

Санкт-Петербургский государственный университет

ЛЕБЕДЕВ Владислав Леонидович

Выпускная квалификационная работа

Мониторинг нарушенных земель Ленинградской области на основе данных дистанционного зондирования и географических информационных систем

Уровень образования:

Направление *05.04.03 «Картография и геоинформатика»*

Основная образовательная программа *ВМ.5523 «Геоинформационное картографирование»*

Научный руководитель:

заведующий кафедрой картографии и геоинформатики,
канд. техн. наук, Паниди Евгений Александрович

Рецензент:

старший инженер-эколог ЗАО «Экопроект»,
Рябцев Иван Сергеевич

Санкт-Петербург

2023

Содержание

Введение	3
1 Мониторинг нарушенных лесных земель	6
1.1 Нормативное регулирование мониторинга нарушенных земель	6
1.2 Применение данных ДЗЗ для мониторинга лесных земель	7
1.3 Российские разработки мониторинга нарушенных лесных земель	10
2 Методы мониторинга нарушенных лесных земель	12
2.1 Основные методы картографирования нарушенных лесных земель.....	12
2.1.1 Визуальное дешифрирование	12
2.1.2 Применение вегетационных индексов.....	13
2.1.3 Методы автоматизированного дешифрирования.....	15
2.2 Исходные данные ДЗЗ и геоинформационная среда	19
2.2.1 Открытые данные ДЗЗ	19
2.2.2 Геоинформационная платформа Google Earth Engine для исследования нарушения лесных земель	21
2.2.3 Нейронные сети для анализа данных ДЗЗ.....	24
3 Разработка методики картографирования нарушенных лесных земель с использованием Google Earth Engine	26
3.1 Методика картографирования нарушенных лесных земель с использованием Google Earth Engine	26
3.2 Результаты исследовательской работы	29
3.2.1 Сравнение вегетационных индексов NDVI и NBR	29
3.2.2 Классификация с учителем с использованием машинного обучения ..	32
3.2.3 Результаты модели классификации	35
Заключение	40
Список литературы	41
Приложение А	46
Приложение Б.....	47
Приложение В	48
Приложение Г	49

Введение

Деятельность человека, вызванная растущей урбанизацией, расширением сельского хозяйства, лесопользованием, оказывает весомое воздействие на экосистемы Земли. Изменения в землепользовании, вырубка лесов, добыча полезных ископаемых, загрязнение земель являются известными примерами нарушений, которые изменяют естественные ландшафты на антропогенные. Такие преобразования часто приводят к ухудшению качества почвы, утрате биоразнообразия, изменению гидрологических режимов и нарушению экосистем. Признавая серьезность этих проблем, важно проводить детальный мониторинг нарушенных земель для оценки степени и скорости изменений, а также экологических последствий антропогенных нарушений.

Леса как жизненно важные природные экосистемы выполняют множество функций, необходимых как для человека, так и окружающей среды. Они не только обеспечивают нас древесиной, но также служат важными местами обитания диких животных, очищают воздух, которым мы дышим, защищают качество водных ресурсов и действуют как значительные поглотители углерода, смягчая последствия изменения климата. Мониторинг ущерба, наносимого лесам, особенно в результате лесозаготовок и лесных пожаров, стали важнейшей задачей современного лесопользования.

Леса в Ленинградской области занимают 57% территории региона, по данным на 2014 год. Ленинградская область относится к давно освоенным регионам Русского Севера, значительное изменение природных ландшафтов обусловлено главным образом многократной вырубкой лесов и осушением большей части болот.

В Ленинградской области находится 54 особо охраняемые природные территории (ООПТ), занимающие 7,2% площади по данным на 2019 год. Федеральный статус имеют государственные природные заповедники «Нижне-Свирский» и «Восток Финского залива», государственный природный заказник «Мшинское болото», региональный – природные парки «Вепсский лес» и «Токсовский», 27 государственных природных заказников, 18 памятников природы.

Лесная, деревообрабатывающая и целлюлозно-бумажная промышленность – одна из базовых отраслей хозяйства области, при том, что Ленинградская область – один из ведущих регионов на российском Северо-Западе по объёму лесозаготовок, деревообработки и экспорту древесины. Основной район лесозаготовок – северо-восточная часть области [1].

Земли лесного фонда – самая динамичная категория земель, поскольку зачастую расширение других категорий земель происходит за счет лесных или

сельскохозяйственных земель, что является следствием расширения населенных пунктов, строительства промышленных объектов и добычи полезных ископаемых. Понимание масштабов и воздействия этих нарушений крайне важно для обеспечения долгосрочного здоровья, биоразнообразия и функциональности лесов.

Эффективный мониторинг нарушения лесного покрова обеспечивает информацией для принятия обоснованных решений государственные и муниципальные органы власти, специалистов по охране природы и землепользователей. Мониторинг нарушенных лесных земель должен позволить осуществлять устойчивое лесопользование совместно с применением современных информационных технологий.

Мощной информационной поддержкой мониторинга нарушенных земель, в том числе лесных, являются данные дистанционного зондирования (ДЗЗ) при использовании географических информационных систем (ГИС). Данные ДЗЗ – важнейший источник своевременной и оперативной информации о природной среде для создания тематических слоёв в базе данных ГИС. Серверные части традиционных ГИС сталкиваются с серьезными ограничениями масштабируемости в результате появления больших данных в виде спутниковых изображений.

В ходе научно-исследовательской работы реализовано внедрение современных технологий облачных вычислений, позволяющих манипулировать большими пространственными данными вне стационарного традиционных настольных ГИС. Используются алгоритмы машинного обучения для задач дешифрирования нарушенных лесных земель, которые позволяют упростить обработку спутниковых изображений, увеличить скорость и сократить трудозатраты на выполнение поиска и векторизации признаков нарушения лесного покрова.

Объектом магистерской научно-исследовательской работы выбраны нарушенные лесные земли Ленинградской области, предметом – методы картографирования нарушенных лесных земель по данным ДЗЗ с использованием ГИС.

Значимость исследования состоит в автоматизации мониторинга лесных ресурсов с целью своевременного обнаружения негативных процессов на нарушенных лесных землях и прилегающих к ним территориях, и в способствовании принятия управленческих решений ответственными органами власти.

Актуальность темы научно-исследовательской работы обусловлена проблемой оценки состояния и использования земель на фоне неизменно возрастающего техногенного воздействия на земельные ресурсы нашей страны, что приводит к ухудшению земельного фонда и изменению природных ландшафтов.

Цель научно-исследовательской работы – разработка методики автоматизированного мониторинга нарушенных лесных земель по данным дистанционного зондирования Земли на территории Ленинградской области.

Сформулированы следующие задачи для достижения поставленной цели:

- Изучить накопленный российский и зарубежный опыт в решении проблемы картографирования нарушенных лесных земель;
- Сравнить методы картографирования нарушения и восстановления лесных земель по данным ДЗЗ;
- Подобрать наиболее подходящие методы, данные ДЗЗ и геоинформационное программное обеспечение;
- Провести опыт и составить схему разрабатываемой методики;
- Составить картографические материалы и представить результаты методики мониторинга.

1 Мониторинг нарушенных лесных земель

1.1 Нормативное регулирование мониторинга нарушенных земель

Нарушенными землями считаются земли, подвергшиеся неблагоприятным процессам. Нарушение земель относится к деятельности, связанной с добычей полезных ископаемых, геологоразведкой, изысканиями, строительством и другими задачами, которые приводят к нарушению почвенного покрова, гидрологического режима в районе, созданию техногенных форм рельефа и различным качественным изменениям состояния земли.

К нарушенным землям относятся земли, деградация которых привела к невозможности их использования в соответствии с целевым назначением и разрешенным использованием [2]. Таким образом, в процессе изменения природных ландшафтов и образования антропогенных ландшафтов меняется и вид целевого назначения земель. Для устранения негативного воздействия на нарушенные территории применяют комплекс мер по рекультивации таких земель.

Нарушенные земли классифицируют по типам и причинам возникновения, выделяют основные причины, такие как:

- строительство линейных объектов (ЛЭП, автомобильных, железных дорог);
- промышленное лесопользование (промышленная вырубка, в том числе просек);
- гидротехническое строительство (водохранилища, сооружение дамб);
- недропользование (разработка и эксплуатация месторождений полезных ископаемых);
- геологоразведочные работы и другие работы, связанные с нарушением почвенного покрова;
- сельскохозяйственное строительство (агромелиоративное);
- складирование промышленных и бытовых отходов.

В Российской Федерации в нормативно-правовом поле мониторинг нарушенных земель не упоминается как обособленный термин, по факту он является частью государственного экологического мониторинга (государственного мониторинга окружающей среды), установленного в соответствии с Федеральным законом “Об охране окружающей среды”.

При этом в соответствии со статьей 67 Земельного кодекса РФ осуществляется государственный мониторинг земель, в свою очередь он является частью государственного экологического мониторинга и представляет собой систему

наблюдений, оценки и прогнозирования, направленных на получение достоверной информации о состоянии земель, об их количественных и качественных характеристиках, их использовании и о состоянии плодородия почв.

Объектами государственного мониторинга земель являются все земли в Российской Федерации.

При осуществлении мониторинга земель установлены следующие методы [3]:

- анализ данных дистанционного зондирования Земли;
- наземные съемки, наблюдения и обследования (сплошные и выборочные);
- привлечение результатов обновления картографической основы (результатов дешифрирования ортофотопланов или сведений топографических карт и планов);
- анализ сведений, содержащихся в государственном кадастре недвижимости, землеустроительной документации, материалов инвентаризации и обследования земель.

Распространение негативных процессов природного и антропогенного характера приводит к ухудшению состояния земель. Полученные фактические данные государственного мониторинга земель позволяют предотвратить развитие негативных процессов [4].

Согласно Лесному кодексу РФ к лесным землям относятся земли, на которых расположены леса, и земли, предназначенные для лесовосстановления (вырубки, гари, редины, пустыри, прогалины и другие). К нелесным землям относятся земли, необходимые для освоения лесов (просеки, дороги и другие), и земли, неудобные для использования (болота, каменистые россыпи и другие).

Также согласно статье 61.1 Лесного кодекса РФ проводится государственный мониторинг воспроизводства лесов, являющийся аналогично государственному мониторингу земель, частью экологического мониторинга. Мониторинг воспроизводства лесов включает в себя оценку изменения площади земель, на которых расположены леса, а также осуществляется с использованием наземных методов и методов дистанционного зондирования Земли [5].

1.2 Применение данных ДЗЗ для мониторинга лесных земель

Как известно, данные ДЗЗ – важнейший источник своевременной и оперативной информации о природной среде для создания тематических слоёв в базе данных ГИС. Данные ДЗЗ и картографические материалы в большей степени обеспечивают проведение мониторинга земель.

Данные дистанционного зондирования, полученные со спутников, могут использоваться для мониторинга и картографирования использования земель и растительного покрова. Наблюдение изменений чаще всего подразумевает использование разновременных спутниковых снимков, на основе которых проводится анализ и прогнозирование [6].

Изучение и картографирование изменений в природе, природопользовании и степени антропогенного воздействия на природу — задачи мониторинга, которые связаны с построением карт динамики. Основу методов их создания составляет сопоставление разновременных космических снимков и карт, фиксирующих состояние исследуемого объекта на разные даты. Компьютерная реализация этих методов главным образом направлена на обеспечение точного пространственного совмещения материалов и извлечение из них динамической информации [7].

Результатом мониторинга земель вне зависимости от типа всегда является картографическое изображение, демонстрирующее изменения земельного покрова, которое в настоящее время подготавливается средствами ГИС. Выявление распространенности нарушенных земель, зафиксированное на разновременных космических снимках, позволяет выполнить оценку этих изменений. Снимки подвергаются цифровой обработке, визуальному анализу и дешифрированию. ГИС на основе данных ДЗЗ позволяют вычислить площади и сравнить изменения за выбранный период [4, 8].

В ряде публикаций по исследованию мониторинга нарушенных земель используется определение нормализованного разностного вегетационного индекса NDVI [7, 9, 10]. Это простой количественный показатель фотосинтетически активной биомассы, вычисляемый как функция по мультиспектральным снимкам с использованием красного и инфракрасного электромагнитного спектра, регистрируемого ДЗЗ. Высокая фотосинтетическая активность, связанная, как правило, с густой растительностью, ведет к меньшему отражению в красной области спектра и большему в инфракрасной. Отношение этих показателей друг к другу позволяет четко отделять и анализировать растительные от прочих природных объектов. Так, например, используя индекс NDVI, можно судить об эффективности рекультивации после окончания отработки месторождений полезных ископаемых.

На примере Новиковского бурогоугольного месторождения были получены результаты пространственно-временного анализа состояния растительного покрова. На основе использования космических снимков среднего разрешения систем Landsat и Sentinel за период с 1980 по 2020 гг. в геоинформационном ПО были рассчитаны значения

NDVI. Это позволило определить современный состав растительного покрова и проследить динамику и характер сукцессии на пораженной территории [9].

Обычно мониторинг нарушений лесного покрова производится по серии последовательных снимков, полученным в течение вегетационного периода. При использовании летних снимков наиболее информативен средний ИК канал, поэтому для выделения свежих гарей и ветровалов создают мультивременные композиты на основе этого канала [11].

Проблеме мониторинга нарушенных земель разного типа (причин нарушения) уделено несколько диссертационных исследований, например исследование Ю. В. Рябова. В работе автором определена структура и содержание ГИС мониторинга земель, нарушенных несанкционированными свалками. В исследовании ставились такие задачи, как выявление основных факторов воздействия свалок на окружающую среду и земельные ресурсы, определение характеристик спутниковых снимков для обнаружения захламливаемых земель, предложен метод прогнозирования мест возникновения несанкционированных свалок, представлена структура базы данных ГИС, обоснованы подходы к принятию решений о защите земельных ресурсов от возникновения несанкционированных свалок и разработана технология мониторинга земель, нарушенных несанкционированными свалками [12].

В диссертационном исследовании, посвященном автоматизированной оценке пирогенной гибели лесов применялась подобная методика [13]:

- Получение спутниковых изображений высокого пространственного разрешения Landsat-TM/ETM+ и SPOT-HRV/HRVIR;
- Использование кластерного анализа по методу ISODATA с целью разбиения гарей на спектрально однородные участки и последующее выделение пространственно связанных областей;
- Расчет нормализованного разностного вегетационного индекса NDVI для пространственно связанных областей с последующим разбиением на классы;
- Совместный анализ наземных и спутниковых данных о повреждении лесов пожарами.

В результате обзора литературы по направлению мониторинга нарушенных лесных земель отмечено, что в основном для проведения мониторинга используются изображения, полученные вычислением вегетационных индексов, в основном NDVI. В следующих разделах более подробно рассмотрена методика применения этого индекса, а также результативность его применения для автоматизированного мониторинга.

1.3 Российские разработки мониторинга нарушенных лесных земель

Важно отметить коммерческие разработки мониторинга нарушенных земель, в том числе лесных, на основе данных ДЗЗ российской группировки спутников Канопус-В, Ресурс-П и других. Компанией АО «ТерраТех» разработаны мониторинговые продукты для обеспечения лесной хозяйственной деятельности.

Мониторинговый продукт «Естественные и хозяйственные изменения в лесном фонде» будет полезен руководству лесничеств, природоохранным организациям и администрации субъектов РФ и других территориальных единиц. Данные о пожарах, ветровалах, воздействии вредителей и болезней могут быть предоставлены с любой периодичностью, вплоть до ежесуточной. Также продукт поможет выявить темпы и масштаб проведения санитарных рубок, работ по осветлению лесов и других действий.

Данное решение основано на основе мультитимеменных композитов космической съемки высокого пространственного разрешения выявляются изменения в площади лесных насаждений, на основе мультиспектральной съемки и подсчета вегетационных индексов выявляются изменения их качества. Полуавтоматическое выделение изменений позволяет описать объемы сплошных и выборочных рубок, прочистки просек и уходных мероприятий в границах интересующей территории.

