



УДК 528.8; 581.9

МРНТИ 34.29.35; 36.33.27

DOI 10.37238/2960-1371.2960-138X.2025.98(2).87

¹Денисова Н.Ф., ^{2,3}Жусупова Г.К.*, ¹Петрова О.А., ²Чепашев Д.В.,
¹Даумова Г.К., ³Темирбаева К.А.

¹Восточно-Казахстанский технический университет им. Д. Серикбаева,
Усть-Каменогорск, Казахстан

²ТОО «Институт Ионосферы», Алматы, Казахстан

³Казахский Национальный Университет им. Аль-Фараби,
Алматы, Казахстан

E-mail: ndenisova@edu.ektu.kz, zhussupova51@gmail.com*, opetrova@edu.ektu.kz,
d.chepashev@ionos.kz, gdaumova@edu.ektu.kz, kamshat.temirbayeva@gmail.com

КАРТИРОВАНИЕ РАСТИТЕЛЬНОСТИ ВОСТОЧНО-КАЗАХСТАНСКОЙ ОБЛАСТИ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ НА ОСНОВЕ СПУТНИКОВЫХ СНИМКОВ LANDSAT

Аннотация. Целью данного исследования является разработка подхода к картографированию растительности Восточно-Казахстанского региона с использованием методов машинного обучения на основе спутниковых снимков Landsat 8. В статье рассматриваются этапы предобработки данных, выбор оптимальных алгоритмов классификации и оценка их эффективности для различных типов растительности. Процесс предварительной обработки и классификации данных выполнялись на облачной платформе Google Earth Engine. Помимо 17 вспомогательных слоев, содержащих информацию о рельефе местности, в процессе сбора данных были сформированы обучающие выборки, распределенные по 10 классам: широколиственные деревья, хвойные деревья, травянистая растительность, кустарники, открытые земли, сельскохозяйственные угодья, водно-болотные территории, застроенные земли, снег и лед, водные объекты.

В рамках исследования были применены и сравнены различные алгоритмы машинного обучения, такие как метод опорных векторов (SVM), деревья решений (CART), градиентный бустинг деревьев (GTB) и случайный лес (RF). Алгоритмы GTB и RF показали наилучшие результаты с Общей точностью 92% и коэффициентом Каппа 91%, что свидетельствует о высоком качестве интеграции данных дистанционного зондирования с методами машинного обучения. Полученные результаты исследования подчеркивают потенциал методов машинного обучения для мониторинга состояния экосистем, планирования природоохранных мероприятий, управления природными ресурсами и чрезвычайными ситуациями в регионе.

Ключевые слова: Восточно-Казахстанская область; Landsat; Машинное обучение; Классификация растительности; Google Earth Engine; Random Forest.

Введение

Растительный покров является одним из ключевых компонентов окружающей среды, играющих значимую роль в поддержании биоразнообразия, регулировании климата и обеспечении устойчивости природных ландшафтов. В условиях



возрастающего антропогенного воздействия и изменения климата мониторинг и картографирование растительности особенно актуальна. Восточно-Казахстанский регион, обладающий уникальным разнообразием природных зон — от степей до горных лесов, — представляет собой важную территорию для исследования экосистем и мониторинга изменений в растительном покрове. Традиционные методы картографирования растительности, основанные на полевых исследованиях, являются трудоемкими и ограничены в масштабах применения. Современные технологии дистанционного зондирования Земли (ДЗЗ) в сочетании с методами машинного обучения открывают новые возможности для автоматизированного анализа и классификации растительного покрова [1, 2, 3]. Спутниковые снимки, в том числе данные спутника Landsat 8, предоставляют многолетние временные ряды, что позволяет отслеживать изменения в растительном покрове с высокой точностью и детализацией [1, 4].

В последние годы методы машинного обучения, включая опорные векторные машины (SVM), случайный лес (RF), градиентный бустинг деревьев (GTB) и деревья решений (CART) стали широко использоваться для картирования растительности и классификации типов землепользования на основе данных дистанционного зондирования. В частности, RF стабильно демонстрирует лучшие результаты в различных исследованиях, обеспечивая высокую общую точность [1, 3]. SVM также показывает хорошие результаты, особенно при работе с разнородными данными [5]. GTB доказал свою эффективность при классификации сельскохозяйственных угодий и лесных массивов, достигая точности свыше 84%. CART проявил себя особенно хорошо в классификации растительности при выделении участков открытой почвы, обеспечивая точность до 98% [6].

Выбор оптимального алгоритма зависит от множества факторов, включая сложность данных, вычислительные ресурсы и конечные цели. Хотя эти методы продемонстрировали высокую эффективность, их производительность зависит от таких параметров, как пространственное разрешение снимков, количество используемых признаков и тип исходных данных [6]. Проведенные исследования подтверждают эффективность машинного обучения в задачах картографирования растительности и прогнозирования изменений земного покрова на основе спутниковых данных.

Материалы и методы исследования

Методология исследования охватывает несколько ключевых этапов: (1) сбор и начальная обработка входных данных, (2) классификация изображений на основе машинного обучения, (3) оценка точности и определение эффективного алгоритма (Рисунок 1). Начальные этапы исследования включают в себя процедуры обработки и сбора данных для повышения их качества: облачная фильтрация, медианные значения спутниковых изображений и добавление вспомогательных данных. Затем справочные данные делятся в соотношении 70/30 для обучения и проверки модели машинного обучения. В следующем этапе проводится анализ алгоритмов SVM, CART, GTB и RF для выбора наиболее точного и эффективного метода. Оценка точности классификации с учетом Общей точности и коэффициента Каппа для оптимального анализа растительного покрова Восточно-Казахстанской области является завершающим этапом исследования.



