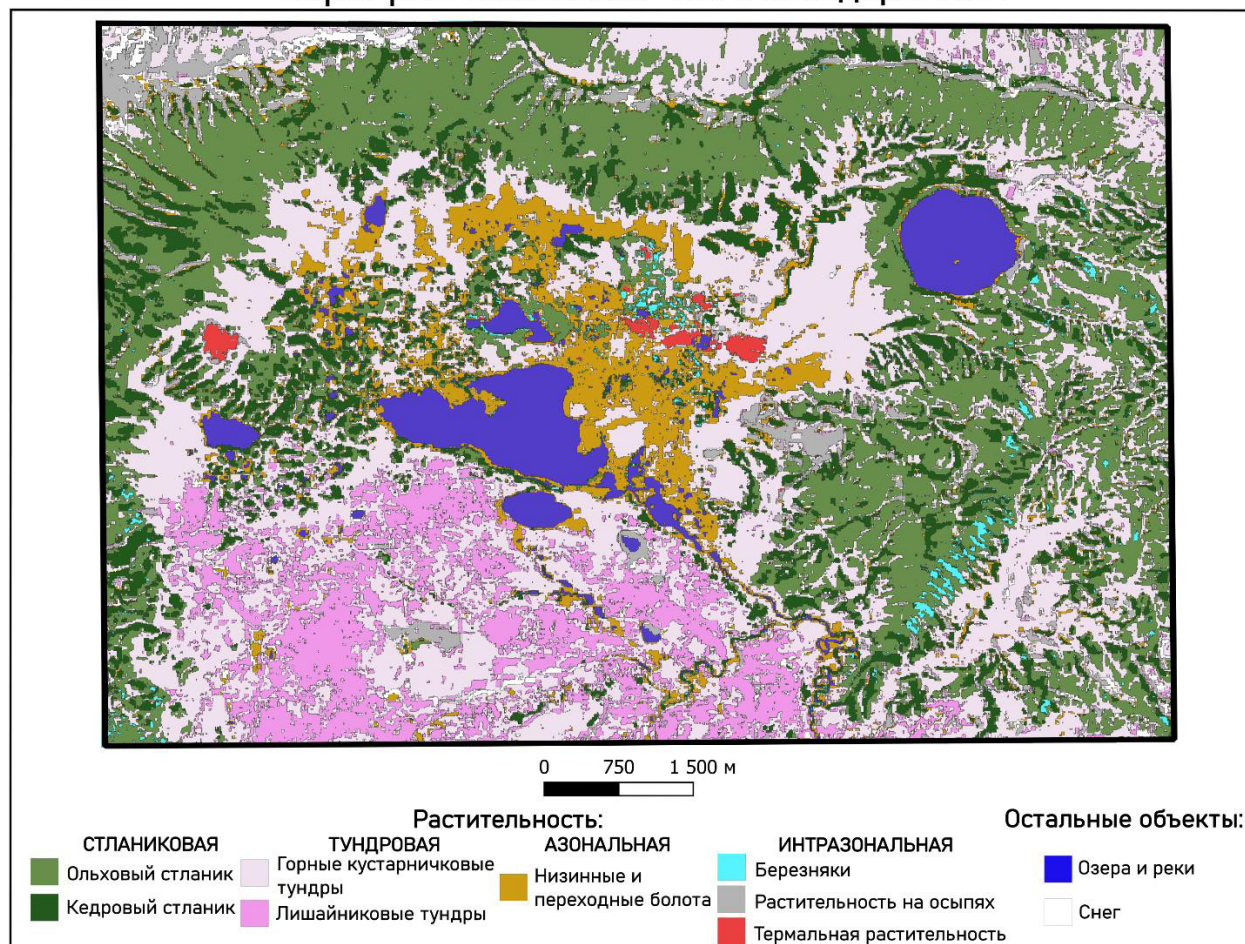


Рис. 7. Картографическая модель растительности части кальдеры Узон  
 Fig. 7. Mapping model of vegetation of part of the Uzon caldera

## ВЫВОДЫ

Авторами была предложена картографическая модель растительности кальдеры Узон, основанная на машинном обучении методом Random Forest. Предварительный анализ кластеризации показал необходимость использования данного метода, а оптимизация параметров обучаемой модели позволила достичь относительно хорошей точности модели в 90 %.

Анализ метрик классификации позволил понять слабые и сильные стороны модели, а также то, в каком направлении стоит ее улучшать. Так, для каждого класса были рассчитаны показатели precision и recall, выявляющие недостатки модели, что поможет в дальнейшем улучшении обучающей выборки. Помимо этого, был рассчитан показатель f1-score.

Не менее важным оказался расчет информативности признаков, показавший важность ЦМР и некоторых каналов Sentinel-2, которые характеризуют растительность.

