|  |  |
| --- | --- |
|  | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ \_\_\_\_\_\_ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

КАФЕДРА \_\_\_\_СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ (ИУ5)\_\_\_\_

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

по дисциплине \_\_\_\_\_\_\_\_\_\_ Технологии машинного обучения\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

по теме\_\_\_\_\_ «Классификация областей леса по спутниковым снимкам» \_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент ИУ5-63Б **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** М.О. Пономарева

(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** А.И. Канев

(Подпись, дата) (И.О.Фамилия)

Консультант **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

(Подпись, дата) (И.О.Фамилия)

*2023 г.*

**Аннотация**

Основной задачей данной научно-исследовательской работы является разработка системы способной классифицировать, сделанные спутником Sentinel-2 на наличие сплошной вырубки (класс 1 в бинарной классификации). В ходе работы будут использоваться средства работы с данными на Python.

Также предварительно будет проводится исследование данных и подготовка их для дальнейшей работы. Из представленного набора данных будут выбираться снимки, соответствующие основной задаче, а именно выбор снимков сплошной вырубке. Будут использованы функции для нормализации яркости и контрастности изображения.

В работе будут использоваться такие модели, как RandomForest и XGBoost. А также проведена работа с фильтрами. Для оценки качества обученной модели будет использоваться метрика Dice.

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc131521360)

[1. Постановка задачи 5](#_Toc131521361)

[2. Исследование данных 7](#_Toc131521362)

[2.1 Описание и структура данных 7](#_Toc131521363)

[2.2 Подготовка данных 9](#_Toc131521364)

[ЗАКЛЮЧЕНИЕ 14](#_Toc131521365)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 15](#_Toc131521366)

# **ВВЕДЕНИЕ**

Цель:

Целью исследования является разработка модели, которая была бы способна с высокой эффективностью и точностью разделять лес на неповрежденные и поврежденные сплошной вырубкой участки.

Задачи:

- описать и оценить исследуемые данные;

- обучить модель разделять участки на поврежденные и неповрежденные;

- визуализировать результирующие маски;

- сравнить две разные модели: RandomForest и XGBoost;

- добавить фильтры;

В настоящее время ведется активная борьба с незаконной вырубкой лесов на территории России, но из-за обширности территорий постоянный мониторинг является трудной задачей. Исследования спутниковых снимков помогут решить не только эту проблему, но и помогут в своевременном обновлении карт лесополосы.

1. **Постановка задачи**

Задачей данной научно-исследовательской работы является классификация сплошной вырубки на некотором количестве спутниковых изображений. Где сплошной рубкой называется одновременная вырубка всех деревьев на отграниченной в натуре площади леса [1, С. 371].

Существует большое количество препятствий, мешающих постоянному мониторингу леса в России. Некоторые из них: отсутствие общегеографических и топографических карты, которые верно отражали бы состояние исследуемых территорий; отсутствие достоверных картографических материалов на некоторые категории лесов, раз в 10 лет составляются лесные карты, причем не всюду, велика доля ошибок при составлении лесных карт [2, С. 2]. Также до сих пор велика проблема незаконных сплошных вырубок, поиск которых необходим для восстановления лесного покрова.

Поэтому в настоящее время анализ изображений, полученных со спутниковых снимков, является наиболее актуальным, за счет своевременного получения информации о вырубках.

Согласно пресс-службе Минприроды России за первое полугодие 2022 года по отношению к аналогичному периоду 2021 года доходы лесного бюджета выросли на 10% и достигли 36,7 миллиарда рублей. Сделано это в том числе для улучшения качества мониторинга лесных территорий, их качественных и количественных характеристик.

В настоящее время в Российской Федерации с помощью анализа спутниковых снимков ведется дистанционный и лесопатологический мониторинг. Согласно словам директора ФГБУ «Рослесинфорг» ведется работа с «Роскосмос». Информация о состоянии лесов будет в оперативном режиме собираться с использованием спутниковых снимков. Съемка из космоса будет обнаруживать то или иное изменение, которое произошло в лесу - вырубка, появление свалки, карьера и так далее - а информационная система автоматически будет определять это изменение как законное или же с признаками незаконного использования лесов. На сегодняшний день контроль за незаконными рубками осуществляется натурно - то есть она должна проводиться лесничествами. Но дистанционный мониторинг на основании космических снимков уже начал внедряться [3].

