

Двадцать вторая международная конференция "СОВРЕМЕННЫЕ ПРОБЛЕМЫ
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА"

**Опыт разработки алгоритма
выделения лесных просек под
линиями электропередач в лесных
ландшафтах на основе данных
Sentinel-2**

Бахрамхан Я.О., НИУ ВШЭ

Ермаков Д.М., ИКИ РАН, ФИРЭ им. В.А. Котельникова РАН

Подольская Е.С., ЦЭПЛ РАН, НИУ ВШЭ

Актуальность исследования

В настоящее время объем пространственных данных настолько велик, что для его обработки нужны эффективные автоматизированные технологии. Такими технологиями призваны алгоритмы машинного обучения. Они позволяют актуализировать уже имеющиеся пространственные данные, что в свою очередь важно при решении практических задач.

В частности, лесные просеки под линиями электропередач (ЛЭП) необходимо поддерживать в нормативном состоянии для бесперебойной передачи электроэнергии в населенные пункты и промышленные объекты, и задача их сохранности становится все более актуальной по мере роста энергопотребления.



Объект и предмет исследования

Объект исследования – лесные просеки под линиями электропередач в районе Богучанской ГЭС, Кежемский и Богучанский район, Красноярский край

Предмет исследования – признаки выделения лесных просек под линиями электропередач на космических снимках



Подложка: CNES/Airbus, Google Earth Pro



Подложка: Esri Satellite

Цель, задачи исследования

Цель - создать алгоритм распознавания лесных просек под ЛЭП для их автоматической идентификации на космических снимках

Задачи:

1. Провести литературный обзор статей по автоматическому распознаванию и особенности дешифрирования лесных просек под ЛЭП
2. Подобрать данные дистанционного зондирования на территорию исследования в оптическом диапазоне и с низкими значениями облачности для более успешного обучения и тестирования алгоритма
3. Создать алгоритм машинного обучения с помощью методов моделирования и программирования на Python
4. Обучить и протестировать получившийся алгоритм
5. Оценить качество работы алгоритма визуально, а также путем сопоставления с векторными данными о сетях линий электропередач от OpenStreetMap

Выбор экспериментальных участков и дешифровочные признаки лесных просек

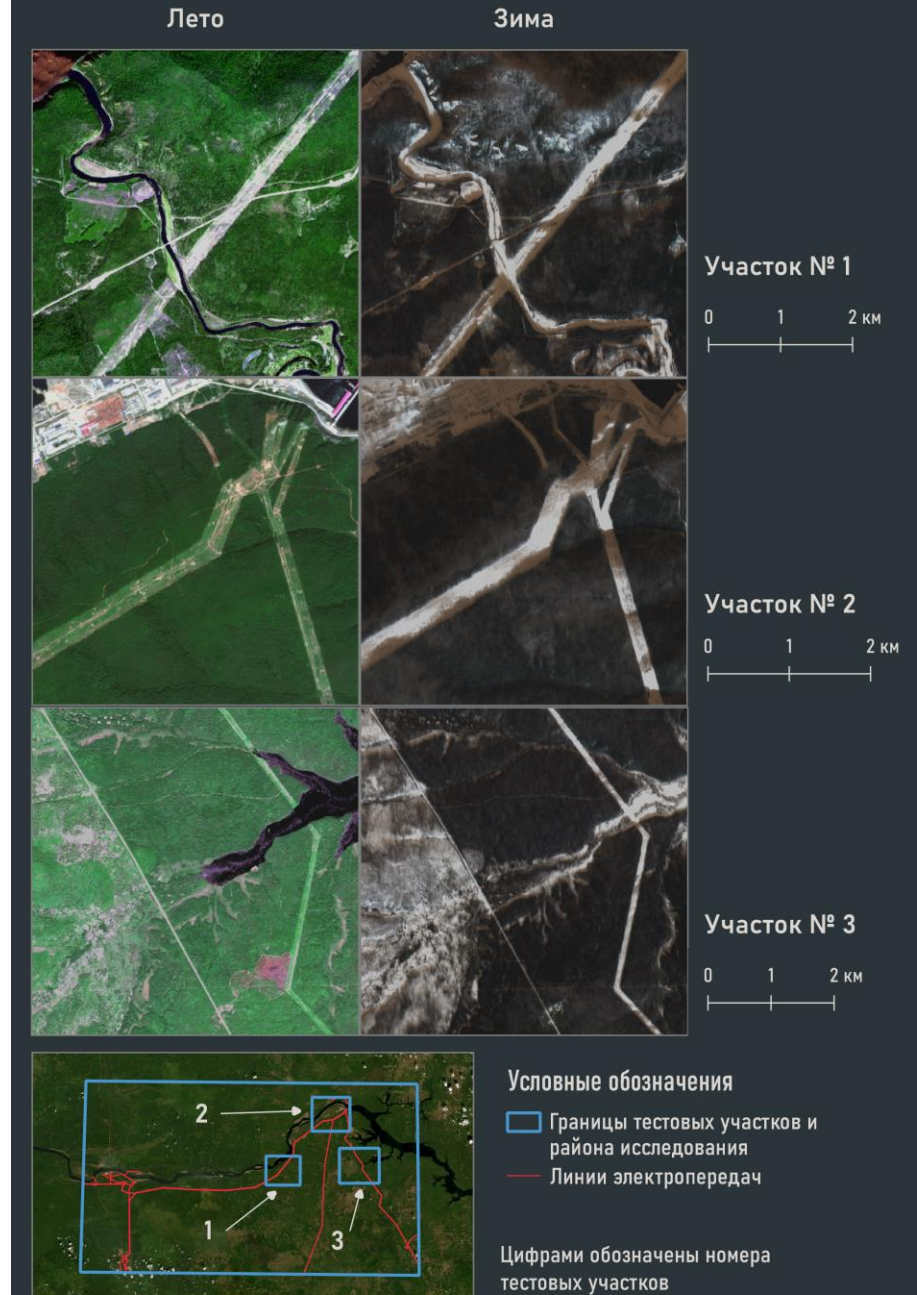
Для тестирования алгоритмов были выбраны три наиболее отличительные друг от друга экспериментальных участка.

Лесная просека под линиями электропередач обладает несколькими признаками, позволяющими её дешифровать от остальных объектов. Среди них можно выделить два:

- Высокая спектральная яркость в коротковолновом инфракрасном диапазоне
- Линейная форма

Кроме того, просека характеризуется низкой степенью шероховатости по сравнению с лесными массивами и редколесьем, а также с антропогенными объектами такими как городская застройка.