Продукт «Незаконная деятельность в лесном фонде» позволяет провести предварительную оценку ущерба от нелегальных лесосек как на уровне отдельных лесничеств, так и на уровне субъектов РФ. Продукт реализован на основе мультитимеменных композитов космической съемки выполняется выделение лесосек, после сравнения с подгруженной информацией о разрешенных лесосеках и квартальной сети формируется выборка нарушений.

Разработанные АО «ТерраТех» мониторинговые информационные системы включают в себя материалы ДЗЗ среднего и высокого пространственного разрешения, полигональные слои, включающие данные о границах участка мониторинга, площади вырубок за пределами разрешенных участков, изменения их площади, аналитические отчеты, поставляемые с заданной периодичностью, включающие тематические карты нелегальных лесосек, таблицы и диаграммы [14].

Другим интересным примером является проект с международным сотрудничеством Финляндии и России «Инновационные информационные технологии для анализа негативного воздействия на леса приграничного региона» (InnoForestView), проводимый с 2014 по 2020 годы. Цель проекта состояла в создании новых возможностей

и передовых технологий для экологичной и чистой окружающей среды, устойчивого управления природой для граждан приграничных регионов [15].

В рамках проекта разработана информационная система, которая представляет собой достаточно универсальный инструментарий, позволяющий создавать удобные средства поддержки принятия решений в области лесопользования, экологии и охраны окружающей среды, а также мониторинга и прогнозирования негативных воздействий на окружающую среду. Информационная система должна была поддерживать анализ состояния приграничных лесов Финляндии и России, которые в том числе проходят в Ленинградской области, что является интересным примером для настоящей научно-исследовательской работы.

Для реализации поставленной цели проекта был осуществлен сбор данных ДЗЗ зарубежных группировок Sentinel-2, российских Ресурс-П и Канопус-В. Проведен сбор достоверных наземных данных и информации с БВС для разработки тематических карт.

К сожалению, отмечено, что более подробной методики, описанной для реализации данных проектов не найдено, скорее всего поскольку они являются коммерческими или в случае второго проекта грантовыми.

2 Методы мониторинга нарушенных лесных земель

2.1 Основные методы картографирования нарушенных лесных земель

2.1.1 Визуальное дешифрирование

Обнаружение рубок леса по космическим снимкам среднего и высокого разрешения не представляет большой сложности за исключением тех случаев, когда рубки занимают небольшую площадь или являются выборочными. Все нарушения лесного покрова (связанные как с лесозаготовками, так и с природными факторами) в целом имеют сходные спектральные дешифровочные признаки.

При частичном или полном нарушении (изъятии) лесного покрова происходит снижение коэффициента отражения в ближней ИК зоне спектра (NIR), обусловленное уменьшением содержания хлорофилла. В то же время растет коэффициент отражения в средней ИК зоне (SWIR), что обусловлено появлением участков открытой почвы. Также на вырубках происходит рост коэффициента отражения в видимой зоне спектра (в красном канале), за счет того, что открытая почва отражает больше солнечной радиации, чем кроны деревьев. Для дешифрирования рубок по снимкам, полученным в вегетационный период, наиболее информативным является средний ИК диапазон (SWIR) и различные варианты цветового синтеза с его участием, например синтез SWIR-NIR-RED. Но поскольку многие спутники или же данные имеют пониженное пространственное разрешение, то для дешифрирования часто используют также синтез NIR-RED-GREEN. Синтез естественных цветов используется реже, так как менее информативен и больше подвержен влиянию атмосферных эффектов [11].

Дешифрирование сплошных рубок не представляет затруднений по причине сильного контраста яркости в сравнении с лесными массивами. В синтезе каналов NIR-RED-GREEN свежие сплошные вырубки будут иметь голубовато-зеленый оттенок, а в синтезе SWIR-NIR-RED – ярко-розовый. Это позволяет надежно отделить их от зарастающих рубок, которые имеют ярко-зеленый цвет в синтезе каналов SWIR-NIR-RED за счет восстановления растительного покрова.

Прямым дешифровочным признаком рубок является их геометрически правильная форма, наиболее характерная для сплошных рубок. Для обнаружения проходных и особенно выборочных рубок леса пространственное разрешение снимков Landsat (30 м) недостаточно. При использовании снимков высокого разрешения (от 10 м) тип рубки определяется по рисунку изображения. При этом для успешного выявления выборочных рубок необходимо пространственное разрешение данных от 5 м [11].

Дешифрование в комбинации спектральных каналов False Colors (каналы Sentinel-2 B5, B4, B3) позволяет выявить ветровалы (буреломы). На изображениях растительность имеет оттенки фиолетового цвета, открытые участки земли отображены ярко салатovým цветом. Несмотря на сходство спектральных характеристик по цветовой гамме между ветровалами и вырубками имеется различие в характере формы таких участков. В отличие от вырубков, участки с ветровалами, вызванными сильным ветром, имеют не стандартную конфигурацию, главным признаком которых является полное отсутствие прямых углов [16].

Отличительной особенностью спектрального образа лесных гарей могут быть более низкие, чем у вырубков, коэффициенты отражения в видимой и в ближней ИК зонах спектра. Таким образом, в синтезе NIR-RED-GREEN гари будут иметь голубовато-зеленый оттенок, но будут значительно темнее вырубков, а в синтезе SWIRNIR-RED – темный розовый. Основным дешифровочным признаком, позволяющим выявить гари, является сложная геометрическая форма, как правило, вытянутая по направлению ветра [11].

2.1.2 Применение вегетационных индексов

В ряде исследований рассматриваются особенности использования индексных изображений при дешифрировании лесного покрова. После выполненных процедур коррекции и синтезирования каналов в истинных и псевдоцветах по снимкам без облачного покрова вычисляются индексы, визуализированные в виде индексных изображений: нормализованный вегетационный индекс (NDVI), почвенный вегетационный индекс (Soil-Adjusted Vegetation Index, SAVI), усовершенствованный индекс растительности (Enhanced vegetation index, EVI), индекс голой почвы (Barren Soil, BS) [17–22].

Индекс EVI представляет собой модификацию индекса NDVI, при этом имеет преимущества, поскольку влияние почвы и атмосферы в значениях данного индекса минимизировано. EVI позволяет оценивать состояние растений, как в условиях густого растительного покрова, так и в условиях разреженной растительности [23].

При определении нарушенности лесных экосистем от пожаров широко используются оценка по хлорофильному индексу – MTCI, индексу вегетации NDVI, нормализованному индексу гарей – NBR, коротковолновому индексу SWVI [24].

Нормализованный индекс гарей NBR (Normalized Burn Ratio) применяется для выявления гарей по двум разновременным снимкам, для его вычисления используются

ближний инфракрасный (NIR) и коротковолновый ИК (SWIR) каналы, поскольку, растительность сильно отражается в NIR спектре и слабо в SWIR, и наоборот, было выявлено, что следы нарушения лесного покрова, содержащие поврежденную землю, наилучшим образом отражаются в SWIR спектре, как показано на рисунке 1. NBR вычисляется по формуле:

$$NBR = \frac{(NIR - SWIR)}{(NIR + SWIR)} \quad (1)$$

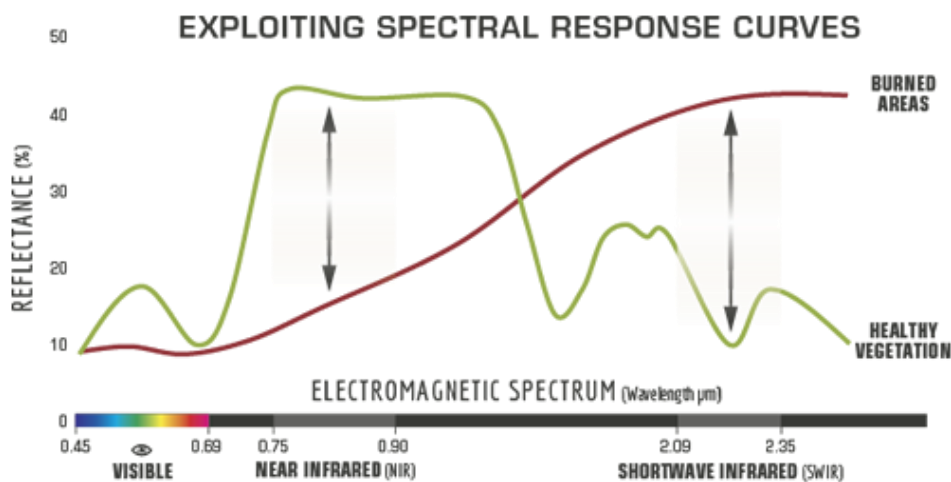


Рис. 1. Кривые спектральной яркости здоровой растительности (зеленым) и нарушенных участков (красным), отражательная способность которых сильно отличается в NIR и SWIR зонах электромагнитного спектра. Источник: [25]

Отмечено в ряде публикаций, чтобы корректно использовать классические индексы для обнаружения гарей, обычно необходимо применять маску для удаления водоемов и облаков. Водные объекты могут быть ошибочно приняты за гары по таким показателям, как NBR, поскольку пиксели имеют такие же или очень близкие к ним значения отражения, как и водные поверхности. Как правило, маска создается для удаления водоемов с использованием индекса нормализованной разности воды (NDWI), который позволяет отделить воду от суши. Как хорошо известно, вода обычно имеет сильное отражение в синем и зеленом спектре волн, поэтому исследователями предложена улучшенная версия NBR, учитывающую отражательную способность воды, названная как NBR+ и вычисляемая по формуле [26]:

$$NBR = \frac{(NIR - SWIR - GREEN - BLUE)}{(NIR + SWIR + GREEN + BLUE)} \quad (2)$$

Индекс NBR является наиболее мощным инструментом для лучшего понимания степени и тяжести пожара при использовании после расчета разницы между условиями до и после пожара. Разностный растр получают с помощью данных, полученных непосредственно перед пожаром, а затем сразу после пожара. NBR менее эффективен, если прошло время и началась регенерация растительности после пожара. С целью корректной переклассификации разностного раstra нарушенности леса на изучаемой территории выделяют валидационные участки [25]. Нормализованный индекс гарей NBR, также хорошо помогает отличить здоровый лес от нарушений лесного покрова в целом благодаря той же разнице отражения в ближнем инфракрасном и коротковолновом инфракрасном диапазонах спектра.

2.1.3 Методы автоматизированного дешифрирования

Методы классификации земельного покрова и землепользования эволюционировали за последние несколько лет. Сегодня существует множество методов классификации с использованием машинного обучения (machine learning, ML). Каждые несколько лет запускаются новые спутники с более высоким пространственным разрешением или даже с большим спектральным диапазоном. Это позволило еще больше модифицировать алгоритмы машинного обучения и разработать различные методы глубокого обучения. Применение алгоритмов машинного обучения повышает качество маскирования облачности на спутниковых снимках, что позволяет повысить точность картографирования нарушений лесного покрова [19].

Некоторые методы машинного обучения существуют уже более десяти лет и используются для задачи классификации земельного покрова (land use/ land cover, LULC). Одними из наиболее широко применяемых являются метод опорных векторов (SVM) и случайных лесов (Random forest, RF) [27].

Среди методов классификации земного покрова можно выделить два основных подхода: обучение с учителем – контролируемая классификация и обучение без учителя – кластерный анализ. Отличием второго является отсутствие необходимости разметки данных для обучения. Среди них можно выделить такие, как метод главных компонент (PCA), метод k-means дающий хорошие результаты, но зачастую для грубого разбиения на классы [17].

Контролируемая классификация реализует стратегию максимального правдоподобия Байеса для нормально распределенных векторов признаков в классах.

Оценки параметров классификатора (векторов средних и ковариационных матриц в классах) получаются на основе обучающих выборок [28].

Методы распознавания с учителем в большинстве современных работ реализованы с помощью свёрточных нейронных сетей, архитектура которых хорошо подходит для обработки изображений в различных задачах дистанционного зондирования, таких как обнаружение объектов и классификация земных покровов [17].

Алгоритм классификации с обучением – это процесс, при котором происходит сравнение значения яркости каждого пикселя с эталонами, которые задаются экспертом (оператором). В результате каждый пиксель относится к наиболее подходящему классу объектов. Классификацию с обучением можно применять, если заранее известно, какие объекты есть на снимке, на снимке выделяется небольшое количество классов и эти классы четко различаются на снимке.

Процесс классификации с обучением включает в себя несколько этапов:

- определение задач обработки снимка и выбор способа классификации;
- выбор эталонных участков;
- проведение классификации и оценка точности.

Термин точность (accuracy) обычно используется для выражения степени «правильности» карты или классификации. Тематическая карта, полученная с методами классификации может считаться точной, если она обеспечивает непредвзятое представление о земном покрове региона, который она изображает. Таким образом, точность классификации обычно понимается как степень, в которой полученная классификация изображений согласуется с реальностью или соответствует «истине» [22].

На оценку точности результатов дешифрирования влияет выбор метода классификации спутниковых изображений. Группой ученых из Австралийского Союза была разработана методика по улучшению точности управляемой тематической классификации спутниковых изображений, выполненной методом управляемой классификации «максимальное правдоподобие» (Maximum Likelihood Classifier, MLC) на основе спутниковых снимков Landsat, была применена пост-классификационная коррекция (Post-Classification Correction, PCC). Этот метод основан на использовании вспомогательных данных (карт, планшетов, полевых данных) и результатов, полученных в процессе исследования [29].

На сегодняшний день Random Forest (RF) считается одним из наиболее широко используемых и эффективных алгоритмов классификации земного покрова по данным дистанционного зондирования [19, 30]. Его преимуществами являются хорошая

обработка зашумленных наборов данных, хорошая производительность и легкость применения, поскольку для оптимизации требуется только два параметра (n_{tree} и m_{try}) по сравнению с другими методами машинного обучения, например SVM.

Алгоритм может быть использован для решения задач как классификации, так и регрессии. Случайный лес имеет тенденцию объединять сотни деревьев решений, а затем обучать каждое дерево решений на другой выборке наблюдений. Окончательные прогнозы случайного леса делаются путем усреднения прогнозов каждого отдельного дерева (рисунок 2) [31].

Результаты исследований показывают, что RF превосходит стандартные подходы к классификации, такие как простое дерево решений, поскольку позволяет усилить дифференциацию между различными категориями изучаемой области. RF относительно устойчив к уменьшению размера обучающей выборки и шума [32].

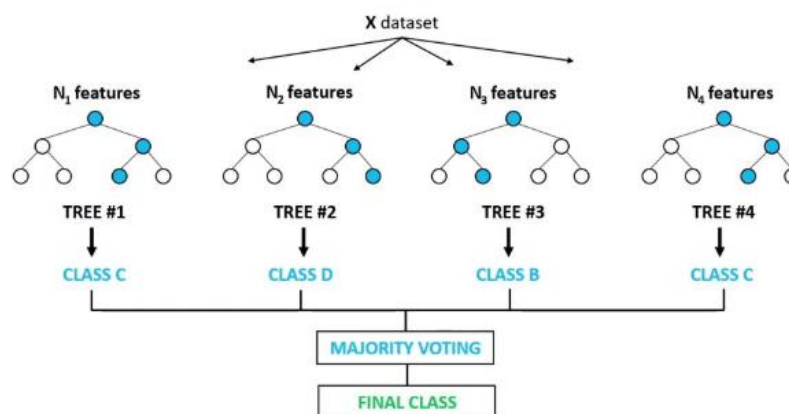


Рис. 2. Пример организации алгоритма случайного леса Random forest. Источник: [32]

Хотя существует другое мнение, например, исследователями из Загреба утверждается, что метод опорных векторов SVM является лучшим классификатором машинного обучения для классификации земного покрова и превосходит метод случайных лесов RF, CatBoost (CB). С другой стороны, SVM более медленный алгоритмом с точки зрения вычислительного времени, необходимого для классификации, а CB – самый быстрый метод машинного обучения [27].