Также проводятся исследования спутниковых снимков в рамках эксперимента «Дубрава», постановщиком которого является Мытищинский филиал МГТУ им. Н. Э. Баумана (Московский государственный университет леса — на момент начала разработки программы эксперимента). Его целью является отработка методов инвентаризации лесов и лесохозяйственной деятельности, определения воздействий на лесной покров природных и техногенных факторов с борта МКС. Для классификации снимков в ходе эксперимента были использованы такие методы как метод k-средних и широкий диапазон метрик, включая расстояние Евклида, косинусную меру или метод максимального правдоподобия (Байеса) [4].

Аналогичные исследования проводились также на тему выборочной вырубки лесов. Целью исследования являлась разработка метода обнаружения по данным дистанционного зондирования Земли высокого разрешения незначительных пространственных изменений на ее поверхности, вызванных антропогенным воздействием. В ходе исследований применялся попиксельный разностный метод, использующий нормализацию изображений на основе поблочного согласования гистограмм, а также построение и анализ разностных кадров на основе оценки согласованности изменения во времени одинаковых по яркости пикселов изображения [5].

Мониторинг лесных массивов с помощью космических снимков с целью контроля вырубок леса осуществлялся также с помощью исследования спектральной отражательной способности лесных насаждений и кустарников. Что в итоге позволило определить площадь вырубок различной давности, а также, используя более ранние снимки, определить типы растительности, росшей на этих местах и их площадь, как в площадных единицах, так и в процентах [2].

# **Исследование данных**

* 1. **Описание и структура данных**

Данные, используемые для исследования, взяты снимки со спутника Sentinel-2 в формате geoTiff, разрезанные на куски 256×256 пикселей. Где geoTiff — это подмножество формата хранения растровых графических изображений tiff, в котором содержится дополнительная геоинформация.

Данные предварительно были разбиты на несколько папок, а именно tiles\_256\_256\_27\_add, tiles\_256\_256\_27\_additionegor, tiles\_256\_256\_27\_additionegor\_noobjmask, tiles\_256\_256\_27\_after\_check, tiles\_256\_256\_27\_windthrows. Посмотрим в качестве примера содержание папки “tile\_256\_256\_27\_after\_check/T40VEL\_pair\_5.tif” (Рисунок 1.1):

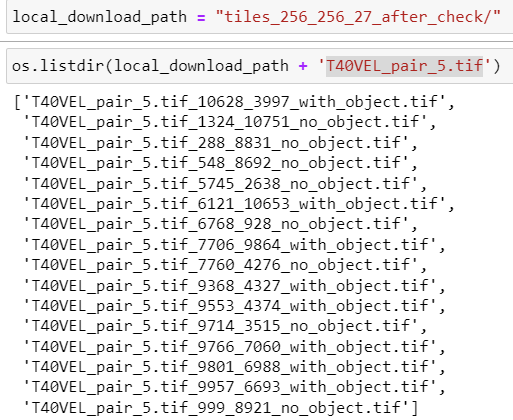


Рисунок 1.1 Папка “tile\_256\_256\_27\_after\_check/T40VEL\_pair\_5.tif”

Видим, что с помощью маркеров with\_object и no\_object снимки разделяются на те, в которых объект присутствует и в которых отсутствует, соответственно.

Объединим пути ко всем изображениям в один список для удобства при дальнейшей работе (Рисунок 1.2):

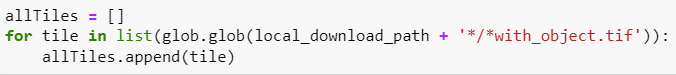


Рисунок 1.2 Объединение путей к изображениям

Выясним размерность изображений (Рисунок 1.3):