Тестовые участки и их расположение на территории исследования

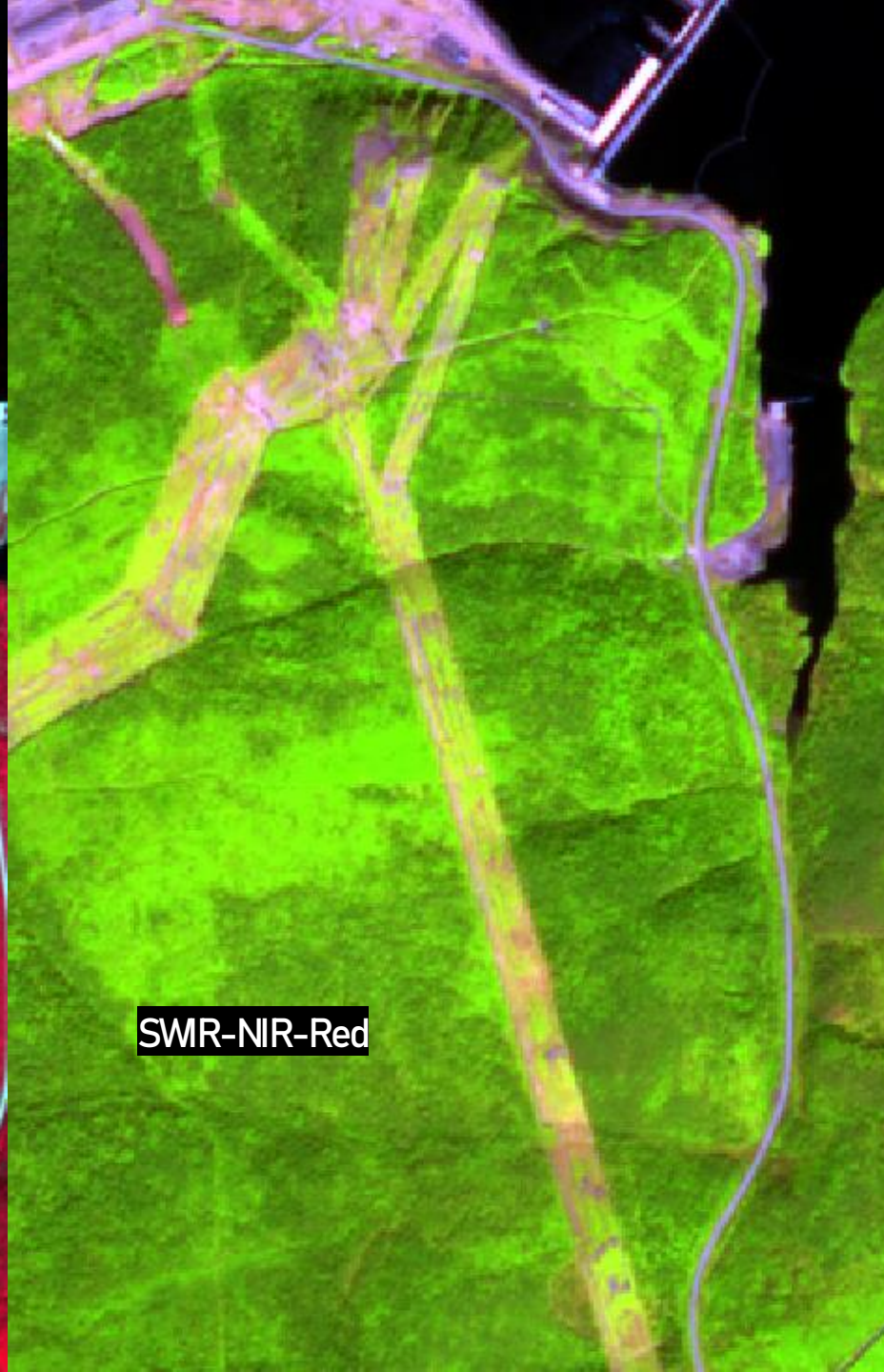




Red-Green-Blue



NIR-Red-Green



SWIR-NIR-Red

Спецификация алгоритма

Алгоритм призван выделить лесную просеку под ЛЭП на основе ее дешифровочных признаков, поэтому для создания комплексного алгоритма выделения лесных просек необходимо использовать алгоритмы, позволяющие выделить лесную просеку на основе каждого признака.

Для выделения просеки под ЛЭП на основе спектральной яркости решено использовать модель машинного обучения, классифицирующая всю территорию экспериментальных участков. Среди моделей для обучения выбраны:

- К ближайших соседей
- Логистическая регрессия
- Метод опорных векторов
- Случайный лес

В качестве классов землепользования для создания карт классификаций наземного покрова выбраны следующие:

1. **Водные объекты**
2. **Леса**
3. **Редколесья**
4. **Луга или свободные от растительности земли**
5. **Антропогенные объекты**

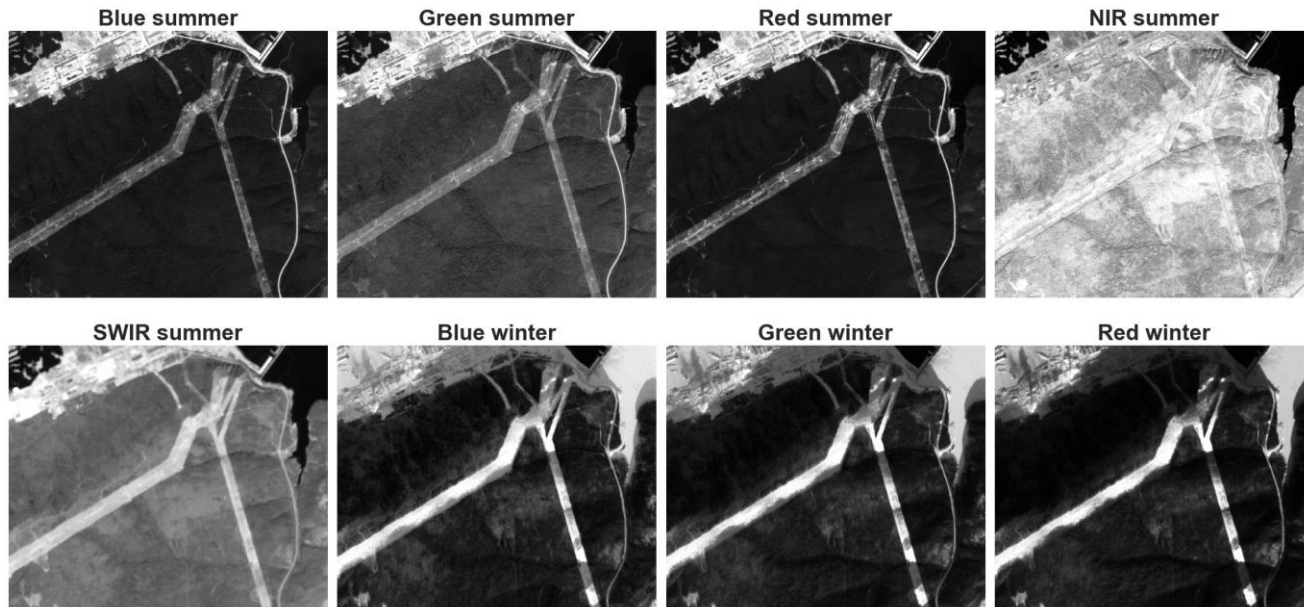
На основе маски лугов (класс № 4), полученной из карты классификации, алгоритм поиска линейных объектов выделяет лесные просеки под ЛЭП. В качестве такого алгоритма выбрано вероятностное преобразование Хафа.