Рисунок 1 – Схема метода исследования
Область исследования

Восточно-Казахстанская область (ВКО) расположена в восточной части Казахстана, в пределах 48°00'-52°00' северной широты и 79°00'-86°00' восточной долготы. Она граничит с Алтайским краем России на севере, Монголией на востоке, Китаем на юго-востоке и занимает территорию в 97,8 тыс. кв. км (3,6% от общей территории Казахстана).

Климат региона резко континентальный, отличается холодными зимами и жарким летом, количество осадков могут достигать 1000–1500 мм в горах, особенно на склонах Алтайских гор, где наблюдаются наиболее влажные условия. В регионе протекают крупные реки, такие как Иртыш, Обь и Ульба, а также расположены значимые озера — Маркаколь, Зайсан, Алаколь и Сасыкколь. Ландшафт региона отличается большим разнообразием: значительную часть территории занимают горные системы Алтая, Калбы и Саур-Тарбагатай, где высоты достигают 3000–4500 метров, включая высочайшую вершину Алтая — гору Белуху. Ландшафты варьируются от равнинных степей и пустынных территорий до горных лесов, альпийских лугов и скалистых вершин [7].

Сбор входных данных

Это исследование было проведено с использованием изображений, полученных со спутника Landsat 8 второго уровня, доступные на облачной платформе Google earth Engine. Landsat 8 Level 2 (Collection 2, Tier 1) – это набор спутниковых данных с разрешением 30 м, предоставляемых USGS, включающий атмосферно-корректированные значения отражательной способности поверхности и температуры земли. В составе изображений содержатся пять спектральных каналов в диапазоне видимого и ближнего инфракрасного излучения (VNIR), два канала коротковолнового инфракрасного спектра (SWIR) и один тепловой инфракрасный канал (TIR) [8]. В ходе исследования на этапе обработки данных был выбран период с июня по август 2024 года для региона ВКО, поскольку в это время вероятность образования облаков и снега минимальна. К сформированному набору из 110 изображений с каналами SR_B2 - SR_B7 была



применена фильтрация облаков с использованием канала «pixel_qa» в снимках Landsat 8 Level-2, что позволило эффективно устранить пиксели с плотными и перистыми облаками, а также с их тенями на поверхности земли [4]. Затем изображения были объединены в единую мозаику и рассчитаны медианные значения всех изображений для устранения возможных аномалий в значениях и влияния шумов (Рисунок 1).

Применение спектральных индексов в качестве вспомогательных данных расширяет массив получаемых данных, что способствует более точному и детализированному процессу классификации [1, 3, 9]. Уравнения в таблице 1 использовались для оценки показателей NDVI, SAVI, EVI, GNDVI, BSI и NDBI.

Таблица 1 – Формулы спектральных индексов

Индексы	Формула
NDVI	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$
EVI	$2.5 * (\text{NIR} - \text{Red}) / (\text{NIR} + 6 * \text{Red} - 7.5 * \text{Blue} + 1)$
GNDVI	$(\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green})$
SAVI	$1.5 * (\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.5)$
BSI	$(\text{SWIR1} + \text{Red}) - (\text{NIR} + \text{Blue}) / (\text{SWIR1} + \text{Red}) + (\text{NIR} + \text{Blue})$
NDBI	$(\text{NIR} - \text{SWIR1}) / (\text{NIR} + \text{SWIR1})$

Индекс NDVI принимает значения в диапазоне от -1 до 1: чем выше значение NDVI, тем более густая и здоровая растительность присутствует на участке, а максимальные показатели обычно соответствуют лесным массивам. Низкие значения указывают на редкую растительность или открытые грунты [9]. В отличие от NDVI, улучшенный вегетационный индекс EVI, обладает преимуществами при анализе густой растительности, поскольку он снижает эффект насыщения и учитывает влияние атмосферных факторов. Кроме того, EVI компенсирует шум крон деревьев, используя корректирующие параметры: коэффициент L (стандартное значение - 0,5) помогает учитывать фон растительности, C1 (стандартное значение - 6) и C2 (стандартное значение - 7,5) используется для компенсации атмосферного рассеяния, а синяя спектральная полоса Blue вводится для учёта атмосферных влияний [9]. GNDVI представляет собой усовершенствованную версию NDVI, в которой вместо красного спектра используется отражательная способность в зеленом диапазоне, что позволяет более точно оценивать содержание хлорофилла в растительности [4]. В исследовании также применен индекс SAVI (Soil-Adjusted Vegetation Index) предложенный Huete [10] как метод коррекции спектральных данных для снижения влияния почвенного покрова на расчёт вегетационных индексов. SAVI включает в себя коэффициент L (стандартное значение - 0,5) учитывающий отражательную способность почвы. SAVI особенно полезен при анализе территорий с редкой растительностью, где значения NDVI могут быть неточны из-за смешивания сигналов почвы и растений. Для выделения застроенных и незастроенных территорий в исследовании были рассчитаны индексы BSI и NDBI. Индекс BSI применяется для идентификации участков с оголённой почвой и мониторинга их сезонных изменений, тогда как NDBI используется для определения урбанизированных зон. Совместное применение этих индексов позволяет повысить точность выявления застроенных территорий [1].

В регионах с горным ландшафтом особенности рельефа значительно влияют на отражательную способность поверхности из-за крутых склонов. Для анализа этого влияния была использована 30-метровая цифровая модель рельефа от миссии Shuttle Radar Topography Mission (SRTM DEM), с помощью которой были получены данные о наклоне (Slope) и ориентации склонов (Aspect). Эти параметры помогают лучше понять,

как рельеф влияет на спектральные характеристики поверхности и могут быть использованы для более точных геопространственных анализов [9, 11].