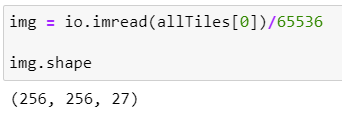


Рисунок 1.3 Характеристики изображения

Видно, что изображение является трёхмерным массивом, и помимо пикселей имеет также 27 других значений. Это каналы “в глубину”, которые учитывались спутником при съёмке. Помимо 3 каналов видимого спектра (синий, зелёный и красный) спутник учитывал также невидимые диапазоны (например, SWIR и VNIR) – всего таких каналов 13 (Таблица 1):

|  |  |  |  |
| --- | --- | --- | --- |
| Канал | Разрешение, м | Центральная длина волны, нм | Описание |
| B1 | 60 | 443 | Ultra Blue (Coastal and Aerosol) |
| B2 | 10 | 490 | Blue |
| B3 | 10 | 560 | Green |
| B4 | 10 | 665 | Red |
| B5 | 20 | 705 | Visible and Near Infrared (VNIR) |
| B6 | 20 | 740 | Visible and Near Infrared (VNIR) |
| B7 | 20 | 783 | Visible and Near Infrared (VNIR) |
| B8 | 10 | 842 | Visible and Near Infrared (VNIR) |
| B8a | 20 | 865 | Visible and Near Infrared (VNIR) |
| B9 | 60 | 940 | Short Wave Infrared (SWIR) |
| B10 | 60 | 1375 | Short Wave Infrared (SWIR) |
| B11 | 20 | 1610 | Short Wave Infrared (SWIR) |
| B12 | 20 | 2190 | Short Wave Infrared (SWIR) |

Таблица 1 – Описание съемки Sentinel-2

В изображениях совмещены пары старых снимков (с нетронутой областью леса) и новых (с изменениями)

В маске разные цифры обозначают разные классы:

* 1 – Сплошная рубка
* 2 – Проходная рубка
* 3 – Лесная дорога
* 4 – Ветровал
* 5 – Пожар
* 6 – Усыхание (повреждение)
* 7 – Выборочная рубка

Теперь определим количество снимков для каждого класса (Рисунок 1.4):



Рисунок 1.4 Количество снимков каждого класса

* 1. **Подготовка данных**

Для исследования будет достаточно выделить всего 300 спутниковых снимков. (Рисунок 2.1):

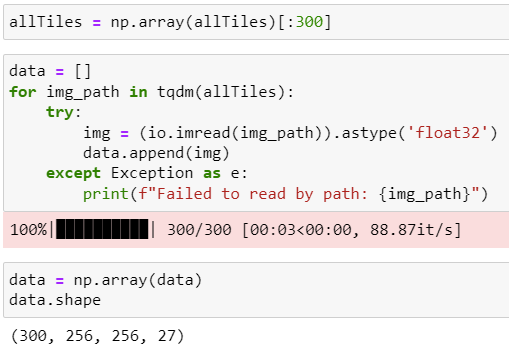


Рисунок 2.1 Выделение необходимых снимков

Проверим сколько снимков из выборки соответствуют сплошной вырубке (Рисунок 2.2):

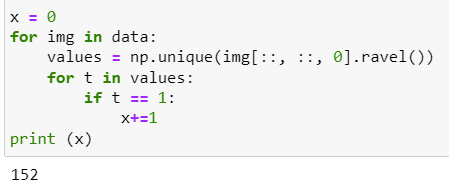


Рисунок 2.1 Количество снимков со сплошной вырубкой

Обнулим метки у снимков, которые не соответствуют выбранному классу, чтобы модель не путала разные виды лесных повреждений (Рисунок 2.3):

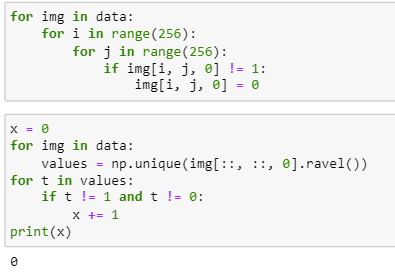


Рисунок 2.3 Обнуление неподходящих изображений

Теперь, чтобы привычнее было смотреть на RGB-изображение создадим функции для выделения старого и нового набора rgb-каналов (Рисунок 2.4):

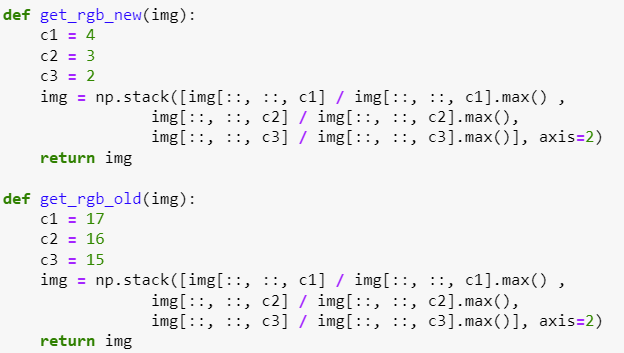


Рисунок 2.5 Выделение RGB-каналов

Введём также функцию для симпатичного отображения маски (show\_mask) и нормализации яркости и контрастности изображения (show\_img\_norm) (Рисунок 2.6):