Выбор признаков

$$SAVI = \frac{(\rho_{NIR} - \rho_{red})}{(\rho_{NIR} + \rho_{red} + L)} * (1 + L)$$

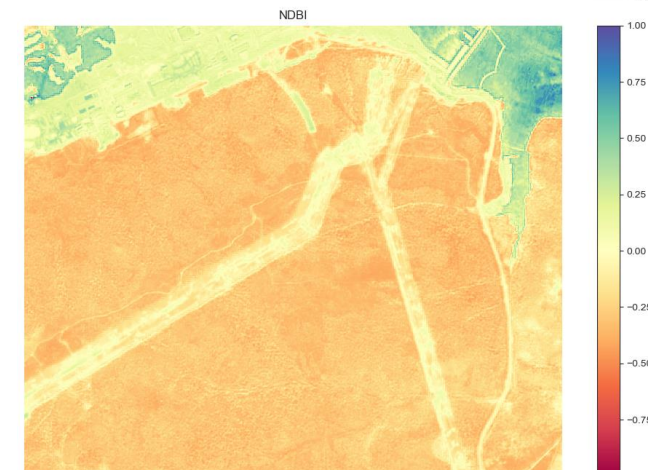
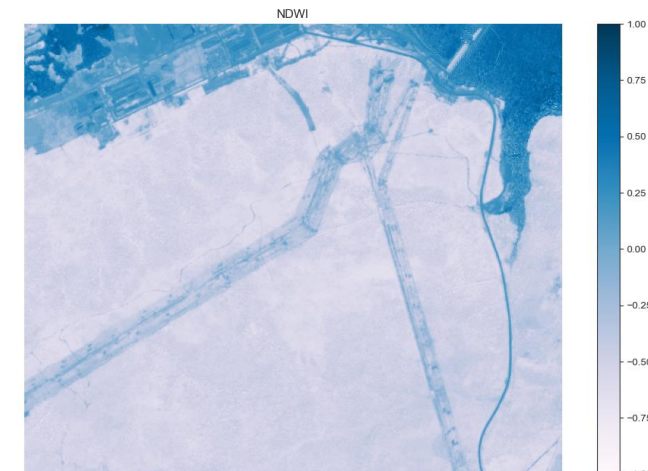
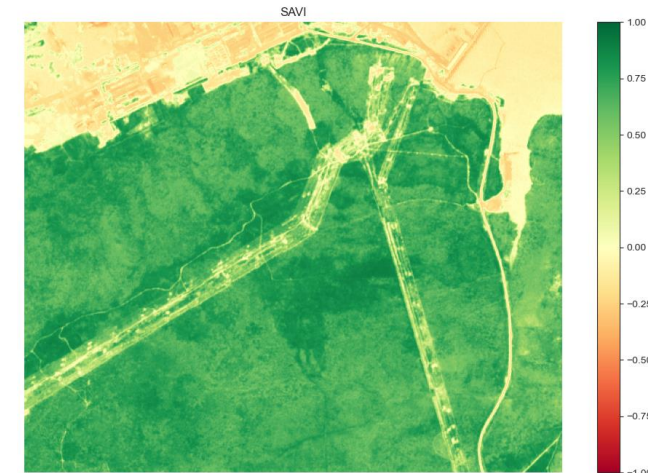
В качестве признаков выбраны:

- Каналы Blue, Green, Red, NIR и SWIR летнего изображения Sentinel-2
- Каналы Blue, Green, Red зимнего изображения Sentinel-2
- Спектральные индексы SAVI (индекс растительности с поправкой на почву), NDWI (нормализованный разностный водный индекс), NDBI (нормализованный разностный индекс городской застройки)



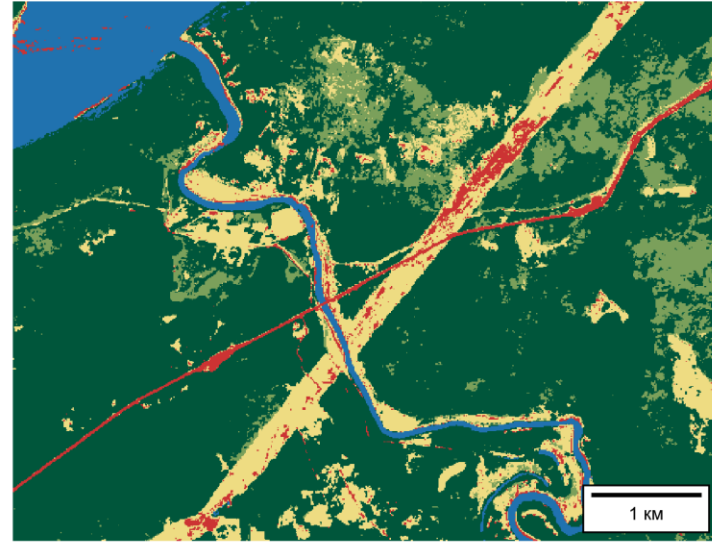
$$NDWI = \frac{(\rho_{green} - \rho_{NIR})}{(\rho_{green} + \rho_{NIR})}$$

$$NDBI = \frac{(\rho_{SWIR} - \rho_{NIR})}{(\rho_{SWIR} + \rho_{NIR})}$$

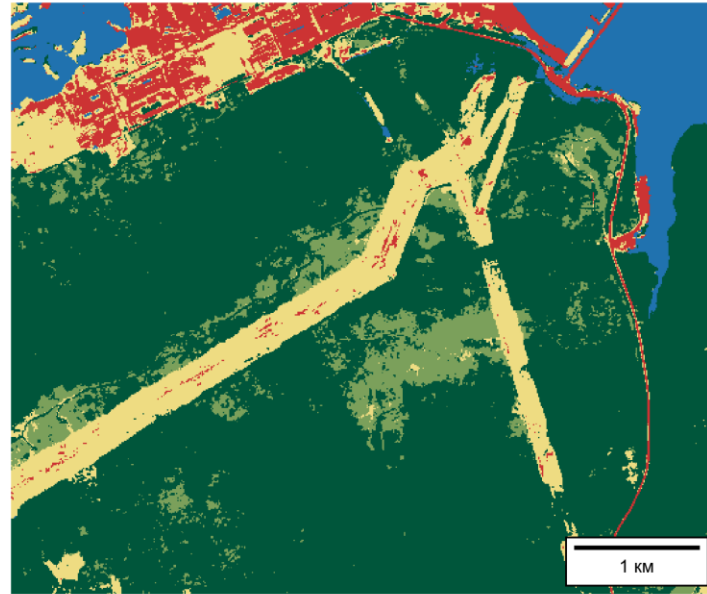


Карты классификации территории экспериментальных участков. Метод создания - k-ближайших соседей

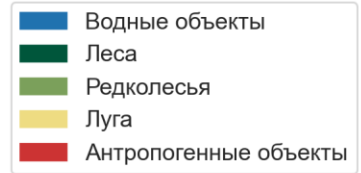
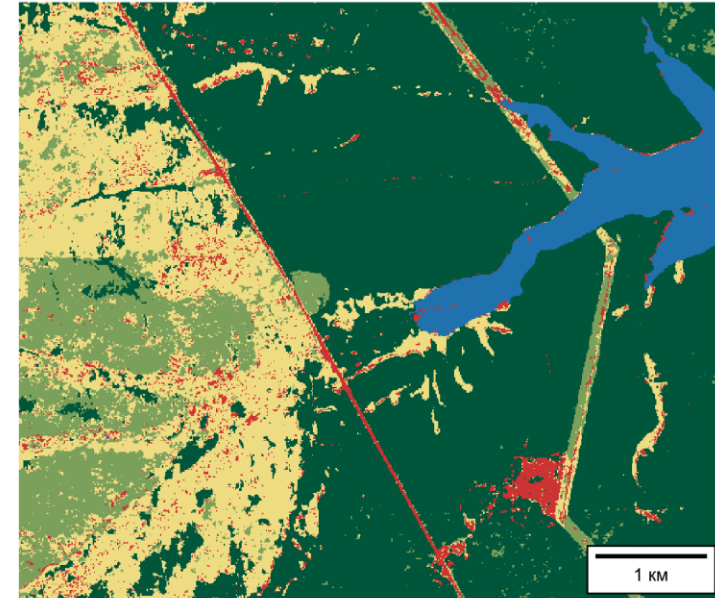
Участок № 1