В исследовании применен набор данных Canopy Height, который представляет собой карты высоты древесного покрова высокого разрешения, созданные с использованием изображений RGB на основе свёрточных нейросетей. Обучение модели, согласно исследованиям [12] проводилось на данных с аэрофотолидара, а также с применением спутниковых снимков Maxar (2017–2020). Использование Canopy Height повышает точность классификации, учитывая пространственное распределение и высотные характеристики растительности. Также в состав входных переменных входит продукт глобального земельного покрова Land Cover с пространственным разрешением 100 м, разработанный в рамках программы Copernicus Global Land Service (CGLS). Этот продукт предоставляет карты земельного покрова 2015-2019 года, построенные на основе спутниковых данных PROBA-V и Sentinel-2. Классификация земельного покрова основана на системе LCCS (Land Cover Classification System) и включает 23 класса, сгруппированных в три уровня детализации [13].

На основе вышеперечисленных данных были созданы 17 входных переменных включающие слои спектральных каналов, спектральных индексов, топографических индексов и вспомогательных данных.

Сбор обучающих данных

Обучающие данные для классификации растительности были собраны на основе сравнении базовой подложки Google Earth Engine (GEE) и спутниковых изображений Landsat 8. Исходные данные были спроецированы на исследуемую территорию и разделены в соотношении 70/30 на обучающую и проверяющую выборки. Такой метод позволяет сбалансировать процесс обучения модели и её последующую проверку, обеспечивая объективную оценку точности классификации [1, 14]. Для определения типа растительности в основном использовался слой Land Cover с помощью которого созданы 1041 обучающих и проверяющих данных. Эти векторные данные разделены на 10 классов земного покрова, в том числе из растительного покрова были добавлены: хвойные деревья, широколиственные деревья, кустарники, травянистая растительность, водно-болотные угодья и сельскохозяйственные земли.

Методы машинного обучения

Для классификации растительности на территории исследования применялись различные методы машинного обучения, включая метод опорных векторов (SVM), деревья решений (CART), градиентный бустинг деревьев (GTB) и алгоритм случайного леса (RF). Эти методы позволяют автоматически анализировать спутниковые снимки, выявляя различные классы земного покрова на основе их спектральных характеристик [2, 4, 5, 6].

Метод SVM используется для разделения данных на несколько классов путём нахождения оптимальной гиперплоскости, обеспечивающей наибольшее расстояние между точками различных классов и линией разграничения. Это делает его особенно эффективным при работе с данными, где границы между классами сложны для традиционных методов [2].

Алгоритмы CART и GTB, основанные на деревьях решений, выполняют последовательное разбиение данных, строя деревья строится по определённым параметрам, таким как критерии роста и обрезки, до тех пор, пока не будет достигнуто оптимальное распределение классов [6]. CART строит одно дерево решений, тогда как GTB использует ансамбль деревьев, обучаемых последовательно для исправления ошибок предыдущих, что делает его более точным, но требующим настройки параметров. В методе Random Forest каждое дерево обучается независимо, а итоговый класс пикселя определяется



голосованием среди всех деревьев, что позволяет снизить вероятность ошибок и переобучения, характерных для одиночных деревьев решений [5, 6].

Для выполнения классификации использовалась облачная платформа Google Earth Engine (GEE), предоставляющая мощные инструменты для геопространственного анализа. Код разрабатывался в JavaScript через встроенный редактор, который позволяет обрабатывать большие объёмы данных и визуализировать результаты.

Оценка точности и важности переменных

В данном исследовании точность классификации оценивалась с помощью метрик Общая точность и коэффициент Каппа, рассчитанных через `confusionMatrix.accuracy` и `confusionMatrix.kappa` в Google Earth Engine. Оценка точности — ключевой этап классификации, определяющий корректность распределения пикселей по классам.

Общая точность отражает долю правильно классифицированных образцов, а коэффициент Каппа измеряет соответствие классифицированных и обучающих данных с учётом случайных совпадений. Значения выше 79% считаются высокой точностью, а Каппа варьируется от -1 (отрицательная согласованность) до 1 (почти полное совпадение) [6, 14].

Для анализа вклада входных переменных применён метод MDI (Mean Decrease Impurity) в GEE, оценивающий влияние признаков на снижение неопределённости в деревьях алгоритмов. Значительное уменьшение неопределённости, связанное с конкретным признаком, указывает на его более высокую важность в процессе классификации. Этот метод помогает выявить ключевые факторы, влияющие на точность модели [15].

Результаты исследования

Карты типов растительности, созданные с помощью алгоритмов машинного обучения на основе 17 наборов входных данных, представлены на рисунке 2. В итоге было распределено 10 классов земного покрова: широколиственные деревья, хвойные деревья, травянистая растительность, кустарники, открытые земли, сельскохозяйственные земли, водно-болотные угодья, застроенные земли, снег и лед, водные объекты. Участки, отмеченные как No Data, включают пиксели, исключенные из анализа в процессе удаления облачного покрова, что составляет 0,437% от общей площади исследования.