При решении задач экологического мониторинга окружающей среды спецификой является, как правило, отсутствие полной и достоверной априорной информации о количестве, типе и спектральных характеристиках наблюдаемых объектов. В этом случае начальным этапом анализа съемочных данных является кластеризация – разделение пикселей изображения на группы (кластеры).

Критерием отнесения пикселей к тому или другому кластеру обычно служит подобие спектральных характеристик. В результате автоматически формируется список классов из статистически однородных групп пикселей [33].

Нарушения структур лесного покрова обнаруживаются при использовании кластерного анализа – классификации «без обучения» методом ISODATA (Iterative SelfOrganizing Data Analysis Technigue) по мультиспектральным снимкам [24].

Говоря о нарушении лесного покрова необходимо рассматривать возможность исследования восстановления лесов в рамках общей динамики роста леса. Алгоритм LandTrendr (Landsat-based detection of trends in disturbance and recovery) позволяет строить график кривой процесса нарушения и восстановления по снимкам миссии Landsat, но на самом деле не только. Он представляет собой метод временной сегментации, который обычно используется для отслеживания как постепенных, так и резких изменений.

Входными данными для алгоритма являются межгодовые временные ряды спектральных каналов или вегетационного индекса, пиксель за пикселем LandTrendr позволяет найти точки перегиба, в которых изменяется траектория значений пикселя по разновременным снимкам, и затем помещает данные в кусочно-линейную модель. В нем используется ежегодный сбор данных в один и тот же сезон с использованием медианных значений коэффициента отражения при минимальной облачности [20, 34].

К настоящему времени опубликовано сравнительно немного работ, где рассматривается применимость алгоритмов обработки изображений, основанных на машинном обучении, для обнаружения нарушений лесного покрова по спутниковым данным. В большинстве случаев алгоритмы сегментации спутниковых снимков средствами глубокого обучения направлены на решение таких задач, как маскирование облачности или классификация типов землепользования [19].

Существуют проблемы автоматизации, связанные с тем, что алгоритм нельзя настроить непосредственно на поиск «нарушенных» и «ненарушенных» земель, можно лишь научить его искать определенные объекты, например сигнал открытой поверхности земли, характерный для сплошных рубок. Однако такой сигнал возможен не только на вырубках, но и в местах схода лавин или там, где имеются открытые скалы среди ненарушенных редколесий. Поэтому необходима ручная проверка автоматически полученных результатов [35]. Также исследователями отмечается, что результаты дешифрирования космических снимков предполагают их полевую верификацию для создания библиотек образов природных объектов исследуемой территории и автоматизированного дешифрирования методами машинного обучения [18].

Развитие технологий машинного обучения, в том числе основанных на использовании глубоких нейро-сетевых моделей, позволяет осуществлять высокоточный автоматизированный мониторинг систем природопользования. Алгоритмы анализа больших геоданных и программные комплексы, функционирующие на их основе, становятся неотъемлемым компонентом инфраструктур пространственных данных. Глубокое машинное обучение (Deep Learning) позволяет снизить стоимость проводимых исследований за счет возможности точной интерполяции и экстраполяции измерений.

К моделям машинного обучения относят рекуррентные и сверточные нейронные сети, автокодировщики, модели глубокого машинного обучения с многослойной системой нелинейных фильтров для извлечения иерархических признаков из данных. При использовании методов и алгоритмов глубокого машинного обучения в анализе пространственных данных возникает ряд проблем. Несмотря на то что глубокие модели способны извлекать информативные признаки из многомерных данных, эффективность этого процесса в значительной мере зависит от большого количества обучающих выборок.

Отдельные пространственные объекты характеризуются разными размерами, спектральными яркостями, структурными особенностями, взаимным расположением по отношению к соседним объектам, что приводит к затруднению извлечения из них надежных и устойчивых признаков. Модели глубокого машинного обучения способны обучиться большому количеству признаков, однако они подвержены проблеме переобучения. Решением изложенных проблем может быть не только совершенствование архитектур моделей глубокого обучения, но и методов оптимального обогащения обучающих наборов данных с применением геосистемного подхода [36].

2.2 Исходные данные ДЗЗ и геоинформационная среда

2.2.1 Открытые данные ДЗЗ

Наиболее подходящими для регулярного мониторинга лесного фонда данными ДЗЗ из космоса являются снимки среднего разрешения (5–15 м), полученные опτικο-электронной аппаратурой в мультиспектральном режиме. Такого разрешения достаточно для выявления большинства нарушенных земель и негативных процессов, таких как вырубки, ветровалы, участки, пострадавшие от пожаров, а высокая повторяемость съемки (вплоть до еженедельной) и большая площадь снимков позволяют эффективно контролировать значительные площади лесных массивов на территории России [37].

Традиционно для мониторинга разнообразных природно-антропогенных явлений используются спутниковые изображения миссии Landsat Геологической службы США (USGS) с пространственным разрешением в видимом и ближнем ИК спектре 30 м/пикс. Большую популярность получил космический аппарат (КА) Sentinel-2 программы Copernicus Европейского космического агентства (ESA) с более высоким пространственным разрешением – 10 м/пикс. Постоянно пополняющийся архив данных обеспечивает пользователей высокоточной, актуальной и доступной информацией, а высокая производительность КА делает его высокоэффективным инструментом для обнаружения и оценки поврежденных участков на значительных по площади территориях лесных массивов.

При решении задач дистанционного мониторинга наземного покрова и отдельных объектов природопользования начинают также активно использоваться изображения высокого пространственного разрешения, полученные с отечественных спутников (Ресурс-П, Канопус-В), которые могут быть использованы для верификации данных, полученных с КА с средним пространственным разрешением [29]. При этом, стоит учитывать, что на данный момент в России спутниковые изображения как высокого и сверхвысокого, так и среднего разрешения не распространяются в открытом доступе.

Исследователями при обработке космических снимков, как правило, используется настольное программное обеспечение ГИС такие как ArcGIS, ENVI, ERDAS, SAGA GIS, специализированные модули QGIS, которые тем не менее имеют существенный недостаток.

Зачастую кроме доступности данных ДЗЗ при картографировании растительного покрова на большой территории возникают две основные проблемы: необходимость обработки «больших данных» (Big Data), или «больших пространственных данных», и доступность безоблачных изображений на большой территории. Обработка больших данных требует значительных объемов памяти и доступа к высокопроизводительным вычислениям.

Для решения задачи хранения и обработки больших пространственных данных широко применяются технологии облачных вычислений (cloud computing). В ходе научно-исследовательской работы было принято решение об использовании для получения и обработки спутниковых изображений с помощью платформы облачных вычислений Google Earth Engine (GEE), поскольку GEE позволяет решить наиболее серьезные проблемы, связанные с картографированием больших территорий.

Благодаря возможностям облачных вычислений пользователь может визуализировать требуемые изображения в графическом пользовательском интерфейсе,

не затрачивая при этом вычислительных мощностей своего компьютера, поскольку обработка данных осуществляется с использованием высокопроизводительных параллельных вычислений и больших дата-центров Google [23, 30, 38].

Пользователи могут анализировать коллекции изображений с помощью веб-редактора кода интегрированной среды разработки (IDE), не загружая эти данные на локальный компьютер. Среди доступных наборов данных в GEE наиболее часто используются данные с оптических спутников и, в частности, временные ряды Landsat продолжительностью более 40 лет, а также платформа располагает коллекциями разновременных снимков Sentinel-1, Sentinel-2, Sentinel-3, Sentinel-5P и MODIS [39].

2.2.2 Геоинформационная платформа Google Earth Engine для исследования нарушения лесных земель

Исследования лесов и растительности являются наиболее часто упоминаемыми задачами, решаемыми с использованием GEE [18, 21, 30, 40]. Выбор облачного геосервиса GEE может быть обусловлен необходимостью сокращения времени обработки большого массива данных и возможностью автоматизации одинаковых задач, с помощью скриптов на языках JavaScript или Python.

Исследователи из Университета Мэриленда использовали Google Earth Engine для разработки карты глобального изменения лесов (проект Global Forest Change), которая предоставляет подробную информацию о масштабах и причинах утраты и прироста лесов по всему миру. Карта ежегодно обновляется и находится в свободном доступе для использования исследователями и другими заинтересованными лицами. А также разработаны алгоритмы картографирования и мониторинга малонарушенных лесных территорий – диких, нетронутых человеком территорий, все так же используя вычислительные мощности GEE [41].

Одна из наиболее выгодных особенностей GEE заключается в том, что платформа может позволить пользователям легко сортировать и фильтровать данные ДЗЗ для конкретного КА для различных пространственных и временных масштабов. Другими словами, пользователи могут выбирать определенные изображения в «коллекции», используя определенные временные, пространственные фильтры, фильтры облачного покрова или другие критерии [21].

Клиентские библиотеки GEE, реализованные на JavaScript, представляют собой прокси-объекты для изображений (`ee.Image`), коллекций (`ee.ImageCollection`, `ee.FeatureCollection`) и других типов данных: числа (`ee.Number`), строки (`ee.String`),

векторные объекты с геометрией (`ee.Geometry`), списки (`ee.List`). Изображения, объединённые по смыслу, группируются в коллекции изображений (`ImageCollection`), которые имеют функционал выбора данных из коллекции путем фильтрации и сортировки по заданным пространственным, временным или другим критериям.

Библиотека API позволяет осуществлять операции над изображениями с использованием растровой алгебры и поддержкой функций более высокого порядка: `map()` и `iterate()`, которые позволяют применять произвольные функции к коллекциям изображений. Для вычисления статистики по коллекциям изображений применяется оператор свёртки `reduce()`, который, к примеру, может агрегировать данные методом скользящего окна [40].

Так же как и в настольной ГИС в GEE есть возможность вычисления вегетационных индексов на интересующую область. Общая методика построения карт вегетационных индексов включает в себя следующие этапы [23]:

- загрузка в каталог Assets файла границ (.shp) интересующей области либо создание простого полигона посредством встроенных функций;
- создание мозаики посредством функции `ee.ImageCollection`, применение фильтрации по периоду съемки и по границам выбранного полигона;
- вычисление растра вегетационных индексов по соответствующим каналам;
- выгрузка растров в формате GeoTiff в Google Drive.

Большой проблемой при исследовании поверхности земли по данным ДЗЗ из космоса, является облачность, так как облака часто могут закрывать довольно большую площадь поверхности, что препятствует качественному автоматическому анализу изображений. Классический метод устранения пикселей облаков предполагает создание маски, при этом пиксели восстанавливаются за счёт не покрытых облаками изображений, полученных в выбранный период съемки. Если безоблачных пикселей не найдется, будет возвращено изображение с пропущенными областями [33].

В GEE реализованы алгоритмы маскировки облаков, основным методом заполнения пробелов в облачных изображениях является использование медианных показателей (метод временной агрегации) [30]. В поисках способа наиболее эффективно устраняющего облачные пиксели, исследователи разрабатывают алгоритмы удаления облаков, например, основанные на обнаружении методом кластеризации k-средних [42]. Стоит отметить, что нет универсального алгоритма, подходящего для любого КА, поскольку сложно учесть все особенности, а некоторые алгоритмы строятся на определённых каналах, которые могут отсутствовать у других спутниковых систем.

Применительно к снимкам КА Sentinel-2, которые получили большую популярность в прикладных исследованиях с целью экологического мониторинга, разработан алгоритм командой исследователей из Google s2cloudless, позволяющий добиться наилучшего результата для получения чистого безоблачного композита. Для этого объединены алгоритм, разработанный Sentinel Hub, с вычислительными ресурсами Google. Набор данных s2cloudless предоставляет гибкий метод точной маскировки облачных пикселей на изображениях уровня 1С (TOA) и 2А (SR) для создания композитных изображений без облаков и выполнения процедур классификации.

В свою очередь алгоритм Sentinel Hub представляет собой маску облачности с использованием канала QA60, который по сути всего лишь двоичный классификатор для плотных облаков и перистых облаков. При использовании QA60 нет способа точно настроить, что считать облаком, а что нет. Изображение без облаков с алгоритмом s2cloudless обеспечивает вероятность присутствия облаков от 0 до 100 процентов, которую можно использовать для настройки агрессивности процедуры маскировки облаков. Данный алгоритм гибкий и настраиваемый, а также позволяет проводить маскировку облаков более точно и за меньшее время [43].

GEE также предоставляет различные вычислительные алгоритмы контролируемой и неконтролируемой классификации: от метода случайных лесов (RF) до алгоритма k-средних. В контексте обнаружения изменений в лесах наиболее часто упоминаемыми методами являются наложение изображений, анализ вектора изменений, ранжирование изображений и анализ основных компонент. Однако было высказано предположение, что при использовании платформы GEE для мониторинга лесного покрова, более эффективным подходом является просто использование разностного изображения вегетационных индексов, вычисленного за определенный интервал [21]. Данное утверждение в ходе научно-исследовательской работы было принято во внимание и сделан собственный вывод о результативности методов.

Одним из главных преимуществ платформы GEE является широкий набор данных (datasets), позволяющих проводить мониторинг изменения земель. В 2022 году компания Google представила совместно с Институтом мировых ресурсов разработку – Dynamic World (рисунок 3), которая представляет собой карту поверхности планеты, обновляемую в режиме реального времени. Сервис является открытым для всех заинтересованных, в Google считают, что инструмент будет полезен для анализа карты на предмет оценки воздействия стихийных бедствий на окружающую среду [44].

Dynamic World — это 10-метровый набор данных о землепользовании и земельном покрове (LULC) в режиме почти реального времени (NRT), который включает

вероятности классов и информацию о метках для девяти классов. Предсказанные классы Dynamic World доступны для коллекции Sentinel-2 L1C с 27 июня 2015 г. по настоящее время. Частота повторного обновления Sentinel-2 составляет от 2 до 5 дней в зависимости от широты. Используется маска для удаления облаков и теней от облаков с использованием комбинации S2 Cloud Probability, Cloud Displacement Index и Directional Distance Transform.

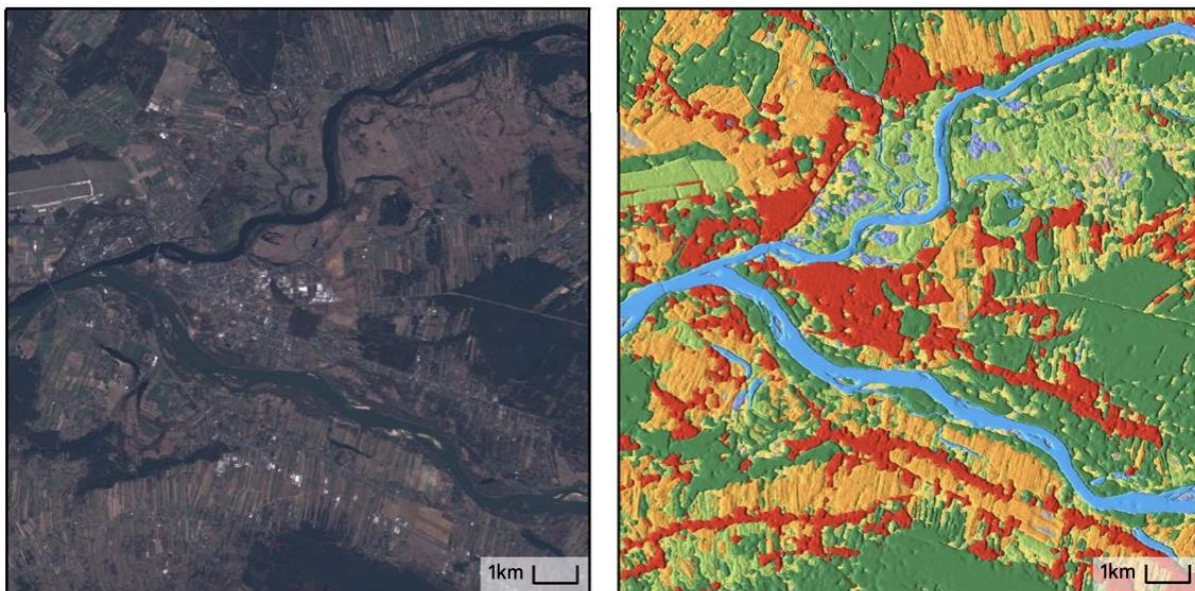


Рис. 3. Пример изображения Sentinel-2 и соответствующего ему продукта Dynamic World, апрель 2021. Источник: [45]

Учитывая, что Dynamic World был реализован совсем недавно и одновременно с настоящей научно-исследовательской работой, такой проект демонстрирует актуальность и важность разработки методов классификации с применением современных технологий на основе облачных вычислений и машинного обучения. При разработке методики в работе исследовано применение продукта Dynamic World для мониторинга и картографирования лесных нарушенных земель.