Участок № 2

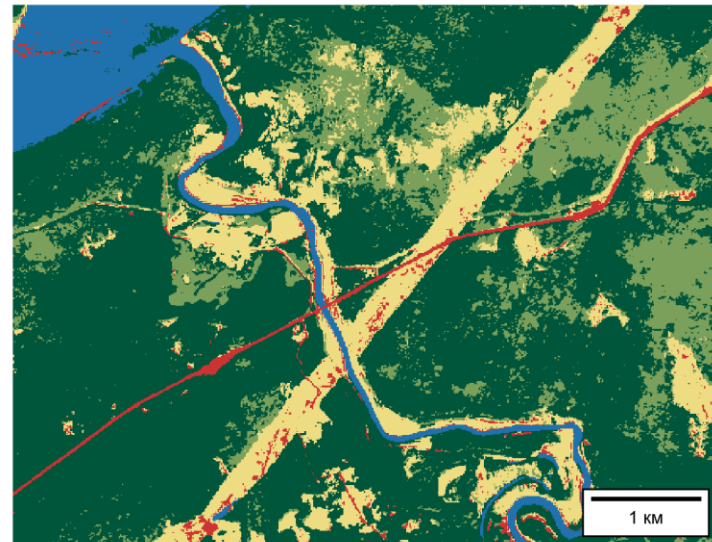


Участок № 3

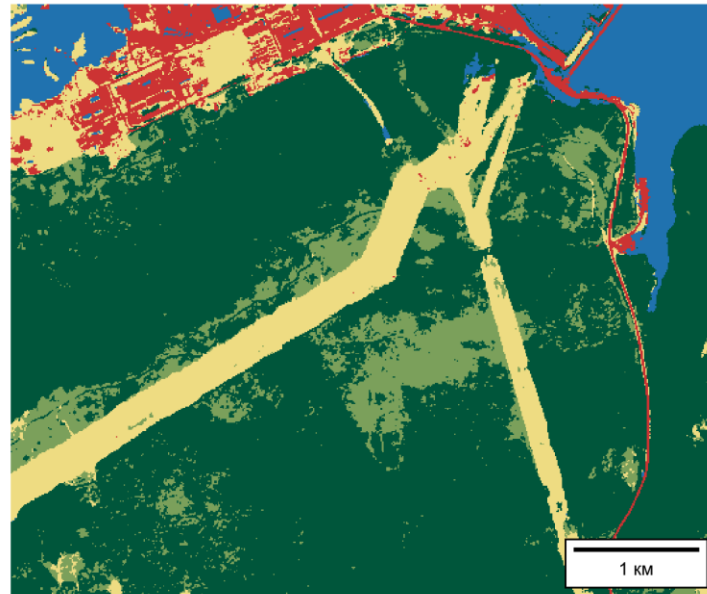


Карты классификации территории экспериментальных участков. Метод создания - логистическая регрессия

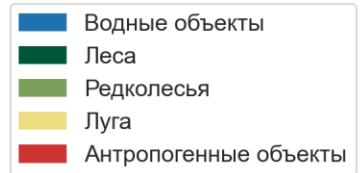
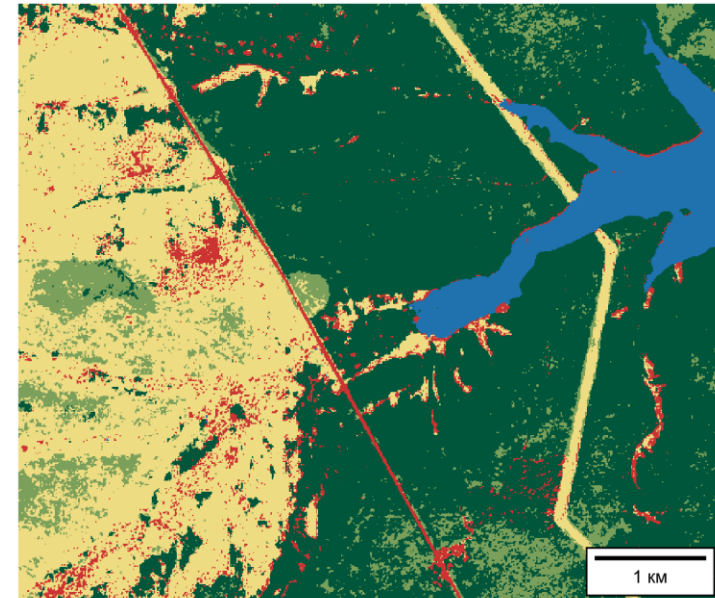
Участок № 1



Участок № 2

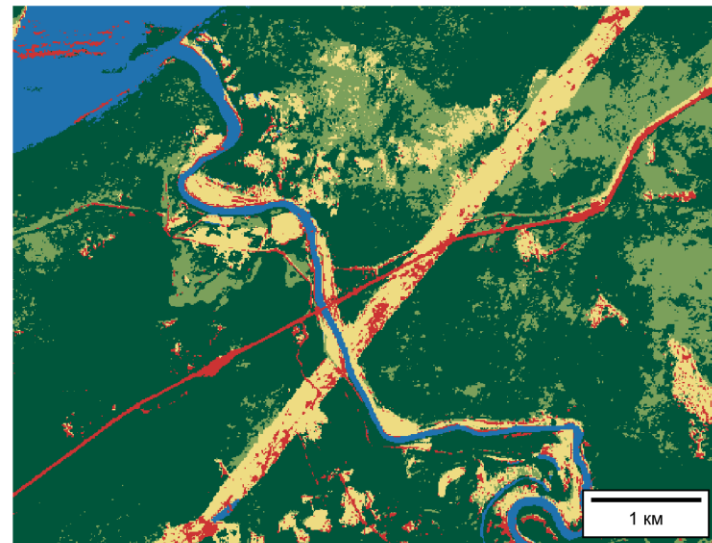


Участок № 3

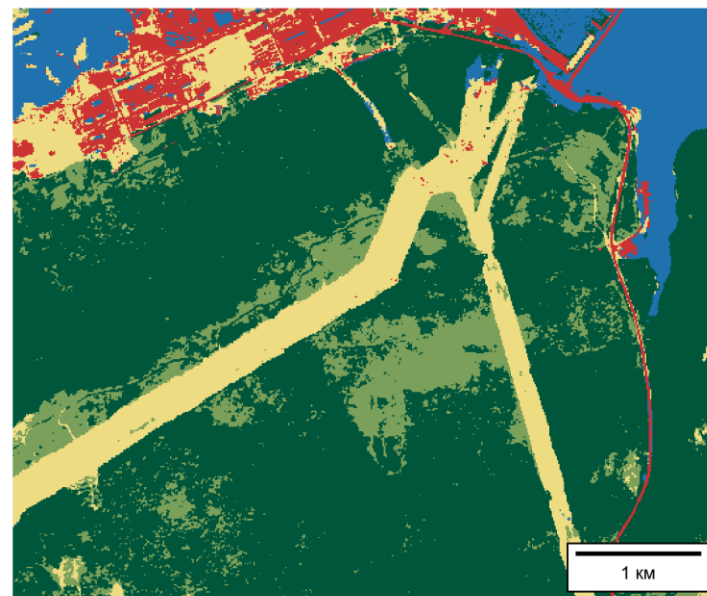


Карты классификации территории экспериментальных участков. Метод создания - метод опорных векторов