Визуальный анализ показывает, что все алгоритмы дают схожее пространственное распределение классов, но отличаются в деталях. Значительную схожесть в результатах алгоритмов можно увидеть в таких классах как водные объекты, снег и лед, открытые земли. Однако для других классов выявлена существенная разница в занимаемой ими площади. Например, алгоритм SVM определяет сельскохозяйственные земли вместо открытых земель вокруг озера Зайсан и допускает ошибки в классификации широколиственных и хвойных деревьев, присваивая большее количество пикселей широколиственным деревьям, что привело к заметному отличию результата от других алгоритмов. В алгоритме CART кустарники и застроенные земли занимают больше площади, в сравнении с результатами остальных алгоритмов, что свидетельствует о более выраженной неоднозначности в классификации. Также путаница между классами водно-болотных угодий и водных объектов, застроенных земель и открытой земли замечается в результатах всех алгоритмов, и обуславливается схожестью в спектральных характеристиках значений этих объектов.

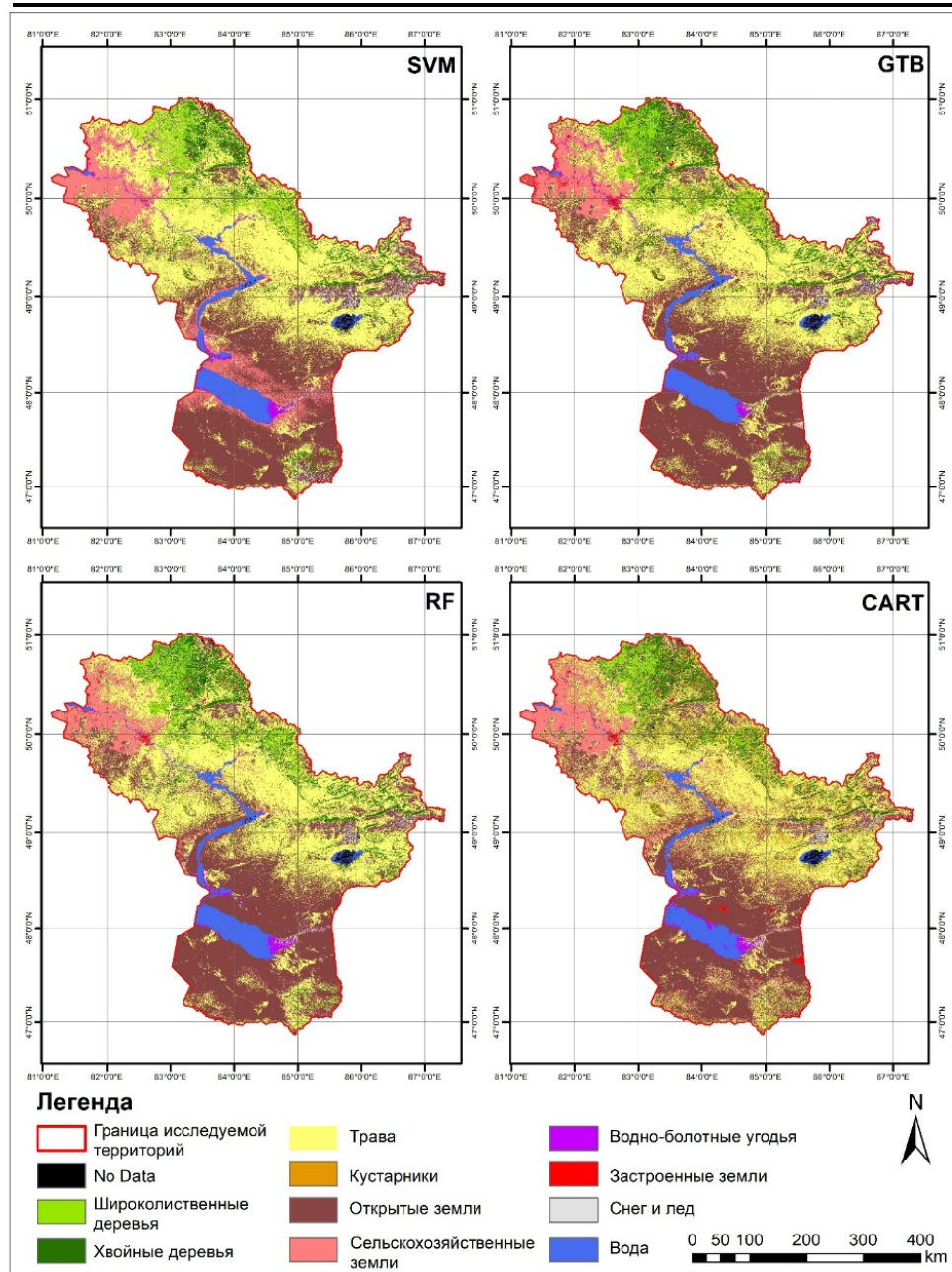


Рисунок 2 – Результат классификаций растительного покрова на территорию ВКО

На рисунке 3 представлены результаты классификации типов земного покрова, полученные с использованием алгоритмов и спутниковых изображений Landsat 8 для двух различных участков (А и В) Восточно-Казахстанской области в масштабе 1:50 000. В верхнем ряду (А) заметны различия в определении лесных массивов. Метод SVM демонстрирует больше неточностей при распределении травянистой растительности и водно-болотных угодий, а также склонен переоценивать площадь широколиственных деревьев, уменьшая долю хвойных. Алгоритмы GTB и RF обеспечивают более детализированное разделение лесных типов, чётко отображая рельефные особенности,

такие как гребни и долины, при этом RF фиксирует больше хвойных деревьев, тогда как GTB чаще классифицирует траву как кустарники. Метод CART характеризуется большим количеством пересечений между классами, что приводит к ошибочному отнесению некоторых территорий к кустарникам или водно-болотным угодьям.