2.2.3 Нейронные сети для анализа данных ДЗЗ

Перспективным направлением исследования проблем мониторинга окружающей среды, в том числе мониторинга нарушенных земель, считается внедрение сверточных нейронных сетей для обработки данных ДЗЗ с целью автоматизации обнаружения объектов, учитывая пространственные закономерности вместе с спектральными характеристиками спутниковых изображений.

Алгоритмы машинного обучения не могут позволить извлечь информацию из изображений не только из спектральных значений в пикселях, но и из пространственных шаблонов пикселей, то есть пространственного контекста. Это становится проблемой для традиционных алгоритмов машинного обучения из-за проклятия размерности.

Если у нас есть изображение со многими каналами и мы хотим, чтобы машина «видела» большие участки пикселей, например 256x256, то количество входных признаков быстро становится слишком большим. Кроме того, алгоритм обучения все еще явно не учитывает пространственную структуру входных данных.

Полностью сверточные нейронные сети (fully convolutional neural networks, FCNN) предназначены для решения этих проблем. Еще в 2018 году Google Earth Engine начал интеграцию с современным машинным обучением на основе сверточных нейронных сетей, добавив экспорт и прием данных в формате обмена данными TensorFlow TFRecord, затем интегрировал Google Cloud для подключения Earth Engine напрямую к пользовательским моделям библиотеки TensorFlow, предназначенной для тренировки нейронных сетей с целью автоматического нахождения и классификации объектов [46].

3 Разработка методики картографирования нарушенных лесных земель с использованием Google Earth Engine

3.1 Методика картографирования нарушенных лесных земель с использованием Google Earth Engine

В ходе научно-исследовательской работы для разработки методики определения нарушенных лесных земель на территории Ленинградской области использованы снимки среднего разрешения из коллекции Sentinel-2 уровня 2A и платформа Google Earth Engine в качестве программного обеспечения. С целью получения и обработки данных спутниковой съемки с использованием платформы разработан скрипт на JavaScript на основе экземпляров, предоставленных в открытом доступе, в том числе открытой документации [47, 48].

В ходе работы рассмотрены и применены к объекту исследования методы, описанные в предыдущем разделе.

В рамках подбора методики ставились следующие задачи:

- Вычисление разности индексов NDVI и NBR по смежным вегетационным периодам (с шагом один год), сравнение результатов, представление выводов о результативности метода;
- Создание тренировочных данных и контрольных точек, реализация классификации с учителем, оценка точности модели;
- Определение разности лесного покрова по набору готовых данных Dynamic World с предсказанными классами и сравнение полученных результатов с результатами собственной модели.

Поскольку результаты применения вегетационных индексов показали себя недостаточно эффективными для автоматизированного дешифрирования нарушения лесного покрова (см. главу 3), в результате научно-исследовательской работы разработана методика мониторинга нарушенных лесных земель с применением классификации с учителем на основе машинного обучения, составлен скрипт на JavaScript, реализованный на GEE. Общая схема методики представлена на рисунке 4.

В начале скрипта загружены границы региона, созданы наборы `ee.FeatureCollection` для создания точек и тренировки модели классификации (см. Приложение X, Y). Заданы функции маскирования облачности на снимках Sentinel-2 по алгоритму `s2cloudless`, функции определения индекса NDVI по нормализованной разности каналов B8 и B4, индекса NBR по разности каналов B8 и B12. Это необходимо

для дальнейшего добавления вычисляемых индексов в коллекцию в качестве новых каналов и добавления их для обучения модели классификации.

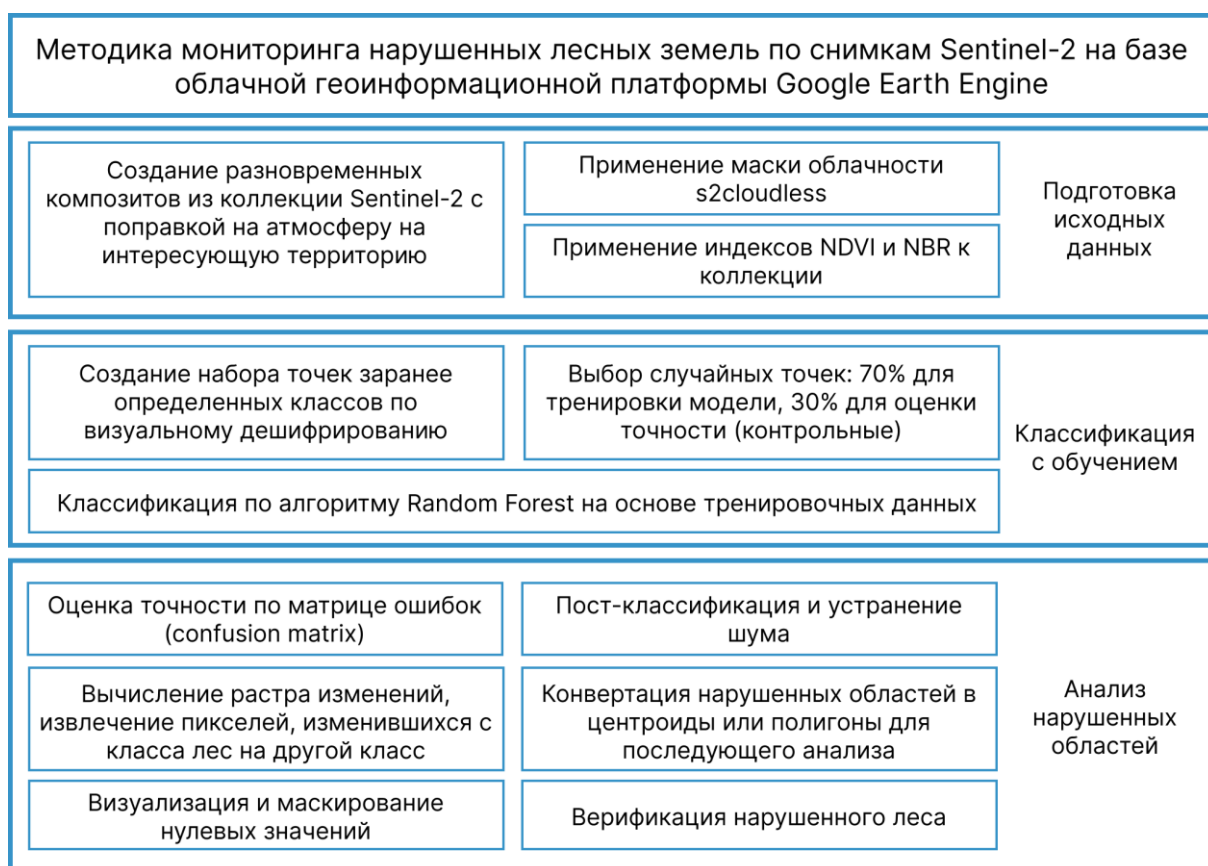


Рис. 4. Общая схема методики мониторинга нарушенных лесных земель, разрабатываемой в научно-исследовательской работе

Итеративно получены композиты из коллекции Sentinel-2 ('COPERNICUS/S2_SR_HARMONIZED'), отфильтрованы по границам региона и датам съемки с 2018 по 2022 годы. Выбран летний вегетативный период съемки с 1 июня по 1 сентября.

Дополнительно в скрипт добавлено получение раstra NDVI, а также результат вычисления разницы изменения лесного покрова на основе продукта Dynamic World (DW) – для сравнения результатов последующей классификации. Для этого выбирается класс лес (trees) из набора DW и вычисляется сколько пикселей классифицировано как древесная растительность с вероятностью выше 70% в заданный предшествующий период съемки, и с вероятностью меньше 50% в период после. Используются вероятности, так как набор данных DW представляет собой каналы с вероятностью классификации того или иного класса в диапазоне от 0 до 1, в сотых долях.

Для проведения классификации в рамках методики создан набор точек (меток) для пяти классов и присвоены целочисленные значения: 0 – вода, water (все постоянные

водные поверхности) , 1 – лес, forest (высокая древесная растительность), 2 – город, urban (урбанизированные территории), 3 – трава, grass (травянистая и низкая растительность), 4 – голая почва, bare (антропогенные и природные объекты без растительного покрова, характеризующиеся высокой отражаемостью в коротковолновой (SWIR) области спектра).

Для создания выборки тренировочных данных использован композит Sentinel-2 за летний период 2021 года. При выборе точек-экземпляров необходимо обращать на такие дешифровочные признаки как геометрическая форма вырубок, которая в некоторых случаях может быть воспринята как пашня или сельскохозяйственный луг, и восстановление леса, дешифрируемого в основном по серии разновременных композитов. Для разметки модели классификации подбирались участки местности, на которых представляется возможным визуальное обнаружение следов вырубок сравнением между собой двух разновременных композита, таких как на рисунке 5.

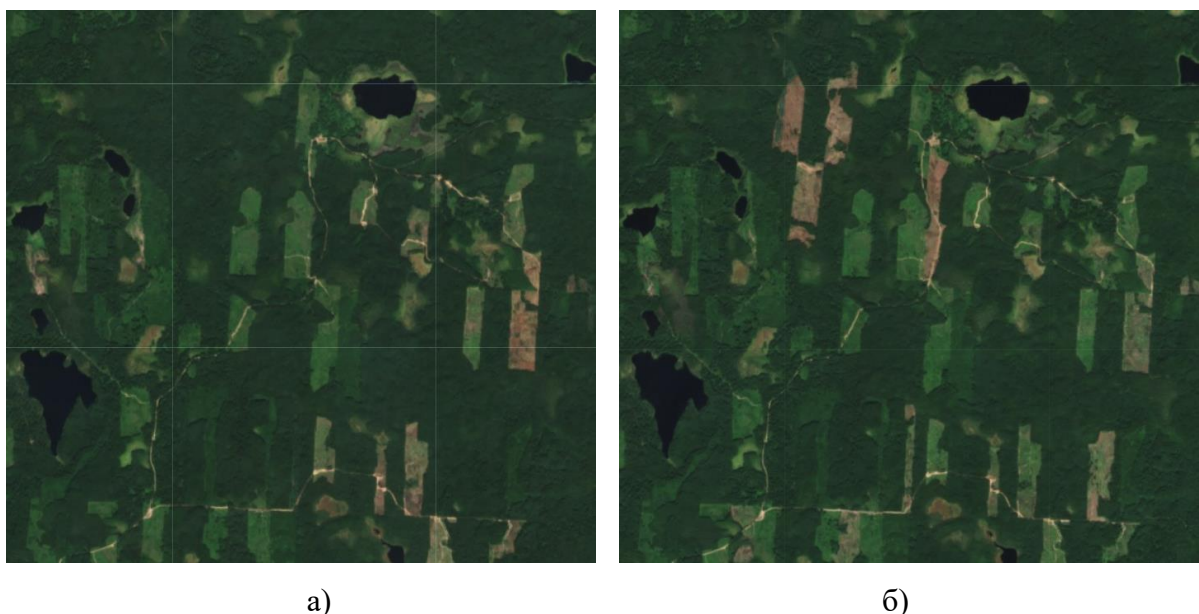


Рис. 5. Поиск эталонных участков вырубок для тренировки модели классификации: а) RGB композит 2021 года; б) RGB композит 2022

Все наборы точек объединены в общий набор. Создана выборка случайным образом из набора 70% точек для тренировки модели и 30% для оценки точности модели, иными словами, контрольные точки. Выбраны каналы, по которым проводится классификация: B3, B8, B12, NBR, NDVI. Выбран классификатор по алгоритму random forest с количеством деревьев – 100. Отмечено, что количество деревьев прямо пропорционально точности классификатора до достижения стабильного состояния (100 деревьев), в котором ошибка обобщения сходится при значениях ниже 10 %. Имеется в

виду, что как только ошибка сойдется, количество случайных величин (деревьев) лишь незначительно изменяет точность классификатора [32].

Для оценки точности модели применяется популярный метод получением матрицы ошибок (confusion matrix, или матрица путаницы) и общей точности. Отмечается важным обращать внимание на точность определения моделью класса лес по количеству контрольных точек, ошибочно определенных как лес, или наоборот, точек, которые должны представлять этот класс, но были отнесены к другому классу.

Из результата классификации извлекается класс, соответствующий лесу, он имеет целые значения 0 или 1, где 1 – лес. итеративно записываются изображения в текущий и предшествующий периоды, а затем вычисляется разность пикселей, которые изменились от класса лес к другому классу по формуле на языке JS, то есть разность которых равна единице:

$$\text{forestLoss} = \text{beforeImage.subtract(afterImage).eq}(1), \quad (3)$$

где: beforeImage – результат определения класса лес за предшествующий период, afterImage – результат определения класса лес за текущий период.

Для устранения шума применяется пост-классификация по методу объединения пикселей функцией ee.Image.connectedPixelCount(), который позволяет отбросить пиксели, являющиеся выбросами или незначительными изменениями, с помощью фильтрации по заданному максимальному размеру окрестности в пикселях.

Конечным результатом являются растры изменения леса к нарушенному состоянию с маскированными нулевыми значениями, на основе которых возможно составлять карты нарушения лесов. Практически любая классификация не идеально точна, поэтому необходима верификация полученных результатов по композитам. Также реализуется конвертация участков нарушений (вырубок и гарей) из растра в вектор, выгрузка геометрий для дальнейшего анализа и верификации в том числе в сторонних ГИС с использованием снимков высокого разрешения или АФС.

3.2 Результаты исследовательской работы

3.2.1 Сравнение вегетационных индексов NDVI и NBR

Такие инструменты анализа, как индексы NDVI или NBR, способны помочь в поиске свежих вырубок, поскольку данные методы основаны на изменении значения спектральной яркости здоровой лесной растительности и мест вырубок. Благодаря разности в ближнем инфракрасном и коротковолновом инфракрасном диапазонах

электромагнитного спектра возможно найти как проходные промышленные вырубки, так и участки, пройденные пожарами. Результаты вычисления растров NBR и NDVI, полученные по данным Sentinel-2 и GEE, представлены на рисунках 6 и 7 соответственно.



Рис. 6. Растр разности индекса NBR, вычисленного за 2021–2022 годы: красным показаны свежие вырубки, светло-зеленым – участки, где происходит восстановление, зеленым – области без изменений

Некоторые исследования сосредоточены на определении того, какой индекс работает лучше, хотя результаты все еще неоднозначны, поэтому в магистерском исследовании применялись оба индекса. Как видно из анализа сравнения полученных карт индексов, оба вегетационных индекса достаточно точно позволяют определить площади нарушенного вырубками леса. Проведенный опыт вычисления разностного индексного изображения NBR, полученного за два смежных вегетационных периода и классифицированного по значениям таблицы 1, показал действительно хорошие результаты.

Таблица 1. Таблица классификации значения пикселей разности NBR

Степень повреждения леса	Значения разности NBR
Восстановление	< -0.1
Не поврежден	-0.1 – 0.1
Средняя степень	0.1 – 0.66
Высокая степень, в т.ч. пожары	> 0.66

Однако эти данные не могут составить в основу автоматизированного мониторинга. Все еще карта разности NDVI или NBR не может быть использована в качестве конечного результата анализа и нуждается в дополнительной интерпретации из-за ложноположительного срабатывания. Проще говоря, необходимо дополнительно визуально дешифровать зоны изменений индекса, поскольку некоторые зоны предполагаемых вырубок могут оказаться, например, полями сельскохозяйственного назначения, на которых в разные периоды съемки может быть собран урожай или наоборот, быть на пике вегетационного периода. А поскольку довольно сложно подобрать снимки без облаков, которые бы составили мозаику, за один и тот же период вегетации, при чем на всех полях одновременно, проблематично, то и нельзя исключить данный фактор из результатов.