В дальнейшем планируется использование большего количества признаков (климатических данных WorldClim и других данных ДЗЗ), которые помогут улучшить классификацию и модель в целом, а также дополнение обучающей выборки, повышение ее сбалансированности и качества.

Полученная картографическая модель может послужить основой для оценки состояния экосистем, биоразнообразия растительных сообществ и видов уникальной

территории исследования. В то же время полученная модель обучения, основанная на Random Forest, может быть использована для других временных периодов на эту же территорию, что позволит проводить мониторинг растительности кальдеры Узон или рассматривать ее в ретроспективе.

### СПИСОК ЛИТЕРАТУРЫ

- Котлов И. П., Черненко Т. В., Беляева Н. Г.* Опыт картографирования типологического разнообразия лесов Московской области. Геоботаническое картографирование, 2024. С. 25–43. DOI: 10.31111/geobotmap/2024.25.
- Мальшева Н. В.* Автоматизированное дешифрирование аэрокосмических изображений лесных насаждений. М.: Издательство МГУЛ, 2012. 154 с.
- Нешатаева В. Ю., Пестеров А. О.* Структура растительного покрова кальдеры Узон (Восточная Камчатка). Известия Самарского научного центра Российской академии наук, 2012. Т. 14. № 1-4. С. 1080–1083.
- Нешатаева В. Ю., Пестеров А. О., Кораблёв А. П.* Растительность Восточного вулканического пояса Камчатки (в пределах Кроноцкого заповедника). СПб.: Марафон, 2021. 67 с.
- Пестеров А. О.* Ценотическое разнообразие и структура растительного покрова Восточного вулканического пояса полуострова Камчатка: Диссертация на соискание ученой степени кандидата биологических наук. СПб., 2017. 275 с.
- Полевицкова Ю. А., Кириллов В. А.* Мультиклассовая сегментация типов растительности с применением облачной вычислительной платформы Google Earth Engine. Информационные технологии и высокопроизводительные вычисления. Материалы VII Международной научно-практической конференции, 2023. С. 18–21.
- Пузаченко М. Ю., Котлов И. П., Черненко Т. В.* Технологическая схема мониторинга природных объектов с использованием ДДЗ и ГИС-технологий. Мониторинг биологического разнообразия лесов России. М.: Наука, 2008. С. 347–359.
- Шашков М. П., Иванова Н. В., Ермолов С. А.* Применимость наборов данных растровых слоев с факторами среды в качестве предикторов для моделирования ареалов дождевых червей. Доклады Международной конференции «Математическая биология и биоинформатика», 2024. С. 228–232. DOI: 10.17537/icmbb24.55.
- Шелухин О. И., Полковников М. В.* Повышение эффективности классификации сетевого трафика мобильных устройств с использованием алгоритма Random Forest. Технологии информационного общества: Материалы XIII Международной отраслевой научно-технической конференции, 2019. С. 415–418.
- Boateng E. Y., Otoo J., Abaye D. A.* Basic Tenets of Classification Algorithms K-nearest-neighbor, Support Vector Machine, Random Forest and Neural Network. Journal of Data Analysis and Information Processing, 2020. V. 8. No. 4. P. 341–357. DOI: 10.4236/jdaip.2020.84020.
- Boitard S., Mialon A., Mermoz S., Rodríguez-Fernández N. J., Richaume P., Salazar-Neira J. C., Tarot S., Kerr Y. H.* Aboveground Biomass Dataset from SMOS L-band Vegetation Optical Depth and Reference Maps. Earth System Science Data, 2025. V. 17. No. 3. P. 1101–1119. DOI: 10.1186/s13021-022-00214-w.
- Mohammadpour P., Viegas D. X., Viegas C.* Vegetation Mapping with Random Forest Using Sentinel-2 and GLCM Texture Feature — A Case Study for Lousã Region, Portugal. Remote Sensing, 2022. V. 14. No. 18. P. 4585. DOI: 10.3390/rs14184585.

*Pal M.* Random Forest Classifier for Remote Sensing Classification. *International Journal of Remote Sensing*, 2005. V. 26. No. 1. P. 217–222. DOI: 10.1080/01431160412331269698.

*Thanh Noi P., Kappas M.* Comparison of Random Forest, K-nearest-neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 2017. V. 18. No. 1. P. 18. DOI: 10.3390/s18010018.

*Tolan J., Yang H. I., Nosarzewski B., Couairon G., Vo H. V., Brandt J., Spore J., Majumdar S., Haziza D., Vamaraju J., Moutakanni T., Bojanowski P., Johns T., White B., Tiecke T., Couprie C.* Very High Resolution Canopy Height Maps from RGB Imagery Using Self-Supervised Vision Transformer and Convolutional Decoder Trained on Aerial Lidar. *Remote Sensing of Environment*, 2024. V. 300. P. 113888. DOI: 10.1016/j.rse.2023.113888.