В нижнем ряду (В) видны различия в классификации водоемов, сельскохозяйственных земель и урбанизированных территорий. SVM завышает долю водно-болотных угодий и практически не выявляет застроенных земель, CART также увеличивает площадь водно-болотных угодий, что может свидетельствовать о переоценке зон этих объектов. Наиболее точные результаты классификации показывает RF, особенно в выделении границ водоемов и урбанизированных зон. В то же время GTB допускает ошибки в распределении застроенных земель и кустарников.

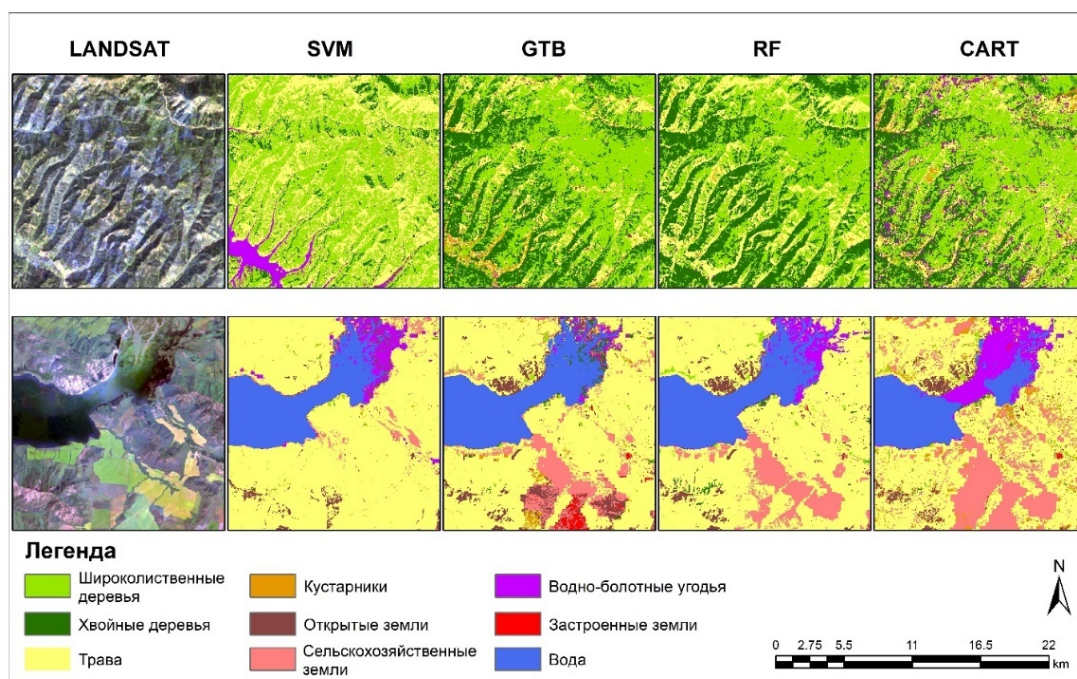


Рисунок 3 – Результат классификаций растительного покрова на территорию ВКО в масштабе 1:50 000

В таблице 2 показано процентное распределение классов земного покрова по четырем методам классификации (SVM, GTB, RF, CART).

Таблица 2 – Площади классов по четырем алгоритмам (в %)

Классы \ Методы	SVM	GTB	RF	CART
Широколиственные деревья	4.535	4.308	3.999	3.682
Хвойные деревья	6.603	8.352	8.922	7.54
Травянистая растительность	38.47	36.794	37.996	31.461
Кустарники	1.034	1.884	0.993	7.089
Открытая почва	31.692	35.687	34.754	31.554
Сельскохозяйственные земли	10.042	5.911	6.161	10.28



Методы	SVM	GTB	RF	CART
Водно-болотные угодья	1.081	0.603	1.167	2.278
Застроенная земля	0.042	0.547	0.328	0.61
Снег и лед	1.586	0.502	0.428	0.4
Вода	4.473	4.97	4.81	4.665

Согласно таблице 2 наибольшие различия наблюдаются в оценке травянистой растительности (от 31.461% в CART до 38.47% в SVM) и кустарников (CART завышает их до 7.089%, тогда как остальные методы дают менее 2%). Открытая почва наиболее выражена в GTB (35.687%) и RF (34.754%), а водно-болотные угодья достигают максимального значения в CART (2.278%) при минимуме у GTB (0.603%). Водные объекты классифицируются стабильно (4.473%–4.97%), тогда как снег и лед учитывая сезонные особенности, варьируются от 1.586% (SVM) до 0.428% (RF).

На таблице 3 показаны результаты сравнения методов классификации (SVM, GTB, RF, CART) по двум показателям: Общей точности (ОТ) и коэффициента Каппа (КК).

Таблица 3 – Результаты оценки точности для четырех алгоритмов (в %)

Методы	SVM	GTB	RF	CART
Общая точность	79.7	92.1	92.4	88.8
Коэффициент Каппа	75.9	90.7	91.1	86.9

По результатам классификации метод RF и GTB показали достаточно схожие результаты, демонстрируя наивысшую точность 92,4% - ОТ и 91,1% - КК, 92,1% - ОТ и 90,7% - КК соответственно. Алгоритм CART показывает немного более низкие результаты 88.8% - ОТ и 86.9% - КК, а SVM имеет самые слабые показатели 79.7% - ОТ и 75.9% - КК, что указывает на его меньшую точность и согласованность.