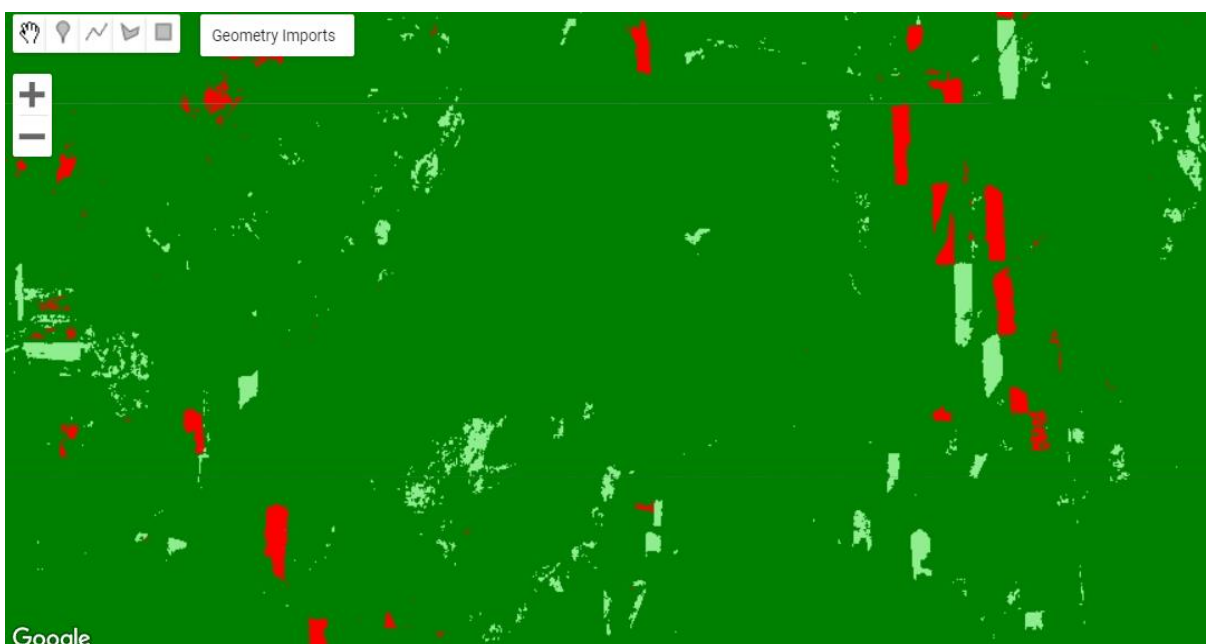


Рис. 7. Растр разности индекса NBR, вычисленного за 2021–2022 годы: красным показаны свежие вырубки, светло-зеленым – участки, где происходит восстановление, зеленым – области без изменений

Но вместе с тем стоит отметить, что растры разности индексов могут быть использованы для целей разработки набора тренировочных данных, поэтому они были добавлены для реализации классификации с обучением. Растры индексов позволяют проводить поиск мест рубок при создании набора тренировочных данных наряду с некоторыми дешифровочными признаками, характеризующимися в первую очередь правильной геометрической формой.

В результате сравнения полученных разностных изображений индексов NDVI и NBR, визуально определено, что их точность практически совпадает. Но и ошибочно

определенные области, которые на самом деле не являются лесом, практически присутствуют на обоих растрах разности вегетационных индексов. Можно отметить, что NDVI определяет немного меньше ошибочно нарушенных лесов.

В одном из исследований, посвященном автоматизированному обнаружению изменений лесного покрова, проведен анализ на основе полученных разностных изображений по обоим индексам и сделан вывод о том, что решение о применении одного из индексов может зависеть от типа оцениваемого нарушения и типа анализируемой растительности [49]. А поскольку индексов существует гораздо больше, может быть проведено исследование о поиске наилучшего индекса для определения нарушения леса, однако, в ходе магистерской работы было установлено, что в этом нет необходимости, поэтому сделан упор на проведение классификации на основе тренировочных данных с применением машинного обучения.

3.2.2 Классификация с учителем с использованием машинного обучения

В данном разделе представлены результаты разработанной методики с применением контролируемой классификации и машинного обучения с использованием геоинформационной платформы Google Earth Engine.

Создан набор данных точек-экземпляров, которые характеризуют наилучшим образом определяемый машинным методом класс, примеры таких меток представлены на рисунках 8 и 9. В каждом наборе класса выбиралось по 50–70 точек.

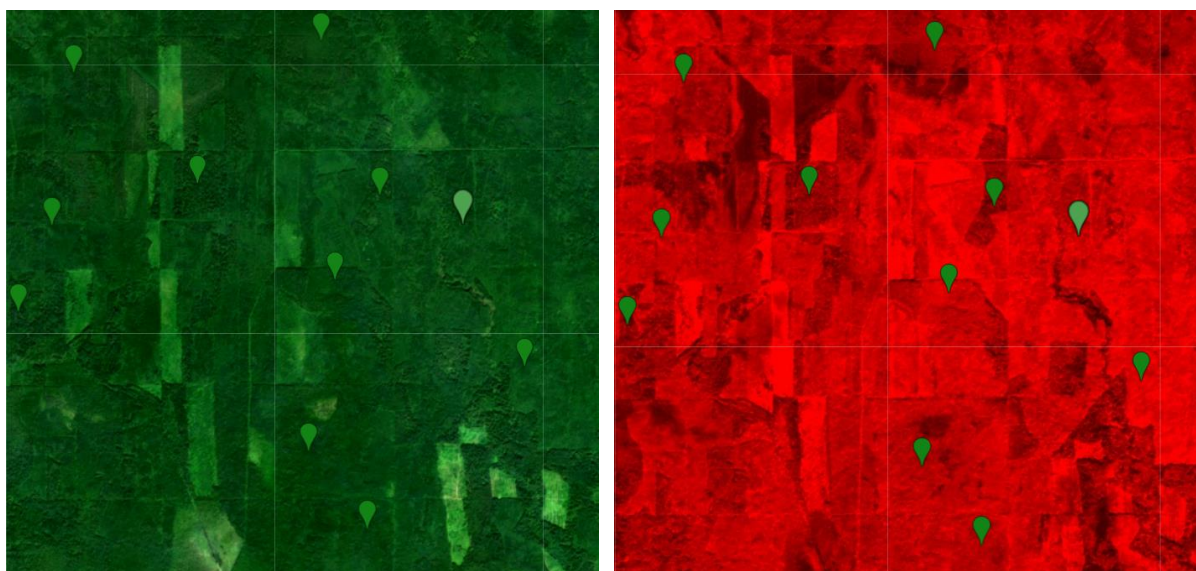


Рис. 8. Тестовые участки класса лес (forest) для обучения модели классификации на подложке композита 2021 год в естественных цветах – слева и NIR-спектре – справа

Выбраны пять классов – водная поверхность, лесная растительность, городская застройка, травянистая растительность и оголенная почва. Класс воды и городской застройки включен для того, чтобы исключить эти территории из анализа в дальнейшем, почвы выбраны исходя из понимания, что свежие вырубki характеризуются нарушенным почвенным покровом, а травянистая растительность – поскольку на местах нарушения может также проходить быстрая регенерация растительности. При создании композитов используются медианные значения пикселей, которые соответствуют всему летнему периоду. Особое внимание уделено выбору меток, соответствующих классу лесной растительности, а также классу голая почва, в который входят места свежих вырубок, характеризующихся отчетливо дешифрируемыми характеристиками отражения, их возможно верифицировать по композитам за текущий и предшествующий периоды.

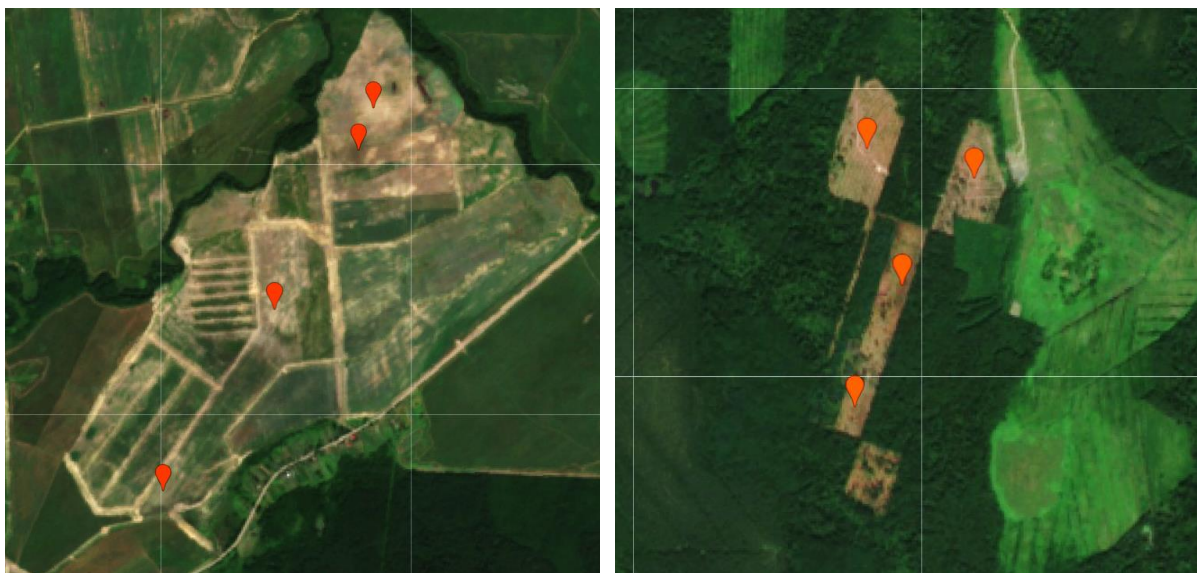


Рис. 9. Тестовые участки класса голая почва (bare) для обучения модели классификации. Используются как места свежих вырубок – справа, так и распаханых полей для более точного обучения модели

По набору тренировочных точек построены средние спектральные сигнатуры с целью подбора каналов для классификации, которые представлены на рисунке 10. На полученном графике видно, что в некоторых каналах значения яркости объектов отличаются между собой (между классами) гораздо отчетливее. Исходя из этого были выбраны для тренировки модели каналы В3, В8, В12. Каналы В2 и В4 не были выбраны, поскольку классы воды и лесной растительности очень близки по своим значениям, и чтобы избежать некорректного определения класса.

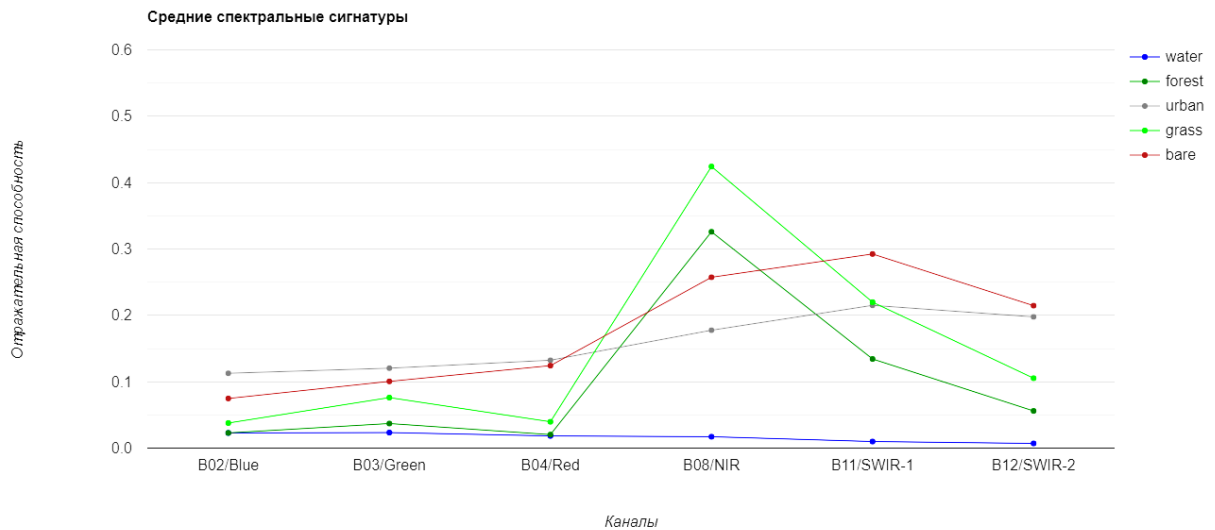


Рис. 10. Средние спектральные сигнатуры классов на основе набора тренировочных данных

Результат классификации по алгоритму случайного леса random forest с количеством деревьев 100 удалось оценить с помощью матрицы ошибок. Матрица ошибок классификации композита 2021 года, по которому и проводилось обучение, представлена на рисунке 11. Для оценки точности использовалось 30% от всех выбранных точек-экземпляров. Полученная общая точность свидетельствует о том, что выборка достаточно однородна по своему составу.

		Классификация (предсказано)					
		0	1	2	3	4	
Действительные классы	0	14	0	0	0	0	1,00
	1	0	19	0	0	0	1,00
	2	0	0	10	0	1	0,90
	3	0	0	0	20	0	1,00
	4	0	0	0	0	14	1,00
		1,00	1,00	1,00	1,00	0,93	0,987
		Точность производителя					Общая точность

Рис. 11. Матрица ошибок (confusion matrix), характеризующая точность модели классификации

Таким образом, тренировочная модель экстраполируется на композиты других периодов съемки и позволяет определить земли, покрытые лесом, а затем сравнить, то есть вычислить разность результатов классификации класса лес и определить области сокращения лесного покрова.

3.2.3 Результаты модели классификации

Для оценки результатов модели классификации получены композиты за 2018–2022 годы, по которым проведена классификация и составлены растры нарушения лесов. Примеры составленных на основе разработанной методики карт нарушения лесов представлены в Приложении А, Б и В. В основном удалось обнаружить зоны сплошных и проходных вырубок, а также просек, очень редко гари. При оценке результатов проводились:

- изучение влияния алгоритмов маскирования облачности;
- верификация зон нарушения леса по композитам, на основе которых и проводилась классификация;
- сравнение результатов классификации с результатами, полученными при использовании Dynamic World.

Облачность влияла резко отрицательно на результаты классификации, из-за чего срабатывало ложноположительное определение мест нарушения леса (рисунок 12).