Рисунок 2.6 Функции для отображения снимков

Теперь посмотрим на результат созданных функций (Рисунок 2.7):

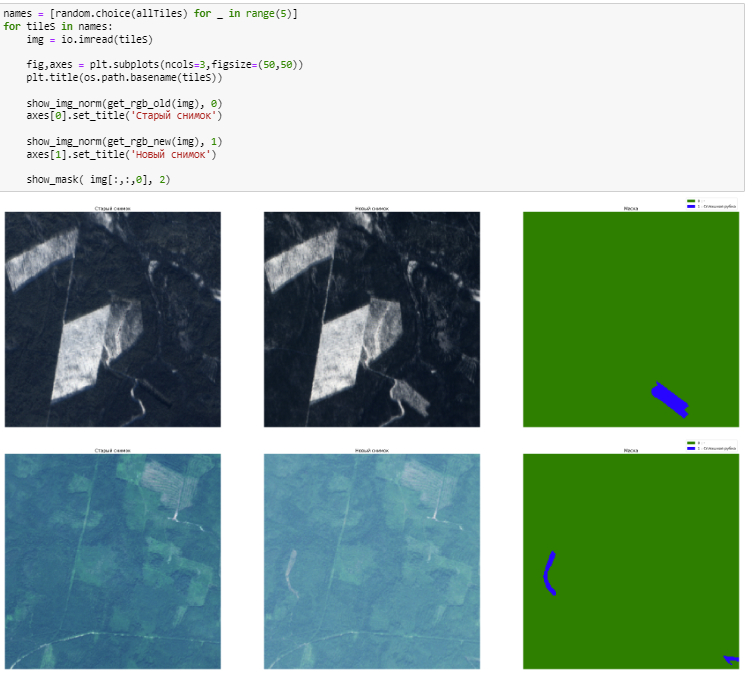


Рисунок 2.7 Результат функций show\_mask и show\_img\_norm

Для задачи выявления нарушений в состоянии лесного покрова будем использовать нормализованный разностный индекс растительности (NDVI). Высокие значения NDVI соответствуют областям, которые больше отражают ближний инфракрасный спектр, что соответствует более густой и здоровой растительности.

У спутников Sentinel 2 NIR и Red соответствуют каналам B8 и B4.

Выделим функцию get\_ndvi и посмотрим на снимки в этой комбинации каналов (Рисунок 2.8):

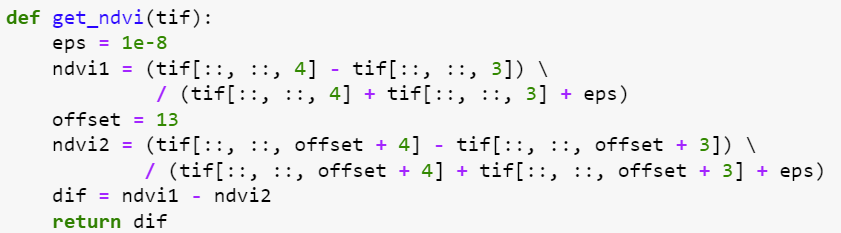


Рисунок 2.8 Функция get\_ndvi

Теперь посмотрим на результат этой функции (Рисунок 2.9):