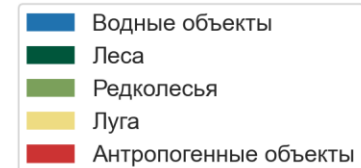
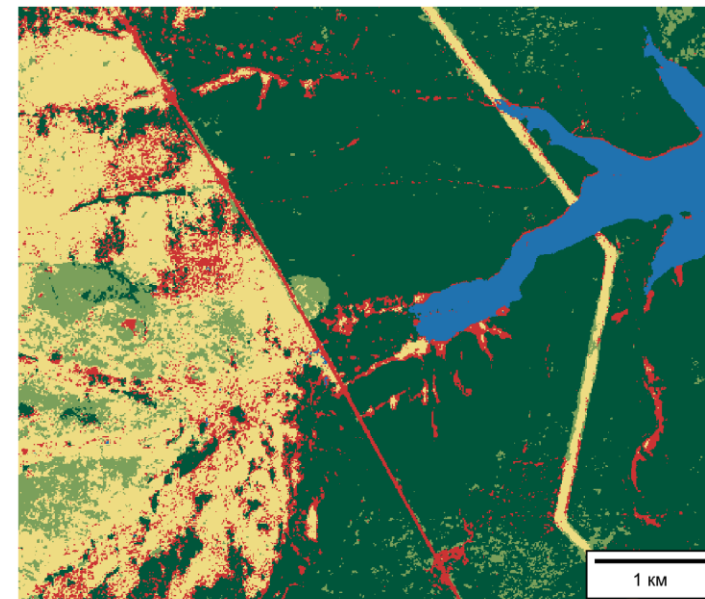
Участок № 1



Участок № 2

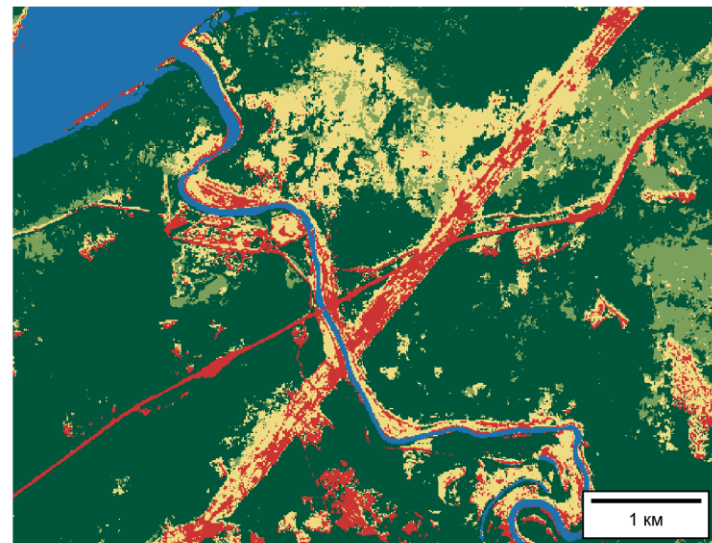


Участок № 3

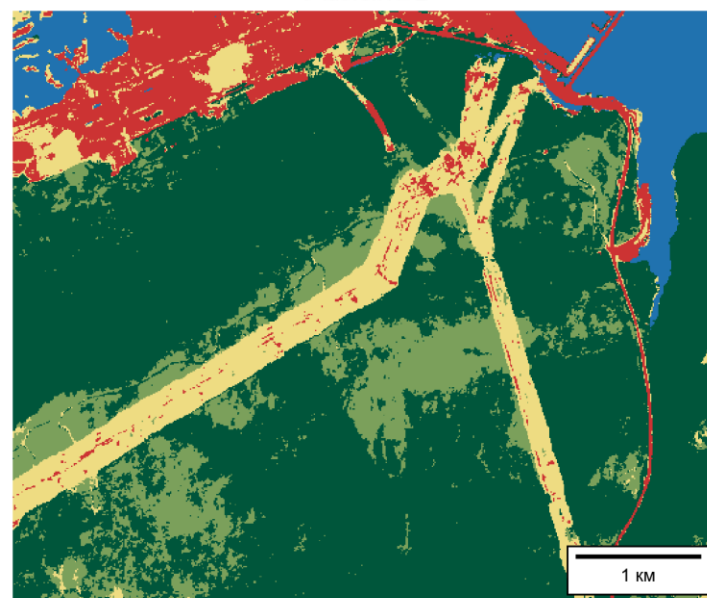


Карты классификации территории экспериментальных участков. Метод создания - случайный лес

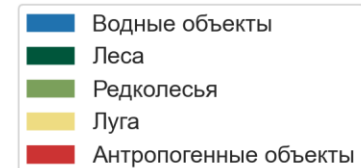
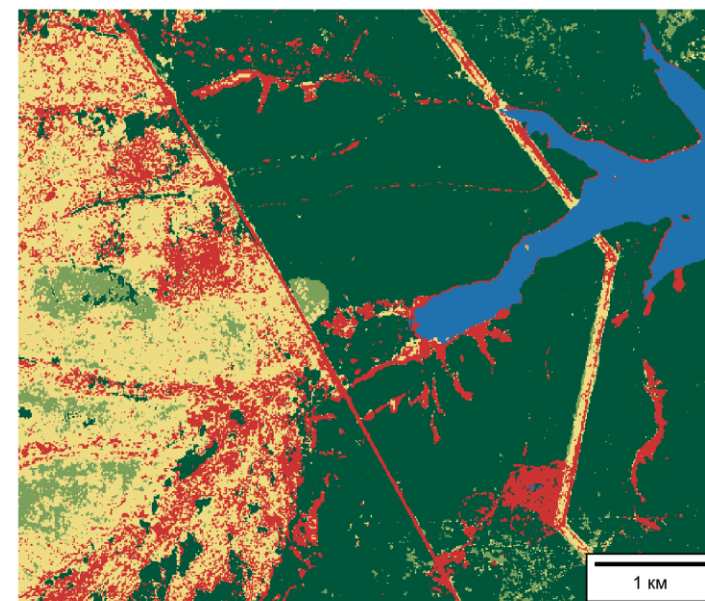
Участок № 1



Участок № 2



Участок № 3

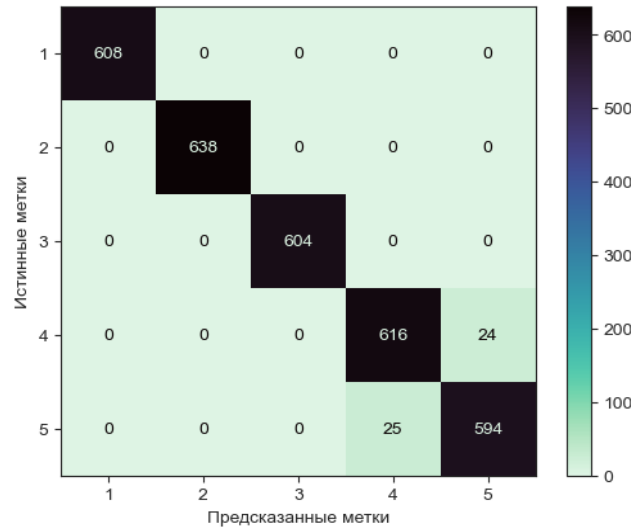


Результаты классификации

Номера классов на матрице несоответствия:

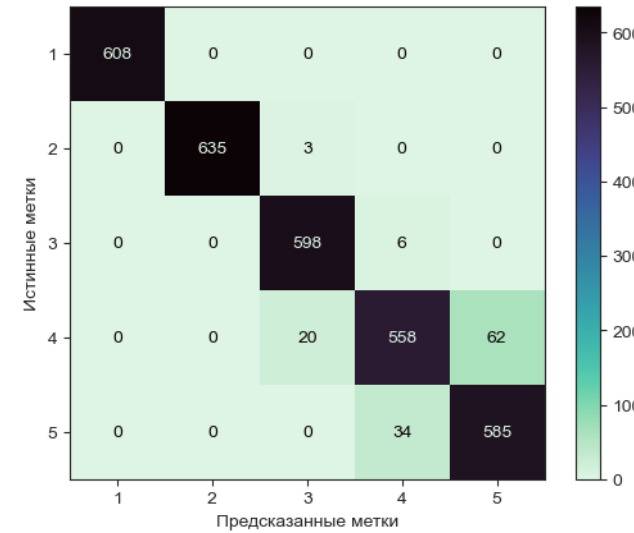
1. Водные объекты
2. Леса
3. Редколесья
4. Луга
5. Антропогенные объекты

К ближайших соседей



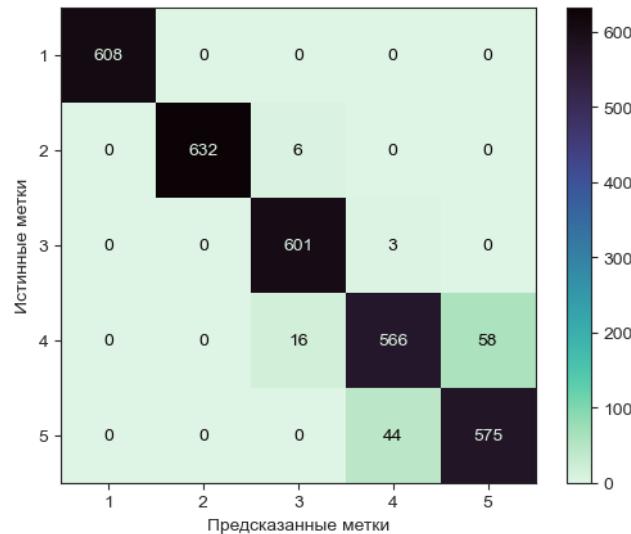
Accuracy: 0.984
Kappa score: 0.98
F1 score: 0.984

Логистическая регрессия



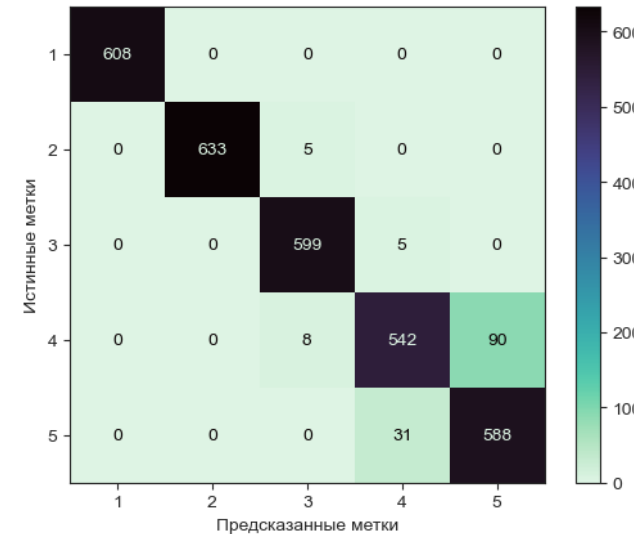
Accuracy: 0.96
Kappa score: 0.95
F1 score: 0.96

Метод опорных векторов



Accuracy: 0.959
Kappa score: 0.949
F1 score: 0.959

Случайный лес

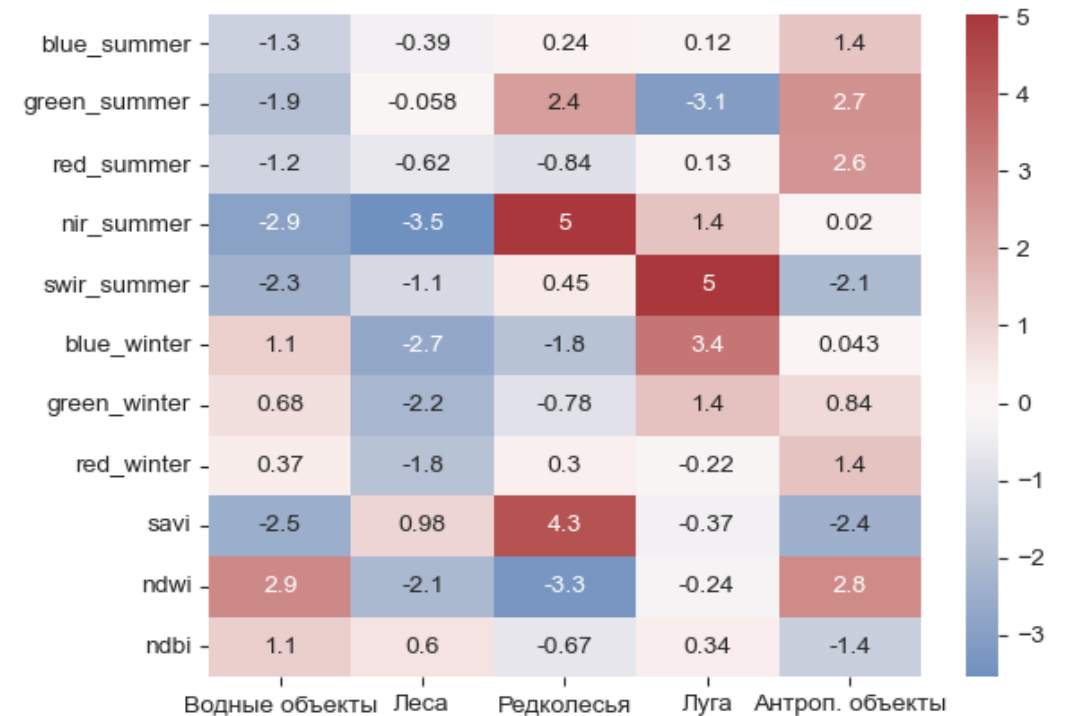


Accuracy: 0.955
Kappa score: 0.944
F1 score: 0.956

Результаты классификации и их интерпретация

По метрикам качества классификации – Ассигасу, Карра score и F1 score – и по матрице несоответствий все модели показали высокое качество выделения лесных просек. Поэтому подбор модели выполнялся на основе карт классификации. Среди моделей самой лучшей для поиска лесных просек на основе спектральной яркости выбрана логистическая регрессия. Ее результаты применены для поиска лесных просек.

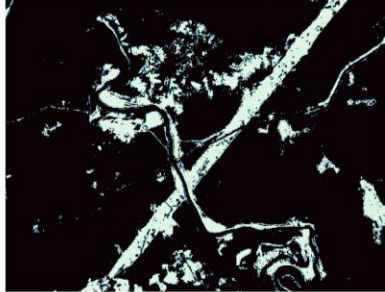
Помимо высокого качества классификации модель логистической регрессии подается простой интерпретации – в результате обучения модели создаются уравнения регрессии, выделяющие каждый из 5 классов объектов. Коэффициенты означают степень значимости признака для выделения определенного класса.