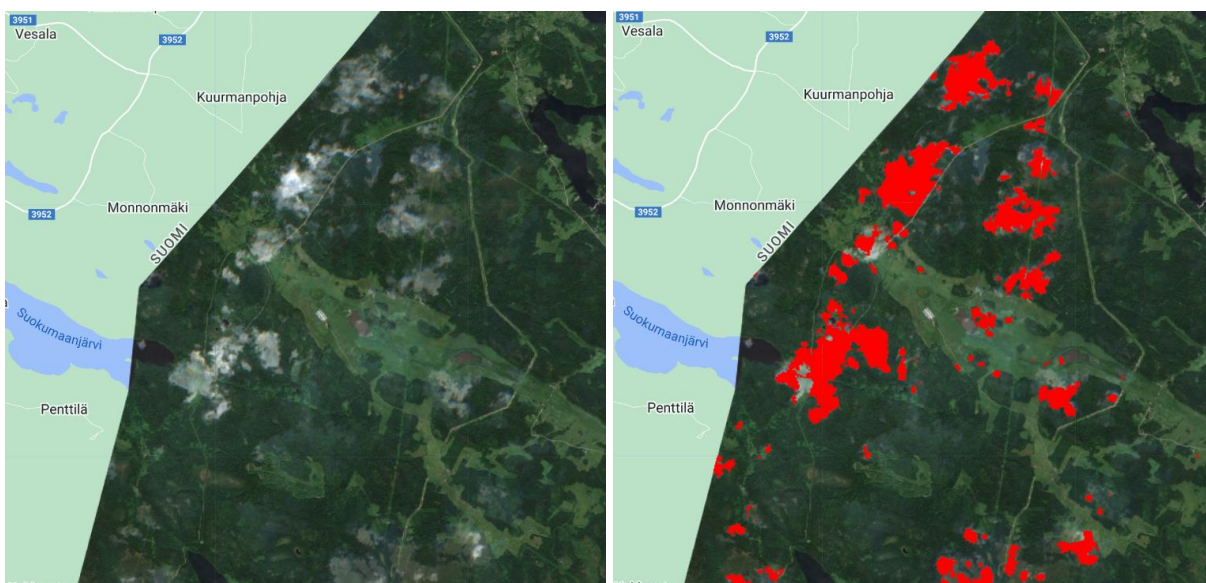


Рис. 12. Ложноположительное срабатывание модели классификации из-за некачественной маскировки облаков (неверное определение мест нарушения леса)

При использовании базовой маски облачности, разработанной Sentinel Hub, и основанной на извлечении облаков по каналу QA60, облака не были удалены полноценно, либо пиксели, которые были заменены с искаженным значением, которое моделью в итоге не были определены как лес. В результате качественного результата

удалось добиться при использовании маски облачности по алгоритму s2cloudless, описанному в предыдущем разделе.

Для верификации областей нарушения лесных земель выбраны несколько областей интереса. Поскольку в Ленинградской области находится более 55 особо охраняемых природных территории (ООПТ), для исследования были дополнительно загружены непосредственно на платформу GEE векторные границы ООПТ области из открытого картографического источника OpenStreetMap.

Например, результаты классификации позволили обнаружить нарушение лесного покрова на территории заказника «Анисимовские озера», произошедшее за 2021–2022 годы, что представлено на рисунке 13.

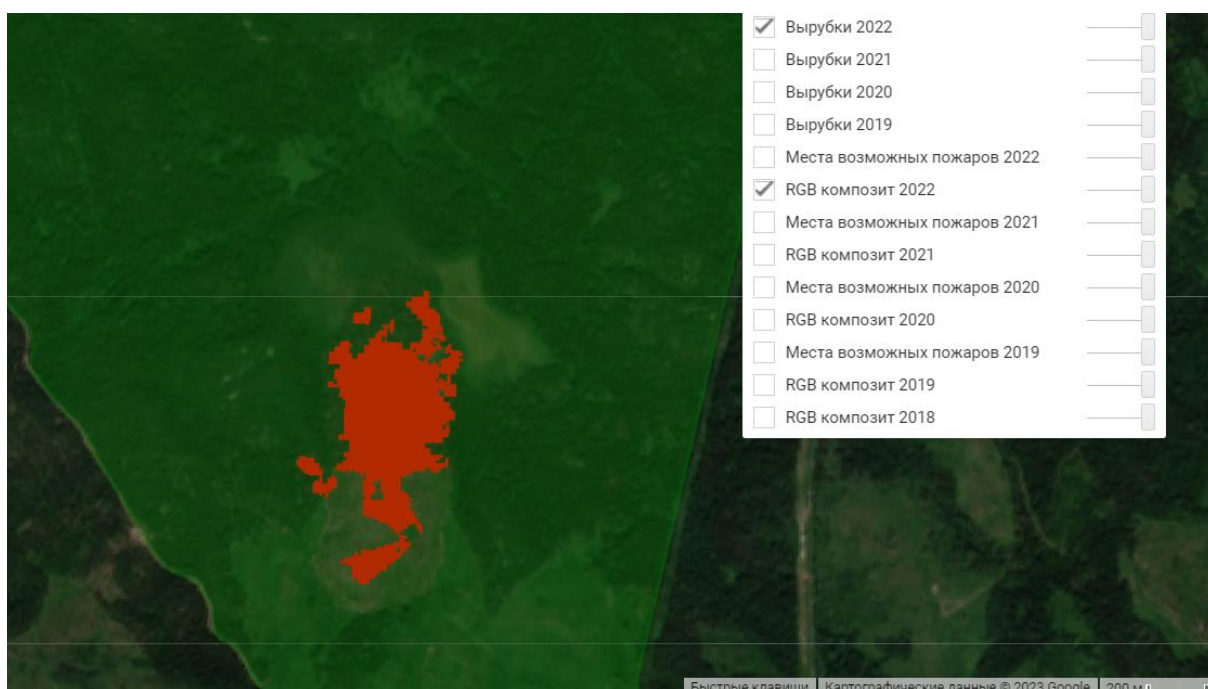


Рис. 13. Выявленный участок гари (вероятно вместе с вырубкой) на территории ООПТ заказника «Анисимовские озера» по данным классификации (веб-интерфейс GEE)

Визуальное дешифрирование на основе композитов (рисунок 14) за соответствующие годы позволяет сделать вывод, что на территории ООПТ произошел пожар, поскольку форма нарушенной области не имеет правильных очертаний и имеет вытянутую форму, скорее всего, по направлению ветра. Также на гарь указывает темный оттенок нарушенной почвы и неполное снятие леса, как бывает при вырубках.

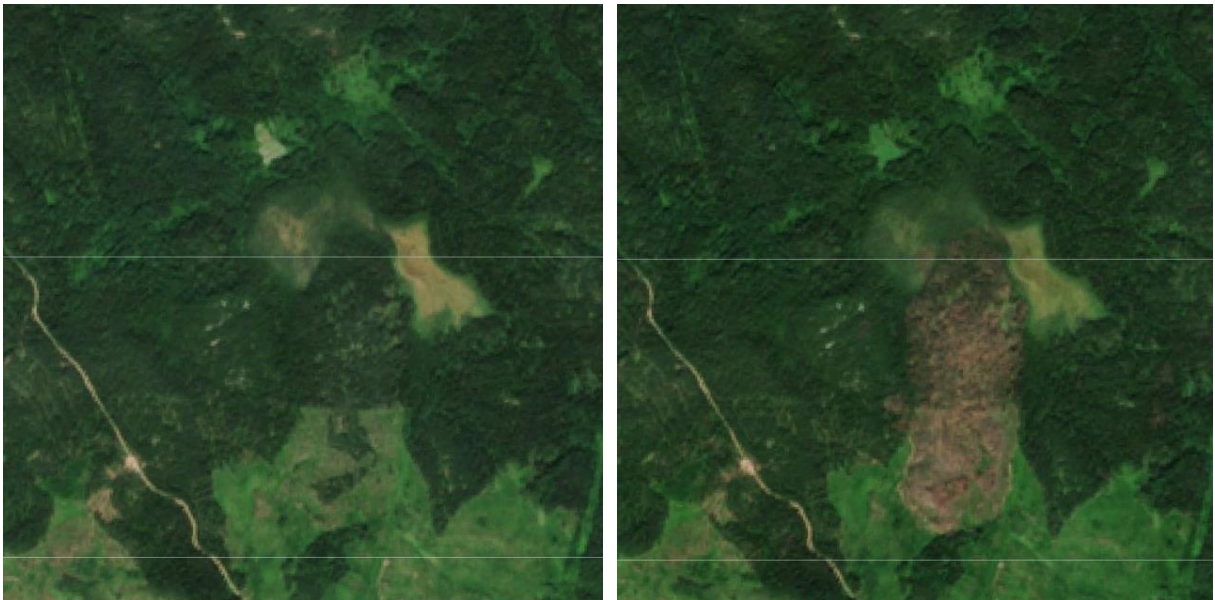


Рис. 14. Композит в естественных цветах до и после нарушения леса в заказнике «Анисимовские озера» за 2021 и 2022 годы соответственно

Полученная модель классификации по снимкам Sentinel-2 может служить основой мониторинга нарушенных земель на территории ООПТ с целью информирования об экологической обстановке в Ленинградской области.

Также выбран для исследования крупный промышленный объект – комплекс по переработке этансодержащего газа и производству сжиженного природного газа, расположенный на западе области в районе поселка Усть-Луга, строительство которого было начато в 2021 году. При строительстве была нарушена довольно большая площадь леса, около 900 гектар (рисунок 15).

По составленной карте нарушения леса, представленной в Приложении Б, можно проводить анализ вырубki леса поэтапно. Данные полученные в результате мониторинга, например, могут позволить провести оценку экологического ущерба от строительства промышленного комплекса.



Рис. 15. Фрагменты композитов в естественных цветах строительства газохимического комплекса в районе поселка Усть-Луга за 2018 и 2022 годы соответственно

В исследовательской работе рассмотрено применение данных набора Dynamic World, доступного на платформе GEE. Вычислены растры, которые представляют собой области пикселей, изменившихся с лесной растительности на нарушенный ландшафт. Результаты также представлены за выбранный период с 2018 по 2022 годы.

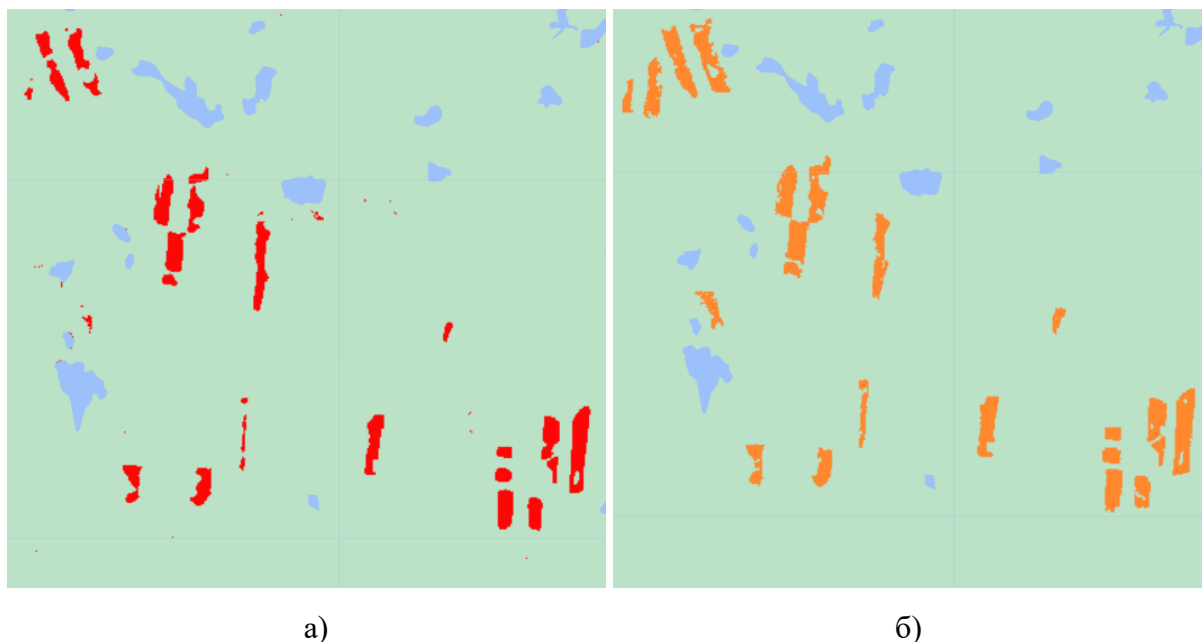


Рис. 16. Результаты определения нарушения леса за 2021–2022 годы с использованием:
а) Dynamic World; б) обученной модели классификации

Сравнивая результат, полученный с использованием DW (рисунок 16) и собственной классификацией на основе размеченных данных, сделан вывод, что набор

данных DW является мощным инструментом для мониторинга нарушенных земель и может быть применен для любого региона России и мира, способным практически безошибочно определять места сокращения леса.

Хотя все-таки было отмечено, что границы по данным набора DW не совсем точно определены и отличаются более сглаженными формами, скорее всего из-за алгоритмов постобработки результатов классификации, выполняемой для отображения результатов DW. Полученные результаты, основанные на разработанной модели, позволили достаточно четко определить границы вырубок и гарей, что дешифрируется по композитам, хотя и нуждаются в окончательной верификации с целью уточнения типа нарушения.

Заключение

В результате проделанной за два года магистерской научно-исследовательской работы, посвященной разработке методики автоматизированного картографирования нарушенных лесных земель:

- проанализирован существующий опыт и методы картографирования нарушения лесных земель;
- проведен опыт и составлена технологическая схема разработанной методики;
- разработан скрипт на языке JavaScript, реализующий методику на базе облачной геоинформационной платформы Google Earth Engine;
- составлены картографические материалы, приложенные к научно-исследовательской работе.

Можно заключить, что классификация на основе машинного обучения в сочетании с Google Earth Engine является мощным инструментом для мониторинга нарушенных лесов. Используя спутниковые снимки и современные алгоритмы машинного обучения, становится возможным автоматически обнаруживать и классифицировать различные типы нарушения лесов, такие как вырубки и лесные пожары.

Применение облачных технологий позволяет анализировать большие наборы данных спутниковых изображений без необходимости загружать на стационарный ПК, что позволяет выявлять и отслеживать нарушения лесов на обширных территориях и в течение продолжительных периодов времени в сжатые сроки, используя распределенные вычислительные мощности. Автоматизация процесса обнаружения и картографирования нарушенных лесных земель упрощает мониторинг, экономя время и ресурсы по сравнению с традиционными методами обработки данных ДЗЗ.

Сочетание этих технологий облегчает создание подробных карт и отчетов о нарушениях лесов, обеспечивая пространственной информацией и поддерживая принятие решений на основе фактических данных для государственных органов власти и органов муниципального самоуправления, исследователей и природоохранных организаций.

Развитие исследования может быть продолжено по направлению обработки данных ДЗЗ, предоставляемых GEE, с применением сверточных нейронных сетей на основе библиотеки TensorFlow, которая предоставляет возможности поиска нарушенных участков леса с большей достоверностью.

Список литературы

1. Ленинградская область – Большая российская энциклопедия [Электронный ресурс]. URL: <https://old.bigenc.ru/geography/text/5718992>.
2. Постановление Правительства РФ от 10 июля 2018 г. N 800 “О проведении рекультивации и консервации земель” [Электронный ресурс]. URL: <https://base.garant.ru/71985800/> (Дата обращения: 18.05.2023).
3. Росреестр Государственный мониторинг земель [Электронный ресурс]. URL: <https://rosreestr.gov.ru/activity/gosudarstvennoe-upravlenie-v-sfere-ispolzovaniya-i-okhrany-zemel/gosudarstvennyu-monitoring-zemel/> (Дата обращения: 18.05.2023).
4. Баянова А. А. Региональные аспекты государственного мониторинга земель // Материалы международной конференции «Agritech-III-2020: Агробизнес, экологический инжиниринг и биотехнологии. – 2020. – 52030с.
5. “Лесной кодекс Российской Федерации” от 04.12.2006 N 200-ФЗ [Электронный ресурс]. URL: https://www.consultant.ru/document/cons_doc_LAW_64299/ee7af8f2c965ebfa961cd92f3446278b87d7678d/#dst834 (Дата обращения: 12.04.2023).
6. Юдин Е.А. Возможности использования ГИС-технологий при мониторинге нарушенных земель // Роль молодых ученых в решении актуальных задач АПК. – 2017. – 245-249с.
7. Лурье И.К. Геоинформационное картографирование: учебник / Лурье И.К. // М.: КДУ – 2008. – С.422.
8. Верхотуров А.А. Разработка систем регионального мониторинга земель на основе атласного картографирования / Верхотуров А.А., Мелкий В.А. // Известия Томского политехнического университета. Инжиниринг георесурсов – 2016. – Т. 327 – № 7 – С.66–83.
9. Верхотуров А.А. Мониторинг нарушенных земель территории Новиковского бурогольного месторождения (юго-восток острова Сахалин) / Верхотуров А.А., Попова А.А. // Регулирование земельно-имущественных отношений в России: правовое и геопространственное обеспечение, оценка недвижимости, экология, технологические решения – 2021. – № 1 – С.175–180.
10. Mahapatra M. Mapping and monitoring of land use and land cover changes using Remote Sensing and GIS techniques / Mahapatra M., Ramakrishnan R., Rajawat A.S. // International Journal of Geomatics and Geosciences – 2013. – Т. 4 – № 1 – С.242–248.
11. Шихов А.Н. Тематическое дешифрирование и интерпретация космических

снимков среднего и высокого пространственного разрешения: учебное пособие/Пермский государственный национально-исследовательский университет / Шихов А.Н., Герасимов А.П., Пономарчук А.И., Перминова Е.С. // Пермский государственный национальный исследовательский университет – 2020.