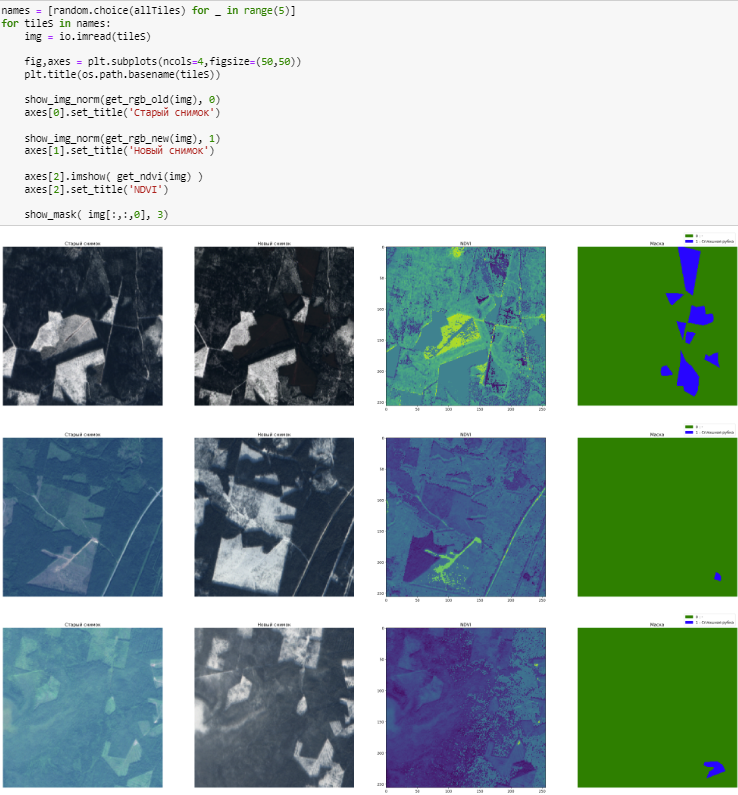


Рисунок 2.9 Результаты с функцией get\_ndvi

Разделим данные на тренировочную и тестовую выборки (Рисунок 2.10):



Рисунок 2.10 Разделение на тестовую и тренировочную выборки

На вход модели будем подавать все 26 исходных каналов (13 старых снимков в разных спектрах, 13 новых), а также NDVI индекс. Отделим 25 изображений для дальнейшего тестирования модели (Рисунок 2.11):

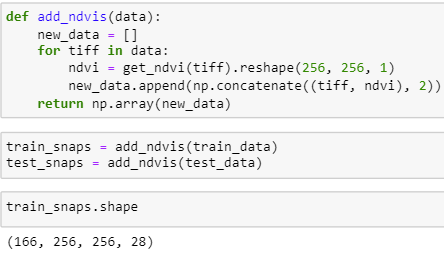


Рисунок 2.11 Выделение 25 снимков для тестирования

На вход классических МЛ-моделей должен подаваться одномерный массив. Так что «вытянем» наши изображения 256х256х28 с помощью reshape (Рисунок 2.12):

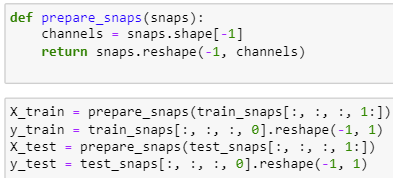


Рисунок 2.12 Применение reshape

На этом этапе закончим подготовку данных.

# **Сравнение моделей RandomForest и XGBoost**

* 1. **RandomForest**

Будем использовать RandomForest для задачи бинарной классификации, чтобы определить наличие или отсутствие повреждения (Рисунок 3.1):

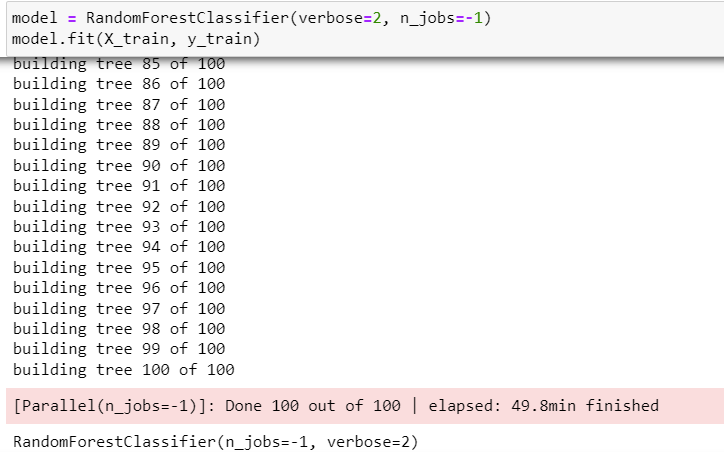


Рисунок 3.1 Применение RandomForest

Импортируем joblib (Рисунок 3.2):