Результаты поиска лесных просек преобразованием Хафа

- Перед поиском линейных объектов преобразованием Хафа маска класса лугов была предварительно очищена от «шумовых» пикселей с помощью операции двоичного открытия.
- К полученной после этого шага маске применен детектор границ Канни, позволяющий выделить границы объектов на бинарной маске.
- Преобразование Хафа выделяет прямолинейные границы на маске детектора, которые в большинстве случаев относятся к границам лесных просек

Маска класса лугов из карты классификации, участок № 1



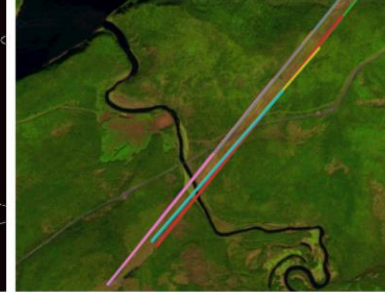
Маска класса лугов, обработанная двоичным открытием, участок № 1



Результат детектора границ Канни, участок № 1



Результат вероятностного преобразования Хафа, участок № 1



Результат выделения лесных просек, участок № 1



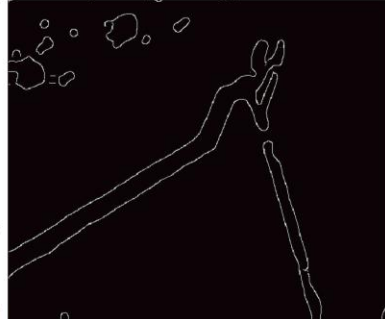
Маска класса лугов из карты классификации, участок № 2



Маска класса лугов, обработанная двоичным открытием, участок № 2



Результат детектора границ Канни, участок № 2



Результат вероятностного преобразования Хафа, участок № 2



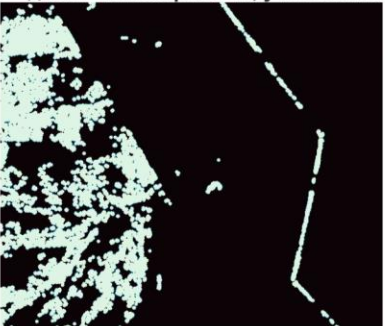
Результат выделения лесных просек, участок № 2



Маска класса лугов из карты классификации, участок № 3



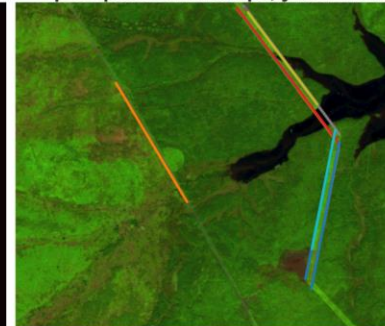
Маска класса лугов, обработанная двоичным открытием, участок № 3



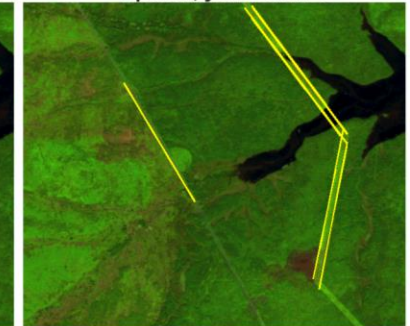
Результат детектора границ Канни, участок № 3



Результат вероятностного преобразования Хафа, участок № 3



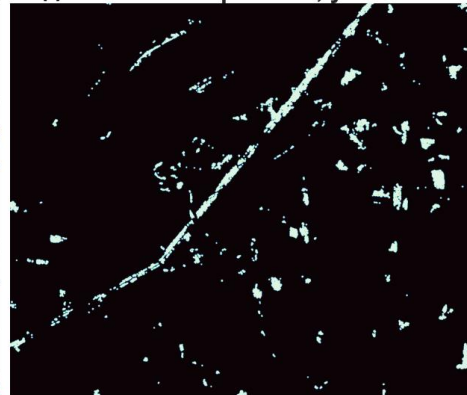
Результат выделения лесных просек, участок № 3



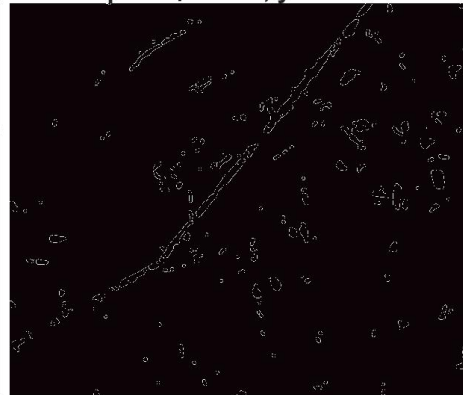
Маска класса лугов из карты классификации, участок № 1



Маска класса лугов, обработанная двоичным открытием, участок № 1



Результат детектора границ Канны, участок № 1



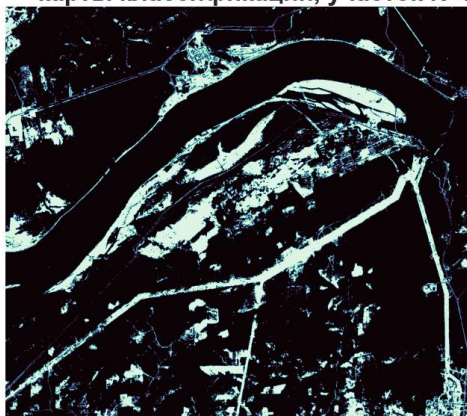
Результат вероятностного преобразования Хафа, участок № 1



Результат выделения лесных просек, участок № 1



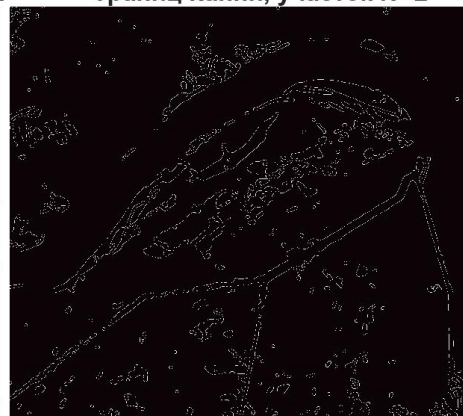
Маска класса лугов из карты классификации, участок № 2



Маска класса лугов, обработанная двоичным открытием, участок № 2



Результат детектора границ Канны, участок № 2



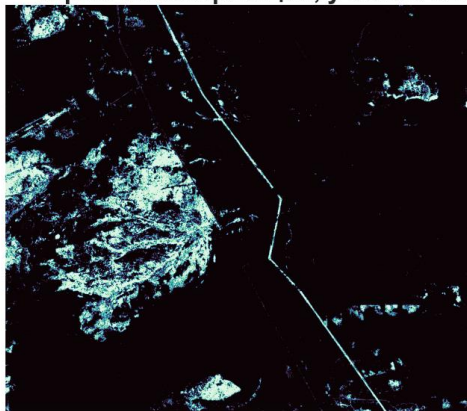
Результат вероятностного преобразования Хафа, участок № 2



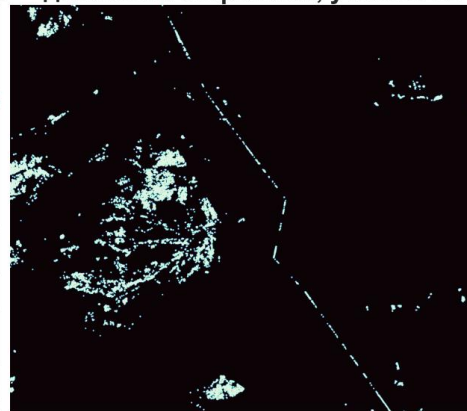
Результат выделения лесных просек, участок № 2