12. Рябов Ю. В. Разработка геоинформационной системы мониторинга земель, нарушенных несанкционированными свалками : автореферат дис. ... кандидата географических наук : 25.00.26 / Рябов Юрий Владимирович. – Санкт-Петербург, 2013. – 19 с.

13. Стыщенко Ф. В. Разработка и применение методики и автоматизированной технологии оценки пирогенной гибели лесов на основе спутниковых данных : автореферат дис. ... кандидата тех. наук : 25.00.34 / Стыщенко Федор Викторович – 2016.

14. Цифровая Земля – Уникальные сервисы на основе данных дистанционного зондирования Земли из космоса [Электронный ресурс]. URL: https://dgearth.ru/product/forest_m_changes/ (Дата обращения: 15.04.2023).

15. Zelentsov V. A. Information system for analyzing negative impacts on forests of the border regions / IOP Conference Series: Earth and Environmental Science, 2021. – 12001с.

16. Мейсунова А.Ф. Оценка влияния антропогенных и природных факторов на состояние лесов с помощью данных дистанционного зондирования в Старицком лесничестве Тверской области / Мейсунова А.Ф., Сметанина Н.Ю. – 2021.

17. Дурандин Д.П. Сегментация типов местности на спутниковых снимках / Дурандин Д.П., Сизов Н.А., Сорокин В.А., Востротина А.В., Омаров Р.З. // Молодой ученый – 2019. – № 26 – С.30–33.

18. Раджабова Р.Т. Использование индексных изображений при дешифрировании растительного покрова Внутригорного Дагестана / Раджабова Р.Т., Алексеенко Н.А., Курамагомедов Б.М., Тажудинова З.Ш., Султанов З.М. // Юг России: экология, развитие – 2020. – № 4 (57) – С.126 –136.

19. Тарасов А.В. Современные методы оперативного картографирования нарушений лесного покрова / Тарасов А.В. // Вестник СГУГиТ (Сибирского государственного университета геосистем и технологий) – 2020. – Т. 25 – № 3 – С.201–213.

20. Meng Y. How can spatial structural metrics improve the accuracy of forest disturbance and recovery detection using dense Landsat time series? / Meng Y., Liu X., Wang Z., Ding C., Zhu L. // Ecological Indicators – 2021. – Т. 132 – С.108336.

21. Jahromi M.N. Google Earth Engine and its application in forest sciences / Jahromi M.N., Jahromi M.N., Zolghadr-Asli B., Pourghasemi H.R., Alavipanah S.K. // Spatial

Modeling in Forest Resources Management: Rural Livelihood and Sustainable Development – 2021. – С.629–649.

22. Nipunajith G. Evaluating Land Use / Land Cover changes in Kundasale DS division using remote sensing & GIS techniques / Nipunajith G. – 2020.

23. Чульдум А.Ф. Получение вегетационных индексов Тувы на платформе Google Earth Engine / Чульдум А.Ф. // Природные ресурсы, среда и общество – 2021. – № 4 – С.51–58.

24. Рожков Ю.Ф. Оценка нарушенности лесных экосистем и их восстановления после пожаров в Олекминском заповеднике (Россия) по космическим снимкам Landsat / Рожков Ю.Ф., Кондакова М.Ю. // Природные ресурсы Арктики и Субарктики – 2021. – Т. 26 – № 2 – С.94–107.

25. Earthlab Work with the Difference Normalized Burn Index [Электронный ресурс]. URL: <https://www.earthdatascience.org/courses/earth-analytics/multispectral-remote-sensing-modis/normalized-burn-index-dnbr/> (accessed: 16.05.2023).

26. Alcaras E. Normalized Burn Ratio Plus (NBR+): A New Index for Sentinel-2 Imagery / Alcaras E., Costantino D., Guastaferrero F., Parente C., Pepe M. // Remote Sensing – 2022. – Т. 14 – № 7.

27. Rumora L. Impact of various atmospheric corrections on sentinel-2 land cover classification accuracy using machine learning classifiers / Rumora L., Miler M., Medak D. // ISPRS International Journal of Geo-Information – 2020. – Т. 9 – № 4 – С.277.

28. Бучнев А.А. Классификация гиперспектральных данных дистанционного зондирования Земли / Бучнев А.А., Пяткин В.П. // Журнал Сибирского федерального университета. Техника и технологии – 2019. – Т. 12 – № 5 – С.536–541.

29. Воробьев О.Н. Пространственно-временной анализ динамики лесного покрова в Среднем Поволжье по спутниковым данным / Воробьев О.Н., Курбанов Э.А., Полевщикова Ю.А., Лежнин С.А. – 2019.

30. Phan T.N. Land cover classification using Google Earth Engine and random forest classifier—The role of image composition / Phan T.N., Kuch V., Lehnert L.W. // Remote Sensing – 2020. – Т. 12 – № 15 – С.2411.

31. Random Forest Classifier Tutorial: How to Use Tree-Based Algorithms for Machine Learning [Электронный ресурс]. URL: <https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/> (Дата обращения: 13.05.2023).

32. Rodriguez-Galiano V.F. An assessment of the effectiveness of a random forest classifier for land-cover classification / Rodriguez-Galiano V.F., Ghimire B., Rogan J., Chica-Olmo M., Rigol-Sanchez J.P. // ISPRS journal of photogrammetry and remote sensing – 2012.

– Т. 67 – С.93–104.

33. Васильева И.К. Метод автоматической кластеризации данных дистанционного зондирования / Васильева И.К., Попов А.В. // *Авиационно-космическая техника и технология* – 2019.

34. Kennedy R.E. Implementation of the LandTrendr Algorithm on Google Earth Engine / Kennedy R.E., Yang Z., Gorelick N., Braaten J., Cavalcante L., Cohen W.B., Healey S. // *Remote Sensing* – 2018. – Т. 10 – № 5.

35. Рогова Н.В. Разработка и апробация методики дистанционного анализа изменений лесного покрова на российском Кавказе за последние полвека / Рогова Н. В., Скворцов В. Э. // *Устойчивое лесопользование* – 2021. – № 1 – С.30–52.

36. Ямашкин С.А. Разработка алгоритма классификации данных дистанционного зондирования Земли с применением методов глубокого машинного обучения для анализа геосистемной модели территории / Ямашкин С.А., Ямашкин А.А., Занозин В.В., Бармин А.Н. // *Геодезия и картография* – 2021. – Т. 82 – № 4 – С.54-64.

37. Абросимов А.В. Российские данные ДЗЗ и потенциал для космического мониторинга / Абросимов А.В., Беленов А.В., Дворкин Б.А., Орлов Т.В., Агольцов А.Ю. // *Геопрофи* – 2017. – Т. 4 – С.42-46.

38. Sidhu N. Using Google Earth Engine to detect land cover change: Singapore as a use case / Sidhu N., Pebesma E., Câmara G. // *European Journal of Remote Sensing* – 2018. – Т. 51 – № 1 – С.486-500.

39. Mutanga O. Google earth engine applications // *Remote Sens.* – 2019. – Т. 11. – № 5. – 591с.

40. Федотова Е.В. Использование Google Earth Engine в системах мониторинга Региональные проблемы дистанционного зондирования Земли, 2021. – 132–135с.

41. Intact Forest Landscapes [Электронный ресурс]. URL: <https://intactforests.org/index.html> (Дата обращения: 20.04.2023).

42. Mateo-Garcia G. Multitemporal cloud masking in the Google Earth Engine / Mateo-Garcia G., Gómez-Chova L., Amorós-López J., Muñoz-Mari J., Camps-Valls G. // *Remote Sensing* – 2018. – Т. 10 – № 7 – С.1079.

43. More accurate and flexible cloud masking for Sentinel-2 images// Medium [Электронный ресурс]. URL: <https://medium.com/google-earth/more-accurate-and-flexible-cloud-masking-for-sentinel-2-images-766897a9ba5f> (Дата обращения: 15.05.2023).

44. Google выпустила сервис с картой Земли, обновляемой в режиме реального времени// Хабр [Электронный ресурс]. URL: <https://habr.com/ru/news/672462/> (Дата обращения: 16.05.2023).

45. Brown C.F. Dynamic World, Near real-time global 10 m land use land cover mapping / Brown C.F., Brumby S.P., Guzder-Williams B., Birch T., Hyde S.B., Mazzariello J., Czerwinski W., Pasquarella V.J., Haertel R., Ilyushchenko S., others // Scientific Data – 2022. – Т. 9 – № 1 – С.251.

46. Down to Earth with AI Platform [Электронный ресурс]. URL: <https://medium.com/google-earth/down-to-earth-with-ai-platform-7bc363abf4fa> (Дата обращения: 15.05.2023).

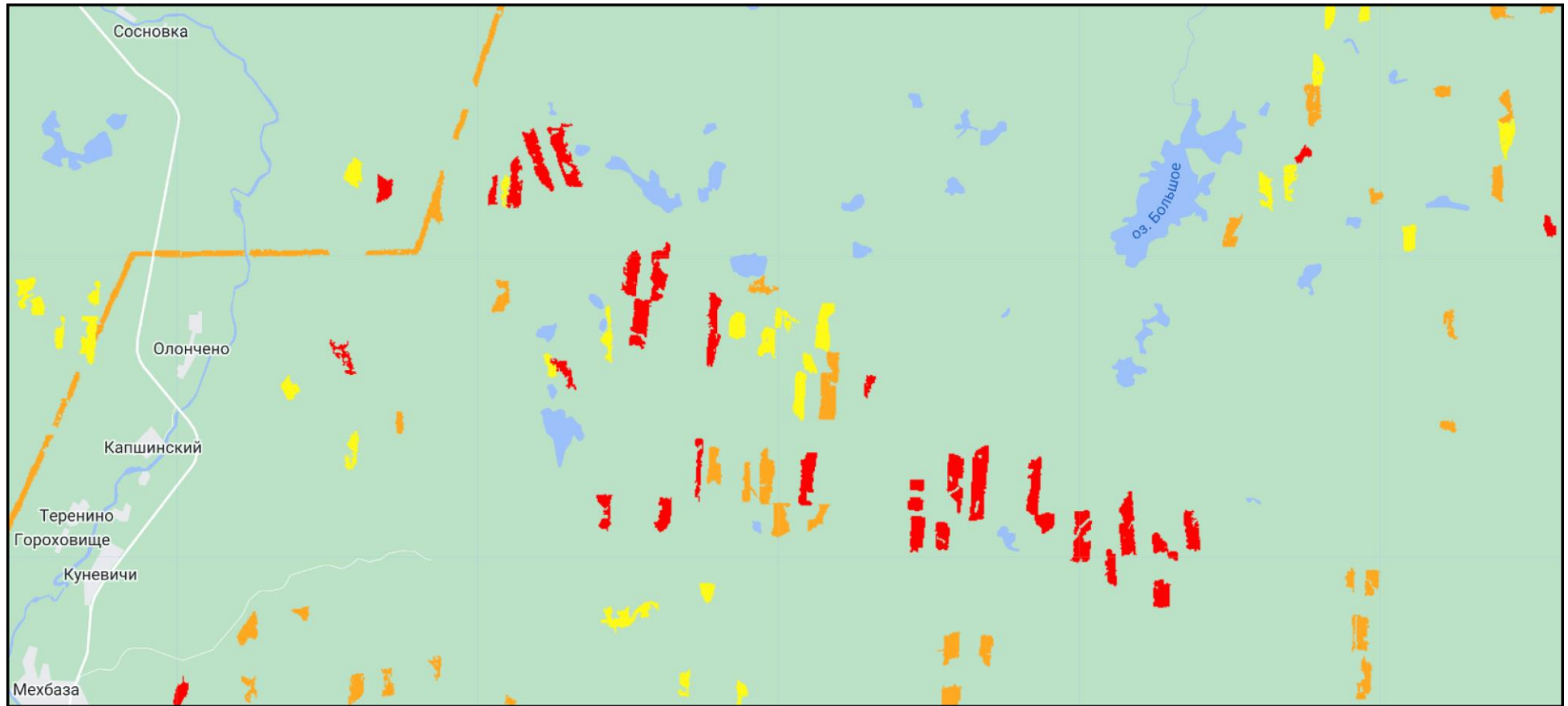
47. Gandhi U. End-to-End Google Earth Engine (Full Course Material) [Электронный ресурс]. URL: <https://courses.spatialthoughts.com/end-to-end-gee.html#module-4-change-detection> (Дата обращения: 15.05.2023).

48. Google Earth Engine – Google for Developers [Электронный ресурс]. URL: <https://developers.google.com/earth-engine/guides> (Дата обращения: 15.05.2023).

49. Alonso Martínez L. Automatic forest change detection through a bi-annual time series of satellite imagery: Toward production of an integrated land cover map / Alonso Martínez L., Picos J., Armesto J. // International Journal of Applied Earth Observation and Geoinformation – 2023. – Т. 118 – С.103289.

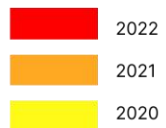
Приложение А

Фрагмент веб-карты нарушенных лесных земель на Google Earth Engine по данным Sentinel-2 за 2020-2022 годы



Условные обозначения

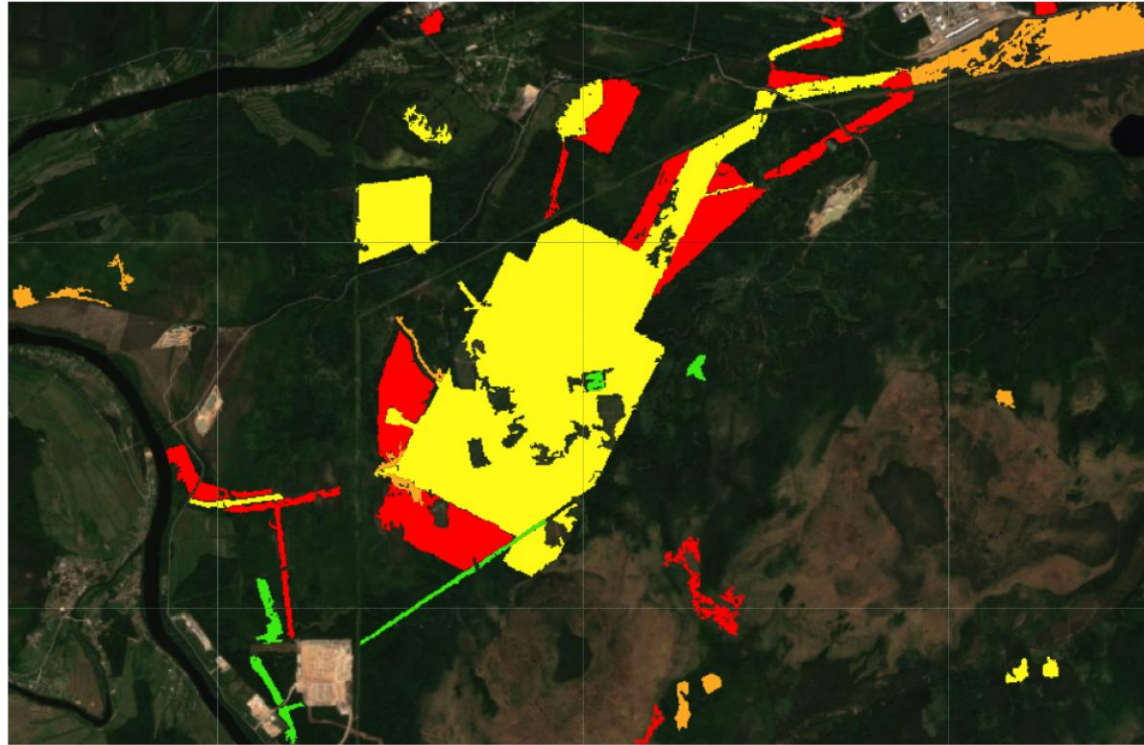
Нарушение леса по годам:



Картографическая основа: Google

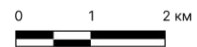
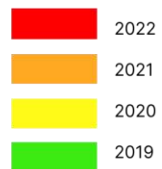
Приложение Б

Фрагмент веб-карты нарушенных лесных земель на Google Earth Engine по данным Sentinel-2 за 2019-2022 годы



Условные обозначения

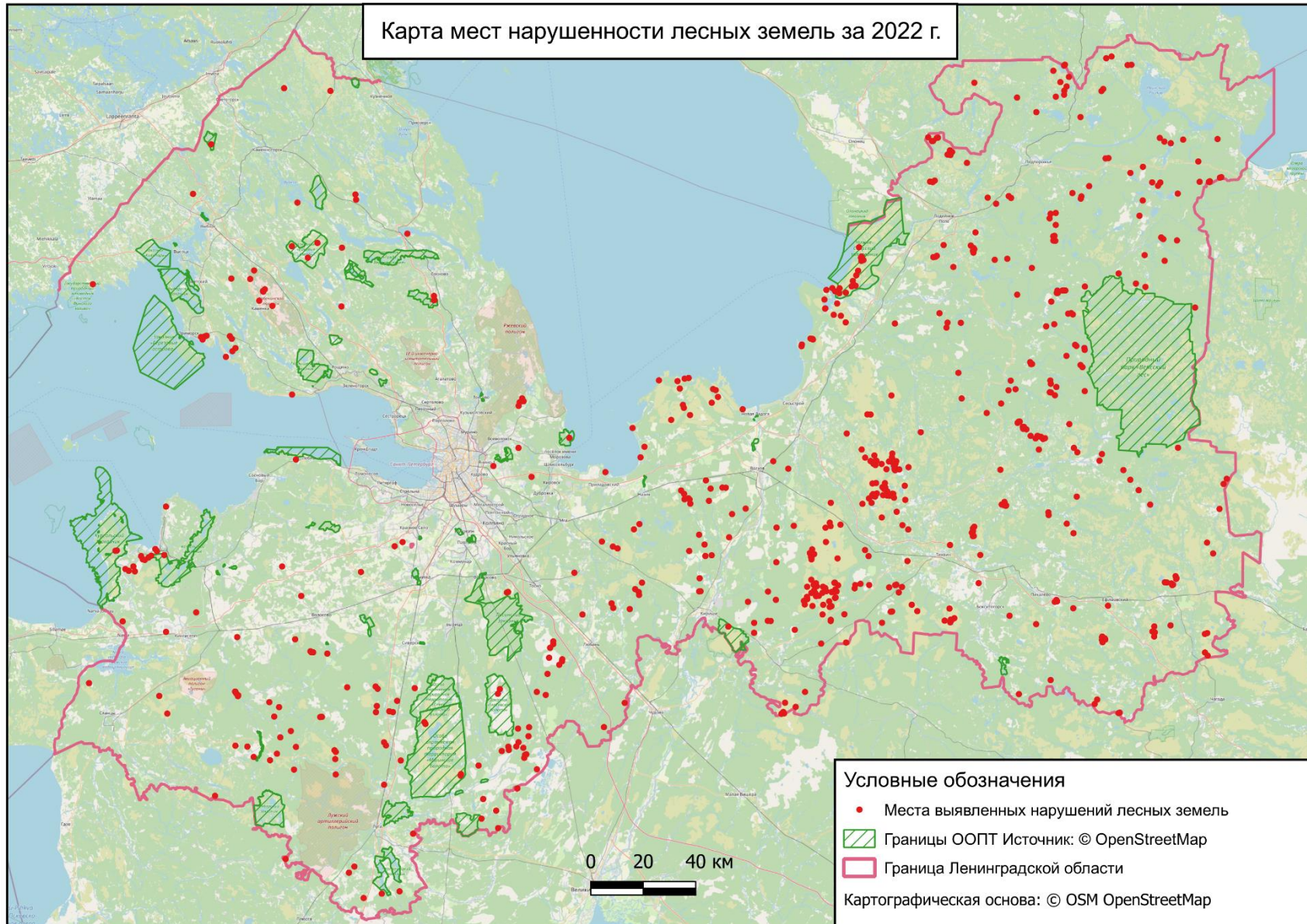
Нарушение леса по годам:



Подложка: композит Sentinel-2 июль 2018

Приложение В

Карта мест нарушения лесных земель за 2022 г.



Приложение Г

```
var protectedArea = ee.FeatureCollection("projects/ee-dicword11/assets/protected_area"),
var region = ee.FeatureCollection("projects/ee-dicword11/assets/lenobl"),
var samples = ee.FeatureCollection("projects/ee-dicword11/assets/GCP");

Map.centerObject(poi, 11);

// Cloud masking using s2cloudless
var REGION = region.geometry();
var CLOUD_FILTER = 100;
var CLD_PRB_THRESH = 30;
var NIR_DRK_THRESH = 0.15;
var CLD_PRJ_DIST = 1;
var BUFFER = 50;

function get_s2_sr_cld_col(region, start_date, end_date) {
  var s2_sr_col = (ee.ImageCollection('COPERNICUS/S2_SR_HARMONIZED')
    .filterBounds(region)
    .filterDate(start_date, end_date)
    .filter(ee.Filter.lte('CLOUDY_PIXEL_PERCENTAGE', CLOUD_FILTER)));

  var s2_cloudless_col = (ee.ImageCollection('COPERNICUS/S2_CLOUD_PROBABILITY')
    .filterBounds(region)
    .filterDate(start_date, end_date));

  return ee.ImageCollection(ee.Join.saveFirst('s2cloudless').apply({
    'primary': s2_sr_col,
    'secondary': s2_cloudless_col,
    'condition': ee.Filter.equals({
      'leftField': 'system:index',
      'rightField': 'system:index'
    })
  }));
}

function add_cloud_bands(img) {
  var cld_prb = ee.Image(img.get('s2cloudless')).select('probability');
  var is_cloud = cld_prb.gt(CLD_PRB_THRESH).rename('clouds');
  return img.addBands(ee.Image([cld_prb, is_cloud]));
}

function add_shadow_bands(img) {
  var not_water = img.select('SCL').neq(6);

  var SR_BAND_SCALE = 1e4;
  var dark_pixels = img.select('B8').lt(NIR_DRK_THRESH *
SR_BAND_SCALE).multiply(not_water).rename('dark_pixels');

  var shadow_azimuth =
ee.Number(90).subtract(ee.Number(img.get('MEAN_SOLAR_AZIMUTH_ANGLE')));

  var cld_proj =
(img.select('clouds').directionalDistanceTransform(shadow_azimuth, CLD_PRJ_DIST *
10)
  .reproject({'crs': img.select(0).projection(), 'scale': 100})
  .select('distance')
  .mask()
  .rename('cloud_transform'));
}
```

```

    var shadows = cld_proj.multiply(dark_pixels).rename('shadows');

    return img.addBands(ee.Image([dark_pixels, cld_proj, shadows]));
}

function add_cld_shdw_mask(img) {
    var img_cloud = add_cloud_bands(img);

    var img_cloud_shadow = add_shadow_bands(img_cloud);

    var is_cld_shdw =
img_cloud_shadow.select('clouds').add(img_cloud_shadow.select('shadows')).gt(0);

    is_cld_shdw = (is_cld_shdw.focal_min(2).focal_max(BUFFER * 2 / 20)
        .reproject({'crs': img.select([0]).projection(), 'scale': 20})
        .rename('cloudmask'));

    return img_cloud_shadow.addBands(is_cld_shdw);
}

function apply_cld_shdw_mask(img) {
    var not_cld_shdw = img.select('cloudmask').not();

    return img.select('B.*').updateMask(not_cld_shdw).divide(10000);
}

// Adding NDVI and NBR bands
function addNDVI(image) {
    var ndvi = image.normalizedDifference(['B8', 'B4']).rename('NDVI');
    return image.addBands(ndvi);
}

function addNBR(image) {
    var nbr = image.normalizedDifference(['B8', 'B12']).rename('NBR');
    return image.addBands(nbr);
}

function processDiff(previousImage, image) {
    var nb = previousImage.select('NDVI').subtract(image.select('NDVI'));
    var nbrReclass = ee.Image([nb.gt(0.1).multiply(1), nb.lt(0.1).and(nb.gte(-
0.1)).multiply(0), nb.lt(-0.1).multiply(2)])
        .reduce(ee.Reducer.sum());
    return Map.addLayer(nbrReclass.selfMask(),
        {min:0, max:2, palette:['green', 'red', 'lightgreen']}, 'Разность NDVI 20' +
y);
}

var previousImage = null;

// Visualizing composites and change detection
// applying Dynamic World collection and NBR difference
for (var y = 18; y <= 22; y++) {
    var START_DATE = '20' + y + '-06-01';
    var END_DATE = '20' + y + '-09-01';

    var s2_sr_cld_col = get_s2_sr_cld_col(region, START_DATE, END_DATE);

    var image = s2_sr_cld_col.map(add_cld_shdw_mask).map(apply_cld_shdw_mask)
        .map(addNBR).map(addNDVI).median().clip(region);
}

```

```

var vis_rgb = {
  bands:['B4','B3','B2'],
  min:0, max:0.3
};

Map.addLayer(image, vis_rgb, 'RGB композит 20' + y);

// NDVI difference
if (previousImage !== null) {
  processDiff(previousImage, image);
}
var previousImage = image;

// Dynamic World difference
var beforeYear = 2000 + y;
var ny = y + 1;
if (y == 22) {
  break;
}
else {
  var afterYear = 2000 + ny;
}

var beforeStart = ee.Date.fromYMD(beforeYear, 6, 1);
var beforeEnd = beforeStart.advance(3, 'month');

var afterStart = ee.Date.fromYMD(afterYear, 6, 1);
var afterEnd = afterStart.advance(3, 'month');

var dw = ee.ImageCollection('GOOGLE/DYNAMICWORLD/V1')
  .filterBounds(region)
  .select(['trees']);

var beforeDw = dw.filterDate(beforeStart, beforeEnd).median();
var afterDw = dw.filterDate(afterStart, afterEnd).median();

var difDw = beforeDw.gt(0.7).and(afterDw.lt(0.55)).clip(region);

var changeVisParams = {min: 0, max: 1, palette: ['white', 'red']};
Map.addLayer(difDw.selfMask(), changeVisParams, 'Вырубки по Dynamic World 20' +
ny);
}

// Supervised Classification with custom training data set
var START_DATE = '2021-06-01';
var END_DATE = '2021-09-01';

var s2_sr_cld_col = get_s2_sr_cld_col(region, START_DATE, END_DATE);

var composite = s2_sr_cld_col.map(add_cld_shdw_mask).map(apply_cld_shdw_mask)
  .map(addNBR).map(addNDVI).median().clip(region);

var samples = water.merge(forest).merge(urban).merge(grass).merge(bare);

var gcp = samples.randomColumn();

var trainingGcp = gcp.filter(ee.Filter.lt('random', 0.7));
var validationGcp = gcp.filter(ee.Filter.gte('random', 0.7));

var training = composite.sampleRegions({
  collection: trainingGcp,

```

```

    properties: ['class'],
    scale: 10
  });

var bands = composite.bandNames();
var numBands = bands.length();
var bandsWithClass = bands.add('class');
var classIndex = bandsWithClass.indexOf('class');

var combinedReducer = ee.Reducer.mean().combine({
  reducer2: ee.Reducer.stdDev(),
  sharedInputs: true});

var repeatedReducer = combinedReducer.repeat(numBands).group(classIndex);

var sampleStats = training.reduceColumns({
  selectors: bands.add('class'),
  reducer: repeatedReducer,
});

var groups = ee.List(sampleStats.get('groups'));

var classNames = ee.List(['water', 'forest', 'urban', 'grass', 'bare']);

var fc = ee.FeatureCollection(groups.map(function(item) {
  var values = ee.Dictionary(item).get('mean');
  var groupNumber = ee.Dictionary(item).get('group');
  var properties = ee.Dictionary.fromLists(bands, values);
  var withClass = properties.set('class', classNames.get(groupNumber));
  return ee.Feature(null, withClass);
})));

// Chart spectral signatures of training data
var options = {
  title: 'Средние спектральные сигнатуры',
  hAxis: {title: 'Каналы'},
  vAxis: {title: 'Отражательная способность',
    viewWindowMode:'explicit',
    viewWindow: {
      max:0.6,
      min:0
    }
  },
  lineWidth: 1,
  pointSize: 4,
  series: {
    0: {color: 'blue'},
    1: {color: 'green'},
    2: {color: 'gray'},
    3: {color: 'lime'},
    4: {color: 'brown'}
  }
};

var bandDescriptions = {
  'B2': 'B02/Blue',
  'B3': 'B03/Green',
  'B4': 'B04/Red',
  'B8': 'B08/NIR',
  'B11': 'B11/SWIR-1',
  'B12': 'B12/SWIR-2'
};

```

```

var chart = ui.Chart.feature.byProperty({
  features: fc,
  xProperties: bandDescriptions,
  seriesProperty: 'class'
})
.setChartType('ScatterChart')
.setOptions(options);

print(chart);

// Classification options
var bands2classif = ['B3', 'B8', 'B12', 'NBR', 'NDVI'];

var classifier = ee.Classifier.smileRandomForest({
  numberOfTrees: 100,
});

var trained = classifier.train(training, 'class', bands2classif);

var beforeImage = null;

function forestChange(beforeImage, afterImage) {
  var forestLoss = beforeImage.subtract(afterImage).eq(1);

  var patchsize = forestLoss.connectedPixelCount(110, false);

  var filtered = forestLoss.focalMode({
    radius: 500,
    kernelType: 'square',
    units: 'meters',
  });

  var connected = forestLoss.where(patchsize.lt(100), filtered);

  return connected.updateMask(connected.neq(0));
}

// Running classification
for (var y = 18; y <= 22; y++) {
  var START_DATE = '20' + y + '-06-01';
  var END_DATE = '20' + y + '-09-01';

  var s2_sr_cld_col = get_s2_sr_cld_col(region, START_DATE, END_DATE);

  var image = s2_sr_cld_col.map(add_cld_shdw_mask).map(apply_cld_shdw_mask)
    .map(addNBR).map(addNDVI).median().clip(region);

  var classifImage = image.select(bands).classify(trained);

  if (y == 21) {
    var test = classifImage.sampleRegions({
      collection: validationGcp,
      properties: ['class'],
      tileScale: 16,
      scale: 10,
    });

    var testConfusionMatrix = test.errorMatrix('class', 'classification');

    print('Матрица ошибок', testConfusionMatrix);
    print('Общая точность', testConfusionMatrix.accuracy());
  }
}

```

```

}

var afterImage = classifImage.eq(1); // Class 'forest' is 1

if (beforeImage !== null) {
  var forestLoss = forestChange(beforeImage, afterImage);

  Map.addLayer(forestLoss, {palette: ['red']}, 'Вырубки 20' + y);

  var vectors = forestLoss.reduceToVectors({
    geometry: region.geometry(),
    scale: 50,
    geometryType: 'centroid',
    eightConnected: false,
    maxPixels: 1e15,
    tileSize: 4,
    reducer: ee.Reducer.countEvery()
  });

  Export.table.toDrive({
    collection: vectors,
    description: 'Centroids_20' + y,
    folder: 'GEE_Folder',
    fileNamePrefix: 'change_sites'});
}
beforeImage = afterImage;
}

Map.addLayer(protectedArea, {color: 'green'}, 'Границы ООПТ');

```