## REFERENCES

*Boateng E. Y., Otoo J., Abaye D. A.* Basic Tenets of Classification Algorithms K-nearest-neighbor, Support Vector Machine, Random Forest and Neural Network. *Journal of Data Analysis and Information Processing*, 2020. V. 8. No. 4. P. 341–357. DOI: 10.4236/jdaip.2020.84020.

*Boitard S., Mialon A., Mermoz S., Rodríguez-Fernández N. J., Richaume P., Salazar-Neira J. C., Tarot S., Kerr Y. H.* Aboveground Biomass Dataset from SMOS L-band Vegetation Optical Depth and Reference Maps. *Earth System Science Data*, 2025. V. 17. No. 3. P. 1101–1119. DOI: 10.1186/s13021-022-00214-w.

*Kotlov I. P., Chernenkova T. V., Belyaeva N. G.* Mapping Experience on the Forest Typological Diversity in the Moscow Region. *Geobotanical Mapping*, 2024. P. 25–43 (in Russian). DOI: 10.31111/geobotmap/2024.25.

*Malysheva N. V.* Automated Interpretation of Aerospace Images of Forest Stands. Moscow: Publishing House of Moscow State University of Forestry, 2012. 154 p. (in Russian).

*Mohammadpour P., Viegas D. X., Viegas C.* Vegetation Mapping with Random Forest Using Sentinel-2 and GLCM Texture Feature — A Case Study for Lousã Region, Portugal. *Remote Sensing*, 2022. V. 14. No. 18. P. 4585. DOI: 10.3390/rs14184585.

*Neshataeva V. Yu., Pesterov A. O.* The Vegetation Cover Structure of Uzon Volcano Caldera (Eastern Kamchatka). *Izvestiya Samarskogo nauchnogo centra RAN (News of the Samara Scientific Center of the Russian Academy of Sciences)*, 2012. V. 14. No. 1-4. P. 1080–1083 (in Russian).

*Neshataeva V. Yu., Pesterov A. O., Korablev A. P.* Vegetation of the Eastern Volcanic Belt of Kamchatka (Within the Kronotsky Reserve). St. Petersburg: Marathon, 2021. 67 p. (in Russian).

*Pal M.* Random Forest Classifier for Remote Sensing Classification. *International Journal of Remote Sensing*, 2005. V. 26. No. 1. P. 217–222. DOI: 10.1080/01431160412331269698.

*Pesterov A. O.* Coenotic Diversity and Structure of Vegetation Cover of the Eastern Volcanic Belt of the Kamchatka Peninsula: PhD dissertation of biological sciences. St. Petersburg, 275 p. (in Russian).

*Polevshchikova Yu. A., Kirillov V. A.* Multiclass Classification of Vegetation Types Using the Google Earth Engine Cloud-Based Platform. *Information Technologies and High-Performance Computing: Proceedings of the VII International Scientific and Practical Conference*, 2021. P. 18–21 (in Russian).

*Puzachenko M. Yu., Kotlov I. P., Chernenkova T. V.* Technological Scheme for Monitoring Natural Objects Using Remote Sensing and GIS Technologies. *Monitoring the Biological Diversity of Russian Forests*. Moscow: Nauka, 2008. P. 347–359 (in Russian).

*Shashkov M. P., Ivanova N. V., Ermolov S. A.* Usability of Raster Spatial Environmental Datasets as Predictors for Species Distribution Modelling of Earthworms. Proceedings of the International Conference “Mathematical Biology and Bioinformatics”, 2024. P. 228–232 (in Russian). DOI: 10.17537/icmbb24.55.

*Sheloukhin O. I., Polkovnikov M. V.* Improving the Classification Efficiency of Network Traffic of Mobile Devices Using the Random Forest Algorithm Technologies of the Information Society. Materials of the XIII International Industrial Scientific and Technical Conference, 2019. P. 415–418 (in Russian).

*Thanh Noi P., Kappas M.* Comparison of Random Forest, K-nearest-neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors*, 2017. V. 18. No. 1. P. 18. DOI: 10.3390/s18010018.

*Tolan J., Yang H. I., Nosarzewski B., Couairon G., Vo H. V., Brandt J., Spore J., Majumdar S., Haziza D., Vamaraju J., Moutakanni T., Bojanowski P., Johns T., White B., Tiecke T., Couprie C.* Very High Resolution Canopy Height Maps from RGB Imagery Using Self-Supervised Vision Transformer and Convolutional Decoder Trained on Aerial Lidar. *Remote Sensing of Environment*, 2024. V. 300. P. 113888. DOI: 10.1016/j.rse.2023.113888.

---