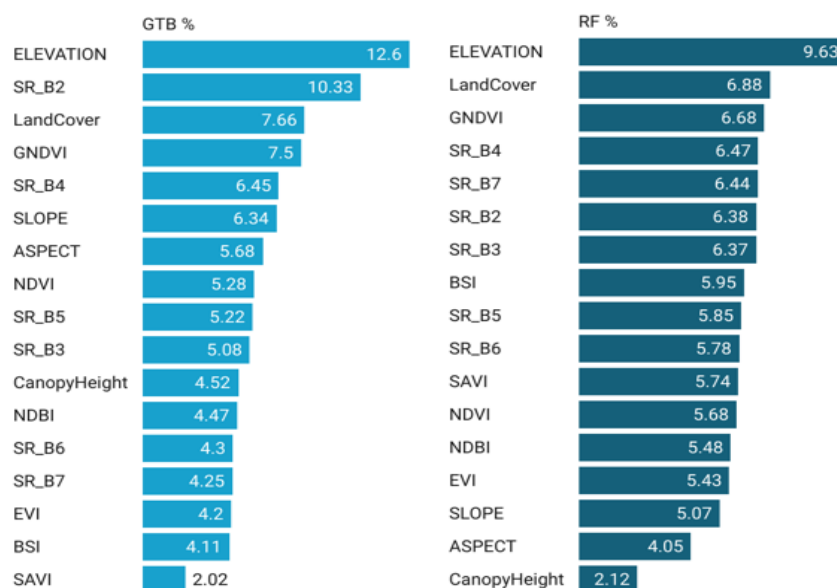


Рисунок 4 – Показатели важности входных параметров

График на рисунке 4 иллюстрирует значимость различных слоев при классификации растительного покрова с применением двух высокоточных алгоритмов по результатам оценки точности – Gradient Tree Boosting (GTB) и Random Forest (RF). Оба метода выделяют высоту местности (ELEVATION) как наиболее влиятельную переменную (GTB – 12.6%, RF – 9.63%). Среди спектральных каналов особенно заметны каналы SR_B2 и SR_B4 у алгоритма GTB, а SR_B4 и SR_B7 у алгоритма RF. Индекс GNDVI также играет значительную роль в классификации, однако другие спектральные индексы – SAVI, NDVI, NDBI, EVI, BSI демонстрируют заметные различия по степени влияния на результаты. Переменные рельефа – SLOPE и ASPECT имеют высокую значимость в алгоритме GTB, чем в RF. Также слой CanopyHeight (высота крон деревьев) значительно важнее для GTB (4.52%), чем для RF (2.12%). Распределение значимости переменных показывает, что GTB более чувствителен к топографическим характеристикам, тогда как RF уделяет больше внимания спектральным особенностям.

Обсуждение

Основная цель данного исследования заключалась в применении различных методов машинного обучения для выбора наилучшего подхода в картировании растительного покрова. В результате были получены актуальные данные о растительности на 2024 год для Восточно-Казахстанской области на основе данных спутников Landsat 8. Согласно полученным результатам, методы ансамблевых деревьев решений RF и GTB, показали лучшие результаты, тогда как SVM и CART оказались менее точными. RF и GTB более точно выделяют растительность, алгоритмы SVM и CART имеют более выраженные погрешности в определении границ классов. Сложности в различении кустарников и деревьев, а также травянистых и сельскохозяйственных участков объясняются схожими спектральными характеристиками этих классов, что и замечается в результатах алгоритмов SVM и CART [2, 3, 4].

Входные параметры были выбраны с учетом предыдущих исследований [1, 3, 6, 14] в области классификации с использованием машинного обучения. Включение таких характеристик, как высота - Elevation, значительно повысило точность классификации, что связано с разнообразием рельефа региона, где леса расположены в горных районах, а степи занимают равнинные территории. Важным фактором стало использование слоя LandCover, который содержит различные типы земного покрова, способствующие улучшению точности классификации. Анализ значимости спектральных индексов показал, что индекс GNDVI, являющийся модификацией NDVI, как и в предыдущих исследованиях [3, 4], существенно повлиял на результаты в обоих методах, а индекс BSI был особенно эффективен в RF для точного выделения сельскохозяйственных и открытых земель. Другие индексы оказывали разное влияние на классификацию в алгоритмах RF и GTB. Влияние переменных подтвердило, что RF имеет преимущество перед GTB благодаря более эффективному использованию спектральных диапазонов, что повысило точность классификации.

Основным ограничением исследования стали возможные искажения спутниковых данных земного покрова в сложных рельефах и зонах с объектами с похожими спектральными характеристиками. Для повышения точности классификации потребуются данные полевых исследований и актуальная наземная информация. Также фильтрация летних снимков с малым уровнем облачности обеспечила стабильные условия освещенности, но ограничила охват всей территории, что привело к появлению пикселей с данными No Data.

Заключение

В данном исследовании предложена методика картирования растительности с использованием четырех различных классификаторов: метод опорных векторов (SVM),

деревья решений (CART), градиентный бустинг деревьев (GTB) и случайный лес (RF), применяемых к данным спутников Landsat 8. Для классификации были задействованы 17 переменных, включая спектральные каналы Landsat 8, спектральные индексы, основанные на тех же данных, топографические слои рельефа из данных SRTM, слой земного покрова LandCover, а также слой высоты кроны деревьев Canopy Height.

По результатам проведенных работ все классификаторы продемонстрировали выше среднего уровень точности, при этом наибольшую эффективность показали алгоритмы RF (92,4% - ОТ и 91,1% - КК) и GTB (92,1% - ОТ и 90,7% - КК). Эти результаты подтверждают высокую надежность и успешность применения методов машинного обучения для картирования растительности в регионах как Восточно-Казахстанская область, где присутствуют морфологически разнообразные ландшафты.