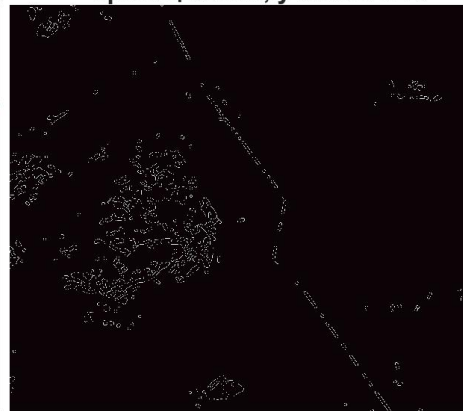
Маска класса лугов из карты классификации, участок № 3



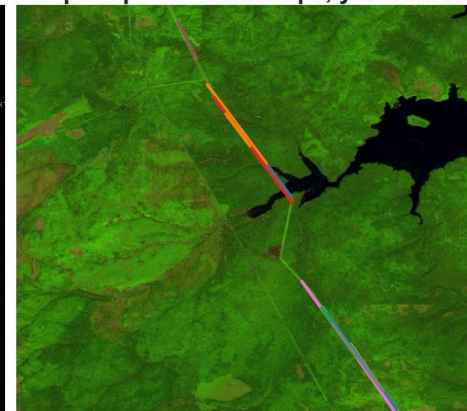
Маска класса лугов, обработанная двоичным открытием, участок № 3



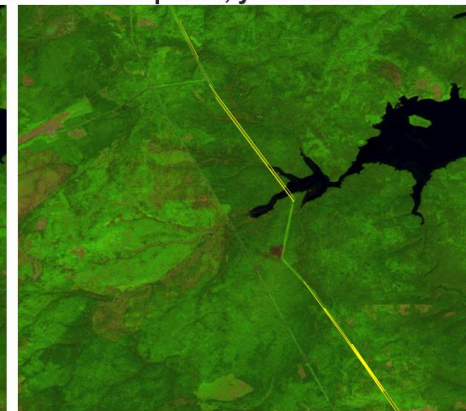
Результат детектора границ Канны, участок № 3



Результат вероятностного преобразования Хафа, участок № 3



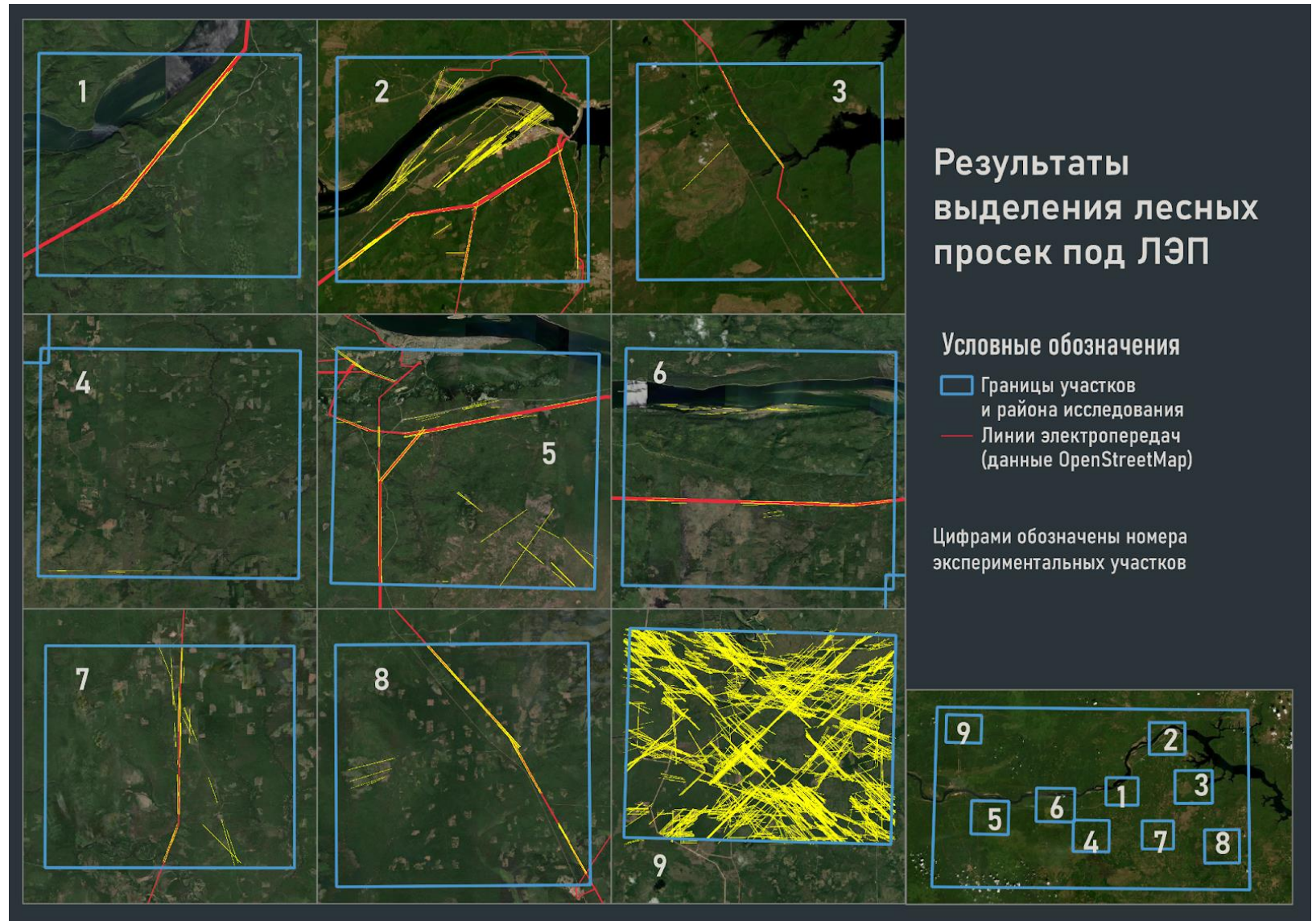
Результат выделения лесных просек, участок № 3



Тестирование алгоритма

Алгоритм был протестирован на шести тестовых участках (участки № 4-9). Для проверки выделения лесных просек использованы данные о расположении ЛЭП от OpenStreetMap, косвенно свидетельствующие о расположении лесных просек.

В целом алгоритм смог успешно выделить лесные просеки практически на всех участках. При этом у алгоритма есть закономерности в результатах – если на участке возникает городская застройка или значительное скопление дорог, то вероятность переоценки лесных просек оказывается выше.



Заключение

В рамках данного исследования выполнена задача разработки алгоритма автоматического распознавания лесных просек в районе Богучанской ГЭС, основанного на логистической регрессии и вероятностном преобразовании Хафа, а также его апробации на различных участках территории исследования. В отличие от алгоритмов, созданных в предыдущих исследованиях, предложенный метод обладает высокой интерпретируемостью и способен выделять лесные просеки под линиями электропередач, опираясь на их дешифровочные признаки. Разработанный продукт отделяет лесные просеки от других объектов на изображении, что делает его перспективным для последующих исследований лесных просек, включая выявление участков с интенсивным зарастанием лесных просек древесной растительностью.

В дальнейшем ожидается продолжение исследования в виде алгоритма, основанного на нейронных сетях, способных с более высокой точностью выделить объекты на космических изображениях. Кроме того, планируется создание единой пространственной системы мониторинга состояния лесных просек под ЛЭП, позволяющая провести постоянный мониторинг этих объектов для предотвращения их зарастания и аварий на высоковольтных линиях. Данную систему можно интегрировать с системой по мониторингу линий электропередач для комплексного мониторинга энергетической инфраструктуры.