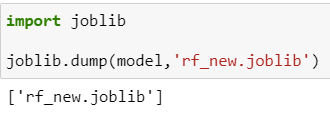


Рисунок 3.2 Импорт joblib

* 1. **Оценка качества модели RandomForest**

Для оценки эффективности модели рассчитаем Dice коэффициент, который показывает попиксельное соотношение между прогнозируемой маской и соответствующей ей истиной. Коэффициент Dice — это двойная площадь их пересечения, деленная на общее количество пикселей в обоих изображениях (Рисунок 3.3):

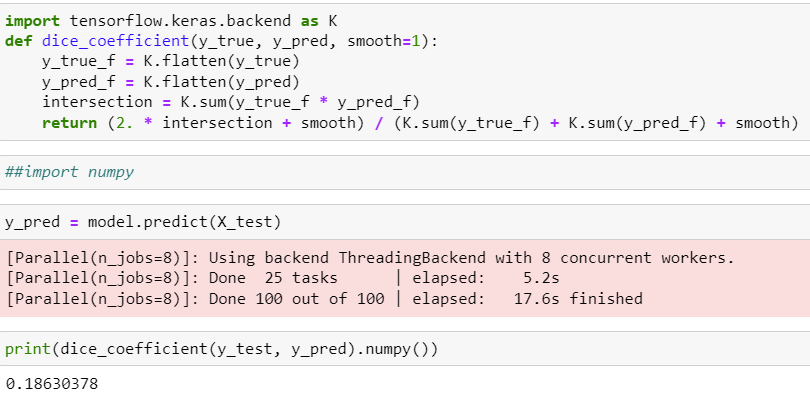


Рисунок 3.3 – Оценка RandomForest

* 1. **XGBoost**

Аналогично можно выполнить обучение другой ML-модели. Мы уже разделили и подготовили датасет в части подготовки данных, так что здесь нам остаётся только обучить модель (Рисунок 3.4):

# **ЗАКЛЮЧЕНИЕ**

В ходе работы были выполнены следующие поставленные ранее задачи:

- Описана структура данных (Был определен формат и размерность изображений, определено количество и размерность их каналов, а также были выделены изображения, соответствующие поставленной задаче, а именно сплошной вырубке)

- Созданы функции для удобного отображения (Такие как, функции для отображения изображений в формате RGB, вывод индексов NDVI, маска изменений лесных участков)

- Осуществлена выборка данных (Выделен набор данных такой, чтобы количество снимков с изображениями сплошной выборки в нем, было оптимальным для решения поставленных задач)

- Проведена очистка данных (Все классы кроме необходимого были удалены)

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Эйтинген Г. Р. Лесоводство. - 6-е изд. - М: Сельхозгиз, 1959. - 416 с.
2. Казарян М.Л., Казарян М.Л., Шахраманьян М.А., Шахраманьян М.А., Шахраманьян М.А. МОНИТОРИНГ ЛЕСНЫХ МАССИВОВ С ПОМОЩЬЮ КОСМИЧЕСКИХ СНИМКОВ-КОНТРОЛЬ ВЫРУБОК ЛЕСА // Современные проблемы науки и образования. – 2015. – № 1-1.
3. Российские леса получат тотальный контроль из космоса // ВЗГЛЯД URL: https://vz.ru/society/2019/9/27/999917.html (дата обращения: 30.03.2022).
4. Батырев Ю.П., Поярков Н.Г., Чумаченко С.И., Беляев М.Ю., Сармин Э.Э., Есаков А.М., Черемисин М.В. Особенности проведения космического эксперимента «Дубрава» с борта Российского сегмента МКС // Лесной вестник / Forestry Bulletin, 2022. Т. 26. № 1. С. 135–142. DOI: 10.18698/2542-1468-2022-1-135-142
5. Борзов С.М., Потатуркин О.И. ОБНАРУЖЕНИЕ ВЫБОРОЧНЫХ РУБОК ЛЕСА ПО ДАННЫМ ДИСТАНЦИОННЫХ ИЗМЕРЕНИЙ ВЫСОКОГО ПРОСТРАНСТВЕННОГО РАЗРЕШЕНИЯ // ИССЛЕДОВАНИЕ ЗЕМЛИ ИЗ КОСМОСА. - 2014. - №4. - С. 87-93.