Данное исследование может внести значительный вклад в науку, связанную с устойчивым развитием, оценкой рисков лавинообразования и влиянием экологических изменений на растительность, обеспечивая детальный анализ растительного покрова с помощью машинного обучения и спутникового дистанционного зондирования. Методологию можно применить к масштабированию на весь Казахстан, с подтверждением результатов полевыми исследованиями и использованием данных спутников и дронов. Сочетание лазерного сканирования, снимков с дронов, спутниковых данных и знаний экспертов позволит повысить точность классификации и предоставить дополнительную информацию о различных характеристиках растительности.

Благодарности

Исследование выполнено при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан (ИРН BR21882022 «Исследование лавинной активности в Восточно-Казахстанской области для разработки систем мониторинга и научного обоснования их размещения»).

ЛИТЕРАТУРА

- [1] Tassi, A., Gigante, D., Modica, G., Di Martino, L., & Vizzari, M. (2021). Pixel-vs. Object-based landsat 8 data classification in google earth engine using random forest: The case study of maiella national park. *Remote sensing*, 13(12), 2299. <https://doi.org/10.3390/rs13122299>
- [2] Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 18. <https://doi.org/10.3390/s18010018>
- [3] Mohammadpour, P., Viegas, D. X., & Viegas, C. (2022). Vegetation mapping with random forest using sentinel 2 and GLCM texture feature—A case study for Lousã region, Portugal. *Remote Sensing*, 14(18), 4585. <https://doi.org/10.3390/rs14184585>
- [4] Junaid, M., Sun, J., Iqbal, A., Sohail, M., Zafar, S., & Khan, A. (2023). Mapping lulc dynamics and its potential implication on forest cover in malam jabba region with landsat time series imagery and random forest classification. *Sustainability*, 15(3), 1858. <https://doi.org/10.3390/su15031858>
- [5] Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International journal of remote sensing*, 39(9), 2784-2817. <https://doi.org/10.1080/01431161.2018.1433343>
- [6] Ouma, Y., Nkwae, B., Moalafhi, D., Odirile, P., Parida, B., Anderson, G., & Qi, J. (2022). Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban LULC features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 681-689. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-681-2022>

- [7] Сайт акимата Восточно-Казахстанской области. [Электронный ресурс]. – Режим доступа: <https://www.gov.kz/memleket/entities/akimvko>
- [8] Сайт каталога данных Google Earth Engine. [Электронный ресурс]. – Режим доступа: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2?hl=ru
- [9] Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: a case study in high-density cypress forest. *Sensors*, 7(11), 2636-2651. <https://doi.org/10.3390/s7112636>
- [10] Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE transactions on geoscience and remote sensing*, 33(2), 457-465. <https://doi.org/10.1109/TGRS.1995.8746027>
- [11] Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... & Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of geophysics*, 45(2). <https://doi.org/10.1029/2005RG000183>
- [12] Tolan, J., Yang, H. I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., ... & Couprie, C. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300, 113888. <https://doi.org/10.1016/j.rse.2023.113888>
- [13] Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N. E., ... & Tarko, A. J. (2020). Copernicus global land service: Land cover 100m: Version 3 Globe 2015-2019: Product user manual. <https://doi.org/10.5281/zenodo.3938963>
- [14] Adelabu, S., Mutanga, O., & Adam, E. (2015). Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. *Geocarto International*, 30(7), 810-821. <https://doi.org/10.1080/10106049.2014.997303>
- [15] Agarwal, A., Kenney, A. M., Tan, Y. S., Tang, T. M., & Yu, B. (2023). MDI+: A flexible random forest-based feature importance framework. *arXiv preprint arXiv:2307.01932*. <https://doi.org/10.48550/arXiv.2307.01932>

REFERENCES

- [1] Tassi, A., Gigante, D., Modica, G., Di Martino, L., & Vizzari, M. (2021). Pixel-vs. Object-based landsat 8 data classification in google earth engine using random forest: The case study of maiella national park. *Remote sensing*, 13(12), 2299. <https://doi.org/10.3390/rs13122299>
- [2] Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 18. <https://doi.org/10.3390/s18010018>
- [3] Mohammadpour, P., Viegas, D. X., & Viegas, C. (2022). Vegetation mapping with random forest using sentinel 2 and GLCM texture feature—A case study for Lousã region, Portugal. *Remote Sensing*, 14(18), 4585. <https://doi.org/10.3390/rs14184585>
- [4] Junaid, M., Sun, J., Iqbal, A., Sohail, M., Zafar, S., & Khan, A. (2023). Mapping lulc dynamics and its potential implication on forest cover in malam jabba region with landsat time series imagery and random forest classification. *Sustainability*, 15(3), 1858. <https://doi.org/10.3390/su15031858>
- [5] Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International journal of remote sensing*, 39(9), 2784-2817. <https://doi.org/10.1080/01431161.2018.1433343>
- [6] Ouma, Y., Nkwae, B., Moalafhi, D., Odirile, P., Parida, B., Anderson, G., & Qi, J. (2022). Comparison of machine learning classifiers for multitemporal and multisensor mapping



of urban LULC features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 681-689. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-681-2022>

[7] Sait akimata Vostochno-Kazahstanskoy oblasti. [Website of the Akimat of the East Kazakhstan region.]. – Retrieved from: <https://www.gov.kz/memleket/entities/akimvko>

[8] Sait kataloga dannyh Google Earth Engine. [Google Earth Engine Data Catalog Site]. – Retrieved from: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2?hl=ru

[9] Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: a case study in high-density cypress forest. *Sensors*, 7(11), 2636-2651. <https://doi.org/10.3390/s7112636>

[10] Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE transactions on geoscience and remote sensing*, 33(2), 457-465. <https://doi.org/10.1109/TGRS.1995.8746027>

[11] Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... & Alsdorf, D. (2007). The shuttle radar topography mission. *Reviews of geophysics*, 45(2). <https://doi.org/10.1029/2005RG000183>

[12] Tolan, J., Yang, H. I., Nosarzewski, B., Couairon, G., Vo, H. V., Brandt, J., ... & Couprie, C. (2024). Very high resolution canopy height maps from RGB imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300, 113888. <https://doi.org/10.1016/j.rse.2023.113888>

[13] Buchhorn, M., Smets, B., Bertels, L., De Roo, B., Lesiv, M., Tsendbazar, N. E., ... & Tarko, A. J. (2020). Copernicus global land service: Land cover 100m: Version 3 Globe 2015-2019: Product user manual. <https://doi.org/10.5281/zenodo.3938963>

[14] Adelabu, S., Mutanga, O., & Adam, E. (2015). Testing the reliability and stability of the internal accuracy assessment of random forest for classifying tree defoliation levels using different validation methods. *Geocarto International*, 30(7), 810-821. <https://doi.org/10.1080/10106049.2014.997303>

[15] Agarwal, A., Kenney, A. M., Tan, Y. S., Tang, T. M., & Yu, B. (2023). MDI+: A flexible random forest-based feature importance framework. *arXiv preprint arXiv:2307.01932*. <https://doi.org/10.48550/arXiv.2307.01932>

**Денисова Н.Ф., Жусупова Г.К., Петрова О.А., Чепашев Д.В., Даумова Г.К.,
Темирбаева К.А.**

LANDSAT СПУТНИКТИК СУРЕТТЕРІ НЕГІЗІНДЕ МАШИНАЛЫҚ ОҚЫТУДЫ ПАЙДАЛАНА ОТЫРЫП, ШЫҒЫС ҚАЗАҚСТАН ОБЛЫСЫНДАҒЫ ӨСІМДІКТЕРДІ КАРТОГРАФИЯЛАУ

Андатпа. Бұл зерттеудің мақсаты – Landsat 8 спутниктік суреттері негізінде машиналық оқыту әдістерін қолдана отырып, Шығыс Қазақстан облысының өсімдіктерін картаға түсіру тәсілін әзірлеу. Мақалада мәліметтерді алдын ала өңдеу кезеңдері, жіктеудің оңтайлы алгоритмдерін таңдау және өсімдіктердің әртүрлі түрлері үшін олардың тиімділігін бағалау қарастырылады. Деректерді алдын ала өңдеу және жіктеу процесі Google Earth Engine бұлттық платформасында жүзеге асты. Жер бедері туралы ақпаратты қамтитын 17 көмекші қабаттардан басқа, деректерді жинау барысында зерттеу аймағы үшін оқу үлгілері де жасалып, 10 класқа жіктелді: жалпақ жапырақты ағаштар, қылқан жапырақты ағаштар, шөптесін өсімдіктер, бұталар, ашық жерлер, ауылшаруашылық жерлері, сулы-батпақты жерлер, елді мекендер, қар мен мұздар, су объектілері.



Зерттеу барысында векторлық қолдау әдісі (SVM), шешім ағаштары (CART), градиенттік бустинг ағаштары (GTB) және кездейсоқ орман (RF) сияқты әртүрлі машиналық оқыту алгоритмдері қолданылып, салыстырылды. GTB және RF алгоритмдері 92% Жалпы дәлдікпен және 91% Каппа коэффициентімен ең жоғары нәтижелер көрсетті, бұл қашықтықтан зондтау деректерін машиналық оқыту әдістерімен біріктірудің жоғары сапалығын растайды. Зерттеу нәтижелері аймақтағы экожүйелердің жағдайын бақылау, қоршаған ортаны қорғау шараларын жоспарлау, табиғи ресурстар мен төтенше жағдайларды басқару үшін машиналық оқыту әдістерінің әлеуетін көрсетеді.

Кілт сөздер: Шығыс Қазақстан облысы; Landsat; Машиналық оқыту; Өсімдіктердің жіктелуі; Google Earth Engine; Random Forest.

Denissova N., Zhussupova G., Petrova O., Chepashev D., Daumova G., Temirbayeva K.
VEGETATION MAPPING OF EAST KAZAKHSTAN REGION USING MACHINE
LEARNING BASED ON LANDSAT SATELLITE IMAGES

Annotation. The aim of this study is to develop an approach for mapping vegetation in the East Kazakhstan region using machine learning methods based on Landsat 8 satellite imagery. The paper examines data preprocessing stages, the selection of optimal classification algorithms, and the evaluation of their effectiveness for different vegetation types. The process of pre-processing and classification of data was carried out on the Google Earth Engine cloud platform. In addition to 17 auxiliary layers containing information about the terrain, training samples were formed during the data collection process and distributed into 10 classes: broadleaf trees, conifers, herbaceous vegetation, shrubs, open land, agricultural land, wetlands, built-up land, snow and ice, and water bodies.

Various machine learning algorithms were applied and compared in the study, including Support Vector Machines (SVM), Classification and Regression Trees (CART), Gradient Tree Boosting (GTB), and Random Forest (RF). The GTB and RF algorithms demonstrated the best results, with an Overall accuracy of 92% and a Kappa coefficient of 91%, indicating the high quality of integrating remote sensing data with machine learning methods. The obtained research results highlight the potential of machine learning methods for ecosystem monitoring, environmental planning, natural resource management, and emergency response in the region.

Keywords: East Kazakhstan region; Landsat; Machine Learning; Vegetation Classification; Google Earth Engine; Random Forest.